



Non-random assembly of bacterioplankton communities in the subtropical North Pacific Ocean

Alexander Eiler^{1,2,†}*, Darin H. Hayakawa^{1,3,†} and Michael S. Rappé¹

¹ Hawaii Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaii, Kaneohe, HI, USA

² Limnology, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

³ Department of Microbiology, University of Hawaii, Honolulu, HI, USA

Edited by:

Peter R. Girguis, Harvard University, USA

Reviewed by:

J. Michael Beman, University of California, Merced, USA

Frank Stewart, Georgia Institute of Technology, USA

Jed Fuhrman, University of Southern California, USA

*Correspondence:

Alexander Eiler, Limnology, Department of Ecology and Genetics, Norbyvägen 18D, SE-75236 Uppsala, Sweden.

e-mail: alexander.eiler@ebc.uu.se

[†] Alexander Eiler and Darin H.

Hayakawa have contributed equally to this work.

The exploration of bacterial diversity in the global ocean has revealed new taxa and previously unrecognized metabolic potential; however, our understanding of what regulates this diversity is limited. Using terminal restriction fragment length polymorphism (T-RFLP) data from bacterial small-subunit ribosomal RNA genes we show that, independent of depth and time, a large fraction of bacterioplankton co-occurrence patterns are non-random in the oligotrophic North Pacific subtropical gyre (NPSG). Pair-wise correlations of all identified operational taxonomic units (OTUs) revealed a high degree of significance, with 6.6% of the pair-wise co-occurrences being negatively correlated and 20.7% of them being positive. The most abundant OTUs, putatively identified as *Prochlorococcus*, SAR11, and SAR116 bacteria, were among the most correlated OTUs. As expected, bacterial community composition lacked statistically significant patterns of seasonality in the mostly stratified water column except in a few depth horizons of the sunlit surface waters, with higher frequency variations in community structure apparently related to populations associated with the deep chlorophyll maximum. Communities were structured vertically into epipelagic, mesopelagic, and bathypelagic populations. Permutation-based statistical analyses of T-RFLP data and their corresponding metadata revealed a broad range of putative environmental drivers controlling bacterioplankton community composition in the NPSG, including concentrations of inorganic nutrients and phytoplankton pigments. Together, our results suggest that deterministic forces such as environmental filtering and interactions among taxa determine bacterioplankton community patterns, and consequently affect ecosystem functions in the NPSG.

Keywords: 16S rRNA gene, assembly, community, gyre, oligotrophic, time series

INTRODUCTION

The bulk of the world's ocean is oligotrophic and, in terms of both biomass and ecosystem function, is dominated by microorganisms (Cho and Azam, 1988). Microbes with diverse, yet specific physiological requirements and ecological roles mediate nearly all energy and material fluxes within this vast biome (Azam, 1998; Herndl et al., 2008). These fluxes have been suggested to be products of multiple interactions within the microbial community, rather than those of single populations (Little et al., 2008; Strom, 2008). Therefore, an understanding of the processes that determine the composition of bacterioplankton communities is imperative in order to formulate accurate estimates and models of open ocean biogeochemistry. Despite increasing efforts to understand emergent patterns in the abundances and distributions of microbial taxa in seawater (e.g., Morris et al., 2002; Fuhrman et al., 2006; Johnson et al., 2006; Carlson et al., 2009; Eiler et al., 2009), the mechanisms that determine patterns observed in entire microbial communities remain largely unknown (Fuhrman and Steele, 2008; Steele et al., 2011). This stands in contrast to macroorganismal ecology, where the ecological mechanisms that underlie and maintain community patterns (beta-diversity) and (alpha-) diversity (often termed "assembly rules") have been a central theme (e.g., Keddy and

Weiher, 2001). Patterns in alpha- and beta-diversity of macroorganisms have been suggested to be the result of the interplay between regional and local processes: regional processes include dispersal, historical legacies, speciation, extinction, and geographic and climatic conditions, whereas local processes include interactions among organisms, and between organisms and their local environment.

Recently, microbial ecologists have begun to adopt many theoretical ecological frameworks upon which the microbial world can be explored and, eventually, predicted (e.g., Horner-Devine et al., 2004; Green and Bohannan, 2006; Martiny et al., 2006; Prosser et al., 2007). For example, microbial ecologists are now exploring whether microbial communities are distinct in (or specific to) different habitats, and if metrics of microbial diversity (alpha- and beta-diversity) show explainable patterns comparable to those of macroorganisms (e.g., Fuhrman et al., 2006; Horner-Devine et al., 2007; Prosser et al., 2007).

The balance between stochastic and deterministic processes in determining the structure of pelagic microbial communities in the open ocean has received little direct attention. If stochastic processes dominate, then open ocean communities may appear as random combinations of taxa, whereas, if deterministic

processes dominate, then predictable patterns of taxa distributions and abundances may be observed. In order to test if community assembly is mainly established non-randomly and thus by deterministic forces, computer simulations can be used that compare simulated, randomly assembled communities with experimental data. Such an assessment can be achieved by analyzing taxa co-occurrence patterns via the approach proposed by Diamond (1975) and modified by Gotelli and McCabe (2002). Using this approach, non-random co-occurrence patterns have been shown to be the rule for macroorganisms (Gotelli and McCabe, 2002), and have also been demonstrated for microorganisms in a number of habitats (Horner-Devine et al., 2007). Environmental filtering and competition (Diamond, 1975; Gotelli and McCabe, 2002), and/or biogeographic and evolutionary history (Vuilleumier and Simberloff, 1980; Cracraft, 1988) are thought to be important deterministic processes that determine community composition.

However, in order to explore the mechanisms that determine community composition, it is not sufficient to describe patterns in community structure and to demonstrate that co-occurrence patterns are random or non-random; the explicit mechanisms that underlie the emergent community patterns need to be identified. Statistical tools allow for the exploration of community drivers through the correlation and/or correspondence among taxa and between taxa and environmental variables.

In the present study, we aimed to explore patterns in bacterioplankton community structure within water samples collected from various depths over a 3-year period at Station ALOHA in the North Pacific subtropical gyre (NPSG), the world's largest oligotrophic marine ecosystem (Karl, 1999). Our analysis revealed that deterministic mechanisms prevail in assembling the bacterioplankton community at Station ALOHA, and that associations among taxa, as well as between taxa and environmental variables such as nutrient and pigment concentrations, are common and predictable.

MATERIALS AND METHODS

SITE DESCRIPTION AND ANCILLARY DATA

Station ALOHA (22°45'N, 158°00'W) is located in the NPSG 100 km north of Oahu, Hawaii. The Hawaii Ocean Time-series (HOT) program¹ (Karl and Lukas, 1996) has made repeated observations of the hydrography, chemistry, and biology of the water column at Station ALOHA approximately once per month since October 1988. Sampling and analytical protocols are described elsewhere (Dore and Karl, 1996; Karl and Lukas, 1996; Letelier et al., 1996; Karl et al., 2001; Church et al., 2005). Metadata used for multivariate statistical analyses were downloaded from the HOT Data Organization and Graphical System (HOT-DOGS) website². Environmental variables used in this study are listed in **Table A2** in Appendix.

SAMPLE COLLECTION AND DNA EXTRACTION

Seawater samples were collected from discrete depths using 12 L polyvinyl chloride (PVC) bottles attached to a CTD rosette sampler, and sub-sampled into acid-rinsed 4 L polypropylene bottles.

Three different sampling protocols were used throughout the study period (November 2004 to July 2008). For samples from cruises conducted between 2004–2006 (HOT cruises 165–185), 1–2 L of each sub-sample was filtered through 25 mm diameter, 0.22 µm pore-sized hydrophilic polyvinylidene difluoride (PVDF) membrane filters (Durapore; Millipore Corp., Billerica, MA, USA). The filters were placed in 2.0 mL microcentrifuge tubes containing 500 µL of Tris–EDTA (TE) buffer, immediately quick frozen in liquid nitrogen, and stored at –80°C until further processing. DNA was extracted and stored at –20°C following the protocol described by Church et al. (2009). For samples from cruises 186–193, 0.5–4 L of each sub-sample was filtered onto 13 mm diameter, 0.2 µm pore-sized polyethersulfone membrane filters (Supor-200; Pall Corp., Port Washington, NY, USA). The membrane filters were subsequently stored in 2 mL screw-cap tubes containing 250 µL of DNA lysis buffer [20 mM Tris HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1.2% (vol/vol) Triton X100, and 20 µg/mL lysozyme] at –80°C until extraction. For HOT cruises 197–200, ca. 4 L sub-samples were filtered onto 25 mm diameter, 0.2 µm pore-sized polyethersulfone membranes (Pall Corp.), and subsequently stored at –80°C in 500 µL of DNA lysis buffer. DNA was extracted from cruises 185–200 using the Qiagen DNeasy 96 Tissue Kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's recommended protocol for bacteria. Extracts were quantified using the PicoGreen fluorescent assay (Invitrogen Corp., Carlsbad, CA, USA) on a SpectraMax M2 plate reader (Molecular Devices Corp., Sunnyvale, CA, USA).

TERMINAL RESTRICTION FRAGMENT LENGTH POLYMORPHISM ANALYSIS

Two microliter of each DNA extract (equivalent to ca. 1–100 ng of genomic DNA) was used as template for the polymerase chain reaction (PCR) amplification of bacterial small subunit (SSU) ribosomal RNA (rRNA) genes. Individual 50 µL reactions contained 1× PicoMaxx reaction buffer (Stratagene, La Jolla, CA, USA), 200 µM of each deoxynucleotide, 200 nM each of the fluorescently labeled general bacterial SSU rRNA gene oligonucleotide primer 27F-B-6FAM (5'-[6~FAM]AGRGTTYGATYMTGGCTCAG-3') and universal SSU rRNA gene oligonucleotide primer 1492R (5'-GGYTACCTTGTTACGACTT-3'; Lane, 1991) and 0.625 units PicoMaxx thermostable DNA polymerase (Stratagene). Cycling conditions consisted of an initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C denaturation for 30 sec, 55°C annealing for 1 min, and 72°C extension for 2 min, with a final extension step at 72°C for 20 min. PCR products were purified using the QIAquick Multiwell PCR Purification system (Qiagen Inc.), and digested in 20 µL reactions containing 4 µg acetylated bovine serum albumin (BSA), 1X MULTI-CORE enzymatic reaction buffer (Promega, Madison, WI, USA), and 10 units of *Hae*III restriction endonuclease (Promega). After a 6-h incubation at 37°C, the digests were terminated by incubation at 70°C for 20 min, and purified using the Millipore MultiScreen Assay System (Millipore Corp.) in conjunction with Sephadex G-50 Superfine (GE Healthcare, Piscataway, NJ, USA). Approximately 100 ng of the purified products were electrophoresed on an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Operational

¹http://hahana.soest.hawaii.edu/hot/hot_jgofs.html

²<http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html>

taxonomic units (OTUs) were identified as peaks between 30 and 600 bp in length, using GeneMapper v4.0 software (Applied Biosystems, Foster City, CA, USA). Terminal restriction fragment length polymorphism (T-RFLP) profiles were first run through an in-house MATLAB (The MathWorks, Natick, MA, USA) script that corrects for instrumental drift while rounding fractional peaks to whole numbers. In brief, this script identified peaks with ambiguous decimal increments (i.e., $x.3$ – $x.6$) and ensured that they were rounded to the same integer throughout all profiles by using a majority rule (i.e., rounded up if most were $> x.5$ and vice versa). Non-ambiguous decimal increments (i.e., $< x.3$ and $> x.6$) were simply rounded to their nearest integer. The profiles then underwent a limited manual curation in order to further compensate for instrumental drift, and finally normalized using a variable threshold (Osborne et al., 2006). Peaks were transformed into relative units by dividing integrated peak areas by the total peak area for that sample. Diversity (evenness) was estimated from the relative units using the Gini coefficient (Wittebolle et al., 2009). The Gini coefficient provides a proxy for the equality of taxa distributions: the higher the Gini coefficient, the more uneven a community is.

Putative identification of abundant OTUs was performed using clone library data obtained from both coastal and open ocean sites in the vicinity of Hawaii, including Station ALOHA. Representative clones were characterized by Sanger sequencing and fragment sizing after restriction digests to correlate phylogenetic identity with terminal restriction fragment length. Sequences were deposited in Genbank under accession number JN166096 to JN166367.

Statistical analyses. Statistical analyses were performed using the software packages PRIMER 6 (PRIMER-E Ltd., Ivybridge, UK), EcoSim (Gotelli and McCabe, 2002), PAST (Hammer et al., 2001), and MATLAB (The Mathworks). Sorenson (Bray–Curtis) distances were calculated from rank-transformed T-RFLP data using a natural weighting as recommended by McCune and Grace (2002), and visualized using non-metric multidimensional scaling (NMS; McCune and Grace, 2002).

Statistical analyses required the exclusion of samples with missing values. In order to account for such gaps in the availability of metadata, environmental variables were grouped into biological, chemical, and physical data classes in order to increase sample size for analyses (Table A2 in Appendix). Within PRIMER 6, BIOENV procedures were used to explore the correspondence between microbial community structure and environmental metadata (Clarke and Warwick, 2001). In this analysis, the environmental variables were first $\text{Log}(X + 1)$ transformed and normalized by subtracting the mean from each entry of a single variable and dividing by the standard deviation for that variable. This allowed the identification of variables or combinations of variables that best corresponded with the community data. As this analysis was based on a limited number of samples due to missing data, the best fitting combinations were re-analyzed after adding samples with non-missing values (in the identified variable combination) using the RELATE test (Clarke and Warwick, 2001) in PRIMER 6 (PRIMER-E Ltd.). The Analysis of Similarity (ANOSIM in PRIMER 6; Clarke, 1993; Clarke and Warwick, 2001) test was used to examine for differences between bacterial communities from different depth horizons in the water column.

Two taxa co-occurrence indices were calculated using the EcoSim 700 software package (Gotelli and McCabe, 2002). The Checkboard index (*C-board*) calculates the number of taxa that never co-occur, which can be interpreted as evidence for competitive exclusion (segregation) or differences in habitat preference. The second index (*C-score*) is calculated by: $CU = (R_i - S)(R_j - S)$ where R_i and R_j are the number of sites where taxon i and j occur, and S is the number of sites containing both taxa. The *C-score* is then calculated from the mean number of all pair-wise CUs. Monte Carlo simulations were used to test if the observed co-occurrence measures were random following the most conservative permutation procedure. Following Gotelli and McCabe (2002), we calculated a standardized effect size (SES) that allowed comparing the degree of co-occurrence across temporal and spatial scales.

Spearman's rank correlation coefficient (ρ_s) was calculated in MATLAB in a pair-wise fashion between OTUs, using the relative contribution of each OTU to the total community for each sample. Only OTUs that were detected in more than two independent samples were included in the analysis ($n = 397$). A significance level of $p < 0.05$ was used to calculate the percentage of significant positive and negative correlations from the total number of pair-wise comparisons. In addition, Bonferroni correction was used to address the problem of multiple comparisons. In the same way, ρ_s between relative contributions of each OTU and depth were calculated.

To test if the number and distribution of significant pair-wise correlations was different from random, Monte Carlo simulations were performed. Following a permutation procedure where the rank of OTUs within samples was randomized, a total of 1000 randomized matrices were created and subsequently used in pair-wise correlations. The maximum and minimum ρ_s values of correlations within each 0.01 interval (from 1 to -1) were determined from all 1000 randomized matrices.

RESULTS

CO-OCCURRENCE PATTERNS AND CORRELATIONS BETWEEN OTUs

A total of 461 different OTUs, each represented by a specific terminal restriction fragment length, were detected in 296 samples collected over 3.5 years (November 2004 to February 2008) from depths ranging from 5 to 4000 m at Station ALOHA. Of these, 397 OTUs were present in at least two T-RFLP profiles, and so were used for further analyses. Co-occurrence indices calculated from this dataset differed significantly from random simulations: more OTU pairs exhibited larger *C-scores* and *C-boards* than expected by chance (Table 1). A higher *C-score* than expected by chance indicates low co-occurrence, whereas

Table 1 | Co-occurrence indexes (*C-score* and *C-board*) and their standardized effective sizes (SES) for bacterioplankton communities from Station ALOHA sampled during the course of this study ($n = 296$).

Index	Observed score	Simulated mean score	Variance	p	SES
<i>C-score</i>	499	470	0.14	>0.001	77.79
<i>C-board</i>	49536	45825	37378	>0.001	19.20

a lower C -score than expected by chance indicates high co-occurrence. Low co-occurrence can be interpreted as evidence for competitive exclusion (segregation), or differences in habitat preference. Of the Spearman's rank correlation coefficients (ρ_s) from all pair-wise comparisons of OTUs ($n = 78,607$), 27.3% were significant at a 5% significance level, of which 6.6% were negatively correlated ($\rho_s < 0$) and 20.7% were positively correlated ($\rho_s > 0$; **Figure 1A**). Both significantly positive and negative correlations

occurred more often than expected by chance when compared to simulations where the rank order of species were randomized within a community. In addition, maximum ρ_s values were much higher for the observed matrix compared to the simulated matrices (**Figure 1A**). Using Bonferroni correction for addressing the problem of multiple comparisons, the number of significant correlations decreased from 21,485 to 3,276 (3.5% significant positive and 0.5% significant negative correlations).

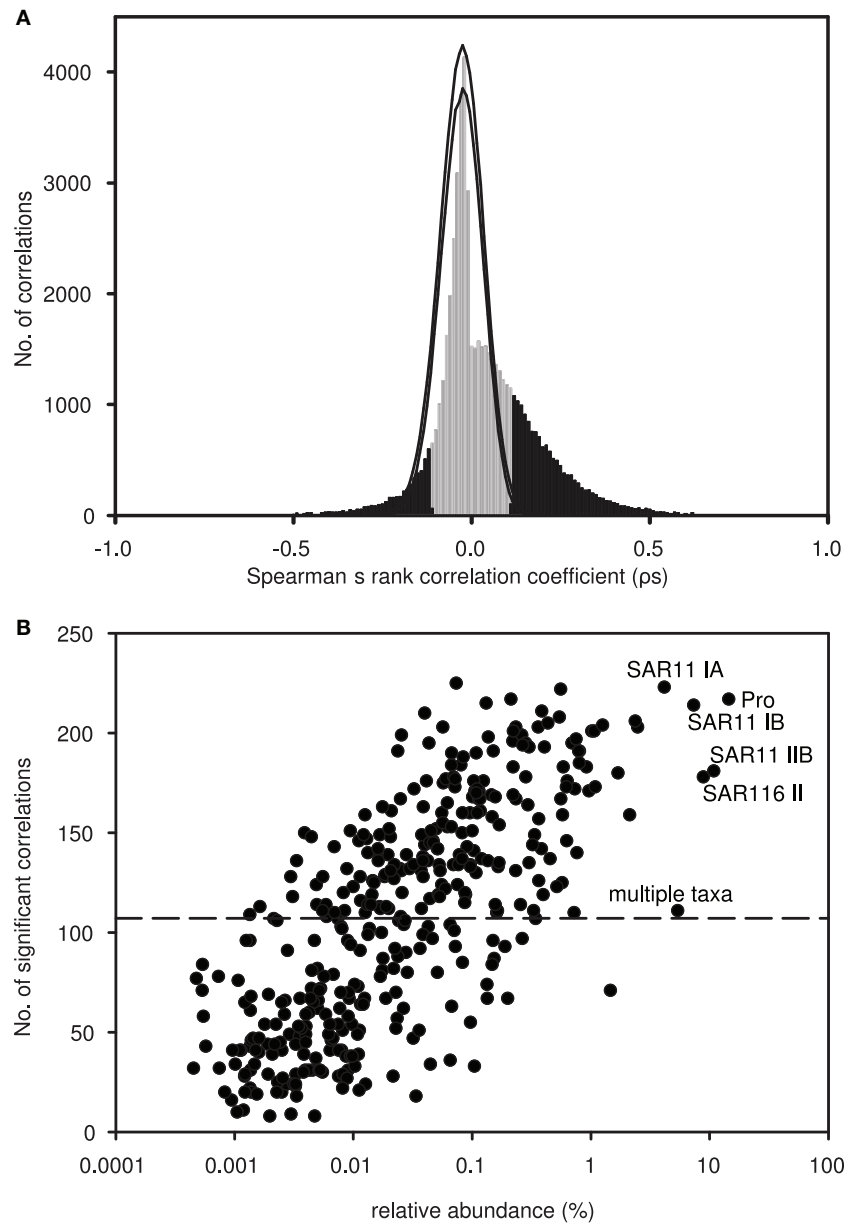
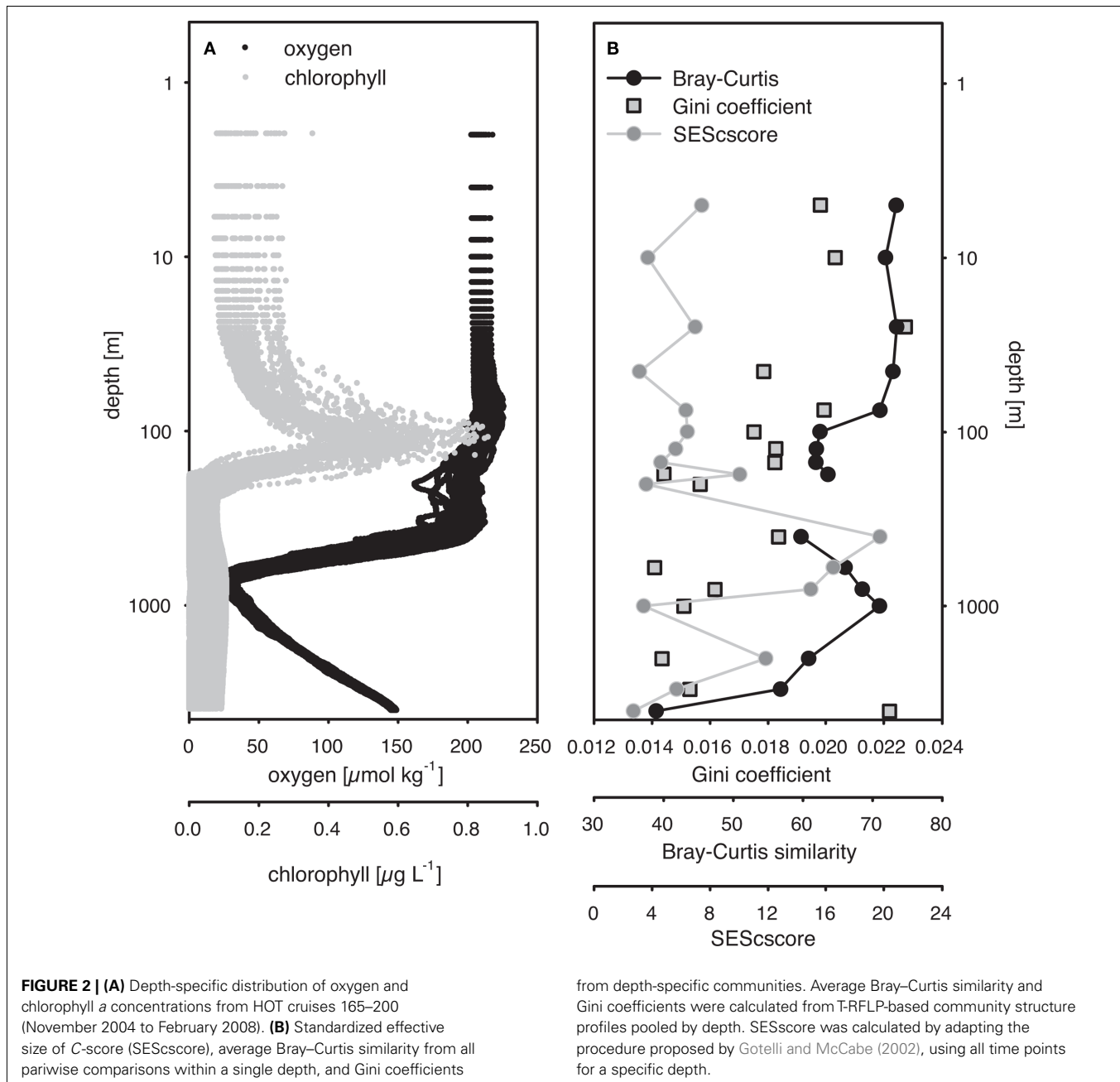


FIGURE 1 | (A) Spearman's rank correlation coefficients (ρ_s) from pair-wise correlations of all OTU pairs ($n = 78,607$), plotted as a frequency histogram using 0.01 intervals. Significant correlations ($p < 0.05$) are indicated in black. Monte Carlo simulations were used to test if the observed histogram was different from random. Following a permutation procedure where OTUs were randomized within sites, a total of 1000 randomized matrices were created, and pair-wise correlations were performed on each matrix. The maximum and

minimum numbers of correlations within each 0.01 interval (from 1 to -1) were determined from all 1000 randomized matrices (hismax and hismin; black lines). **(B)** Plot of the average relative abundance of each OTU, determined via T-RFLP, against the number of significant pair-wise correlations to other OTUs. Putative identities are indicated for the most abundant OTUs. The dashed line indicates the median number of significant pair-wise correlations.

On average, 108 significant correlations were observed per OTU; 22 OTUs were correlated with more than 50% of the 397 OTUs included in the dataset. This subgroup of 22 included the most abundant OTUs, which are putatively identified as *Prochlorococcus* and SAR11 subclades IB and IA. Other abundant OTUs putatively identified as SAR116 subclade II and SAR11 subclade IIB were significantly correlated with >150 OTUs. However, many OTUs contributing <1% in relative abundance (on average) to the total bacterioplankton community exhibited a high number of significant correlations with other OTUs in our dataset (Figure 1B). Hence, we did not observe a significant correlation between the average relative abundance of an OTU and the number of correlations it exhibited.

Variation in the *C*-score along the depth gradient was investigated in detail by calculating the SES from *C*-score calculations grouped by depth (SESScore; Figure 2). A positive SESScore significantly different from random ($p < 0.01$) was observed at all depths, indicating low co-occurrence. The SESScore was not correlated with depth, but exhibited intriguing variability along the vertical axis (Figure 2). The highest value was estimated at 400 m, and decreased again toward the oxygen minimum zone at ca. 800 m. The SESScore did not vary with the size of the data set, as measured by either the number of samples or the number of different OTUs included in a given incidence matrix ($\rho = 0.103$, $p = 0.696$, and $\rho = 0.359$, $p = 0.137$, respectively). It could also be argued that the number of OTUs represents an estimate of



community richness. Hence, this apparent lack of relationship between richness and SEScore, together with evenness and SEScore also not being related ($\rho = -0.279$, $p = 0.276$) indicates that co-occurrence patterns are not related to alpha diversity at least in our study.

The Gini coefficient, an estimate of community evenness, ranged from 0.007 to 0.067 in the 296-sample dataset. These low values indicate a high degree of equality, as a value of 0 indicates complete equality while a value of 1 indicates maximal inequality among OTUs. Spearman's rank correlation analysis revealed a significant negative relationship between the Gini coefficient and depth ($\rho = -0.27$, $p < 0.001$) indicating an increase in community evenness with depth (Figure 2).

CORRESPONDENCE BETWEEN MICROBIAL COMMUNITY STRUCTURE AND ENVIRONMENTAL VARIABLES

In order to visualize potential spatial and temporal patterns in community structure, communities were plotted using NMS to two dimensions based on a Bray–Curtis distance matrix (Figure 3). ANOSIM and RELATE analyses corroborated initial visual observations of depth-specific clustering of bacterioplankton communities at Station ALOHA ($R = 0.613$, $p < 0.001$, and $\rho = 0.67$, $p < 0.001$, respectively). An R -value of 0 resulting from the ANOSIM analysis indicates that there were no differences among the communities from different depths, while an R -value of close to 1 would indicate that the dissimilarities between communities from different depths were larger than those from each individual depth. However, pair-wise ANOSIM revealed that community structure from different depth horizons in the upper 100 m were not significantly different whereas, below 100 m, all depth horizons were significantly different from each other (Table A1 in Appendix). Exceptions were adjunct depths where, in most cases, differences were not significant (Table A1 in Appendix).

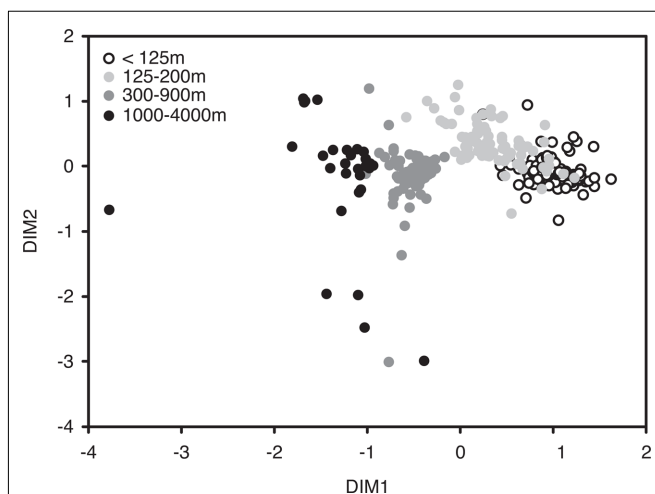


FIGURE 3 | Non-metric multidimensional scaling (NMS) of bacterial community structure based on a T-RFLP analysis of bacterial SSU rRNA genes amplified from bacterioplankton samples collected approximately monthly at Station ALOHA from November 2004 to February 2008 ($n = 296$; stress-value = 12.1).

Average Bray–Curtis dissimilarities between communities from different depths increased with increasing distance along the vertical (i.e., depth) profile. The significant ρ value of the RELATE analysis indicated that communities from more distant depth horizons were more different than communities from adjunct depths. When Bray–Curtis similarities were averaged over the sampling period for each depth, values stayed rather constant above the mixed layer (Figure 2). In the depth layers where the mixed layer and chlorophyll maximum temporally oscillate, Bray–Curtis similarity averages were low. The relative abundance of nearly early half (46%) of all OTUs was significantly correlated with depth ($p < 0.05$), of which 32% were positive and 14% negative (Figure 4).

Unlike the clear depth-specific zonation of bacterioplankton communities, temporal patterns were not readily apparent in NMS plots (data not shown). However, we were able to statistically test for temporal patterns in bacterioplankton community structure, since the sampling schema spanned over more than three consecutive years. When months and seasons were used as discriminating factors in ANOSIM analyses, and when months, seasons, and Julian days were used in cyclic RELATE analyses, significant temporal patterns were observed at various depths. However, significant annual cycles in bacterioplankton community composition were only found in the upper 200 m of the water column (Table 2) and were not detected in the deeper horizons.

BIOENV procedures were employed to search for relationships (correspondence) between community and environmental variability, and to identify the strongest relationships. Since the environmental dataset was incomplete, variables were divided into three categories: physiochemical, biogeochemical, and biological (Table A2 in Appendix). Data from each category was used separately in the BIOENV procedure, and the variable combinations that best matched the patterns in community composition were identified and then analyzed by RELATE after inflating the number of samples (Table 3). Parameters from all three general categories revealed significant correspondence along the sampled spatial and temporal scales (Table 3). This included depth, which represented the single variable that corresponded the best with community composition. Variation in bacterioplankton community structure in the upper 200 m of the water column was highly correlated with concentrations in heterotrophic prokaryotes, *Prochlorococcus*, *Synechococcus*, and pigments such as chlorophyll *a* and phaeopigments. In addition, inorganic nutrients such as nitrite, nitrate, and particulate carbon corresponded well with community composition. Most of these variables were also correlated with depth (Table A3 in Appendix), and annual periodicity in these variables was either weak or not present (data not shown).

DISCUSSION

In this study, we sought to understand the mechanisms that underlie the patterns of community composition and maintain biodiversity over space and time (in other words, “to formulate community assembly rules”). Our results suggest that non-random patterns of co-occurrence are common between bacterioplankton lineages in the NPSG. This is consistent with observations made previously in other ocean provinces (Fuhrman et al., 2006; Gilbert et al.,

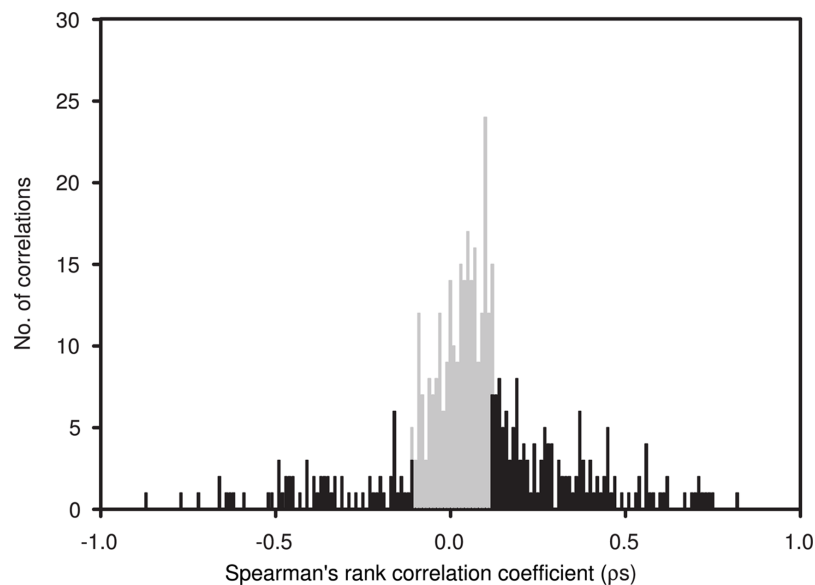


FIGURE 4 | Spearman's rank correlation coefficients (ρ_s) between depth and the relative abundance of individual OTUs, determined by T-RFLP (number of OTUs = 397), plotted in a frequency histogram using 0.01 intervals. Significant correlations ($p < 0.05$) are indicated in black. ρ_s Values indicate the strength of the relationship, with positive values indicating an increase in the relative contribution of an OTU with depth and negative values indicating a decrease.

Table 2 | Statistical analyses of temporal patterns in bacterioplankton community structure at Station ALOHA.

Depth (m)	No.	ANOSIM						RELATE			
		Month		Season		Day		Month		Season	
		<i>R</i>	<i>p</i>	<i>R</i>	<i>p</i>	<i>R</i>	<i>p</i>	ρ	<i>p</i>	ρ	<i>p</i>
5	13	0.076	0.390	0.323	0.012	0.106	0.154	0.178	0.036	0.077	0.143
10	9	0.148	0.277	-0.314	0.833	-0.004	0.402	-0.034	0.479	-0.236	0.750
25	16	0.120	0.314	0.268	0.024	0.136	0.066	0.175	0.018	0.134	0.076
45	21	0.139	0.199	-0.002	0.464	-0.030	0.671	-0.032	0.671	-0.040	0.765
75	28	0.260	0.013	0.162	0.015	0.095	0.048	0.084	0.070	0.106	0.032
100	33	-0.033	0.652	0.024	0.290	0.035	0.153	0.039	0.117	0.041	0.122
125	29	0.146	0.096	0.102	0.062	0.045	0.161	0.062	0.093	0.105	0.018
150	18	0.293	0.066	0.280	0.001	0.124	0.022	0.154	0.007	0.207	0.001
175	26	0.089	0.183	0.070	0.163	0.184	0.002	0.167	0.004	0.163	0.001
200	6	NA	NA	-0.091	0.567	0.420	0.041	0.483	0.018	0.018	0.371
400	13	0.036	0.398	-0.130	0.847	0.015	0.415	-0.085	0.696	-0.111	0.722
600	16	0.037	0.344	-0.068	0.695	-0.057	0.579	-0.029	0.553	-0.111	0.886
800	15	0.120	0.220	-0.034	0.532	-0.008	0.490	-0.008	0.511	-0.042	0.511
1000	16	0.126	0.194	-0.100	0.789	-0.088	0.708	-0.097	0.734	-0.067	0.604
2000	10	-0.079	0.608	0.043	0.342	-0.147	0.716	-0.058	0.554	-0.111	0.635
3000	7	0.100	0.476	-0.019	0.457	-0.193	0.725	-0.013	0.455	-0.016	0.437
4000	8	0.038	0.448	0.144	0.315	-0.085	0.485	-0.081	0.518	-0.036	0.469

Significant seasonal patterns ($p < 0.05$) identified by RELATE and ANOSIM are indicated in bold. Analyses were performed on community data sorted either by seasons, months, or Julian days.

2009; Steele et al., 2011), and demonstrates that both the emergent depth-dependent and temporal patterns in bacterioplankton taxa distributions in the NPSG derive mainly from deterministic

forces, like controlling environmental conditions and taxa interdependencies. Documenting non-random patterns of co-occurrence for members of natural microbial communities is an important

Table 3 | Statistical analysis of correspondence between bacterioplankton community structure and environmental parameters.

Environmental subgroup	Tested variables	No.	ρ	p
Physicochemical	Depth, pH, alkalinity, temperature, salinity	131	0.75	>0.001
	Depth, pH	139	0.82	>0.001
	Depth	296	0.67	>0.001
Biogeochemical	PO ₄ , NO ₂ /NO ₃ , sil, pP, pN, pSi, pC, pC:pN, pSi:sil, pC:pSi	45	0.49	>0.001
	NO ₂ /NO ₃ , pC	98	0.65	>0.010
Biological	Heterotrophic prokaryotes, <i>Prochlorococcus</i> , <i>Synechococcus</i> , chlorophyll <i>a</i> , various pigments	137	0.46	>0.001

The best matching variables in each category (physicochemical, biogeochemical, and biological) resulting from BIOENV procedures are given. These variables were then analyzed by RELATE test to increase the number of samples since the initial data sheet including all variables contained only a limited number of samples ($n < 50$). The ρ and p -values from the RELATE analyses are shown. Sample size (n) indicates the number of samples that were used for each analysis. A $p < 0.05$ denotes a significant correspondence between environmental and community dynamics.

step in establishing generalities for community assembly, and confirms previous predictions regarding the non-random assembly of microbial communities in other environments (Horner-Devine et al., 2007). In this meta-analysis, no clear trends in co-occurrence patterns could be observed in relation to microbial habitats or phylogenetic groups but trends could be attributed to the variability in sample sizes and analysis methods of the different studies.

In our analysis, SESscore values were not related to sample size, even though the number of samples analyzed from each depth ranged from 6 to 33. In addition, proxies for alpha-diversity (GINI index and the total number of OTUs) and beta-diversity (Bray–Curtis similarity) were not related to SESscore values, suggesting that community properties are not likely effecting co-occurrence estimates. Still, we observed depth-specific variability in co-occurrence indices (SESscore) within the predominantly stratified water column of the NPSG that could be attributed to the presence of complex and depth-specific ecological forces structuring bacterioplankton communities in the oligotrophic ocean. High SESscore values appear to coincide with steep, depth-specific gradients in such environmental variables as chlorophylls and oxygen concentrations (Figure 2). When coupled with the observation that bacterioplankton communities are stratified in the NPSG, it appears that deterministic mechanisms, such as environmental selection and competition, are stronger in these deep horizons of the water column than, for example, stochastic mixing events that transport nutrients and organisms through the water column. In addition, these stratified bacterial communities were maintained consistently for 3.5 years. An exception to this was observed above the depth of the mixed layer, where communities were fairly homogeneous (Figure 3; Table A2 in Appendix). This stands in contrast to the deep mixing events that disrupt the stratified bacterioplankton communities at the Bermuda Atlantic Time-series Study (BATS) site in winter, producing large shifts in community composition in the epipelagic and upper mesopelagic (Morris et al., 2005; Treusch et al., 2009).

While simple randomization tests like those applied in this study are useful statistical tools for recognizing non-random patterns of distribution (Gotelli, 2001), they do not allow for the identification of the causal underlying mechanisms shaping bacterioplankton communities in this system (Roughgarden, 1983). If the goal is to predict bacterioplankton community composition

in the oligotrophic ocean, the causal underlying mechanisms need to be disentangled and their relative importance quantified. The application of statistical tools to identify relationships between community structure and abiotic and biotic environmental properties, as well as relationships among bacterial groups, is a first step toward such predictions (e.g., Fuhrman et al., 2006; Gilbert et al., 2009; Treusch et al., 2009). Using permutation-based procedures such as ANOSIM, we observed highly structured bacterioplankton communities along depth gradients, and a high number of individual OTUs either positively or negatively correlated with depth, allowing predictions on their distribution along the stratified water column. Numerous studies have demonstrated that specific groups of planktonic microorganisms are vertically stratified in the open ocean (e.g., DeLong, 1992; Giovannoni et al., 1996; Field et al., 1997; Church et al., 2005; DeLong et al., 2006; Johnson et al., 2006; Pham et al., 2008). While not surprising when considering the existence of strong environmental gradients along the vertical profile (Table A3 in Appendix), disentangling depth from correlations with key environmental parameters remains an important but unresolved task.

In addition to identifying that different depths harbor distinct bacterioplankton communities, we were able to characterize these differences in evenness, SESscore and variance in community similarity (Figure 2B). Community evenness increased with depth, whereas the variability in community composition (as indicated by Bray–Curtis similarities) decreased with depth, clearly showing that the epipelagic bacterioplankton community was more dynamic than the bathylopelagic community. Higher variability in epipelagic communities was also reflected by the subtle but statistically significant seasonal oscillations, suggesting re-occurring temporal patterns as proposed for other ocean sites (Fuhrman et al., 2006; Treusch et al., 2009). There was also a clear correspondence between epipelagic bacterioplankton community structure and environmental variables that do exhibit some seasonality at Station ALOHA (Karl, 2002). The relationship between bacterioplankton community composition and nitrite, nitrate, and particulate carbon concentrations and the abundance of *Prochlorococcus*, *Synechococcus*, and pigments indicate that resource availability (or bottom up control) plays a dominant role in the assembly of bacterioplankton communities in the NPSG (Torsvik et al., 2002). Previous studies also

point to the importance of top down controls such as viruses and grazers (Thingstad and Lignell, 1997; Fuhrman, 1999; Hewson et al., 2006). Besides trophic interactions such as bacterivory and viral lysis, other forms of interdependencies have attracted little attention, especially antagonism and mutualism among open ocean bacterioplankton taxa.

Correlations among taxa may provide preliminary indications regarding the nature of bacterioplankton interactions. For example, a positive correlation may reflect syntrophy or other interdependency between two taxa, though it may also reflect similar responses to controlling environmental conditions, leading to the two taxa following a similar environmental trend. A negative correlation, on the other hand, may reflect either direct competition, or differences in growth constraints between taxa. Still, the higher number of positive than negative correlations among taxa observed in this study could partially result from a higher degree of mutualistic than antagonistic interactions. Metagenomic sequence data obtained along a depth profile in the NPSG support this view, as many genes with potential interactive functions have been identified, such genes encoding for chemotaxis, antibiotics, pili, and flagellate synthesis (DeLong et al., 2006). However, metabolic interactions between microorganisms resulting from the biotransformation of metabolites (e.g., from reduced to oxidized states and vice versa) do not rely on close spatial association, and hence may be far more important than interactions requiring close physical proximity. Many transformations in the carbon, nitrogen, and sulfur cycles are reliant on appropriate cellular and ambient microenvironmental redox conditions (see for example, Paerl and Pinckney, 1996, and references therein). Genome analyses have also emphasized the importance of metabolites for interactions in pelagic marine environments (Tripp et al., 2008). Although many of the identified correlations among taxon-pairs may be the result of such metabolic interactions, experimental validation is required

in order to explore the causality behind the observed correlations. A logical starting point for such studies are the highly correlating taxa, which may represent groups with extensive and strong interactions to other taxa. The high relative abundance of some of these taxa (i.e., *Prochlorococcus* and SAR11) already suggests a key role in open ocean bacterioplankton communities.

CONCLUSION

The non-random temporal and spatial patterns in microbial community structure we observed are an important starting point for investigating microbial ecosystem services in the NPSG (Fuhrman et al., 2006). For example, if interdependencies including mutualism, antagonism, trophodynamics (predation and parasitism), and competition determine community structure, then it can be speculated that these interactions directly underlay ecosystem processes by modifying pathways of energy and material flow, or indirectly by modifying the abundances of organisms or traits with strong ecosystem effects. Thus, understanding community interactions is essential if we want to predict ecosystem functioning of the oligotrophic ocean in an era of anthropogenic induced environmental change.

ACKNOWLEDGMENTS

We thank Matthew J. Church for his comments and suggestions on an early version of this manuscript, and for providing genomic DNA samples from a portion of the HOT cruises investigated here. We thank J. Becker, M. Miller, T. Campbell, S. Yeo, J. Coloma, N. Wagoner, E. Hambleton, and all HOT personnel for assistance in sampling and contextual data. This research was supported by funding from the Center for Microbial Oceanography: Research and Education (NSF Science and Technology Center Award EF-0424599). This is SOEST contribution 8206 and HIMB contribution 1452.

REFERENCES

- Azam, F. (1998). Microbial control of oceanic carbon flux, the plot thickens. *Science* 280, 694–696.
- Carlson, C. A., Morris, R., Parsons, R., Treusch, A. H., Giovannoni, S. J., and Vergin, K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* 3, 283–295.
- Cho, B. C., and Azam, F. (1988). Major role of bacteria in biogeochemical fluxes in the oceans interior. *Nature* 332, 441–443.
- Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Global Biogeochem. Cycles* 23, GB2020.
- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71, 5362–5370.
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143.
- Clarke, K. R., and Warwick, R. M. (2001). *Change in Marine Communities, An Approach to Statistical Analysis and Interpretation*, 2nd Edn. Plymouth: PRIMER-E.
- Cracraft, J. (1988). Deep-history biogeography, retrieving the historical pattern of evolving continental biotas. *Syst. Zool.* 37, 221–236.
- DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689.
- DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N.-U., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R., Chisholm, S. W., and Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311, 496–503.
- Diamond, J. M. (1975). "Assembly of species communities," in *Ecology and Evolution of Communities*, eds M. L. Cody and J. M. Diamond (Cambridge, MA: Harvard University Press), 342–444.
- Dore, J. E., and Karl, D. M. (1996). Nitrite distributions and dynamics at station ALOHA. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 385–402.
- Eiler, A., Hayakawa, D. H., Church, M. J., Karl, D. M., and Rappé, M. S. (2009). Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre. *Environ. Microbiol.* 11, 2291–2300.
- Field, K. G., Gordon, D., Wright, T., Rappé, M., Urbach, E., Vergin, K., and Giovannoni, S. J. (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* 63, 63–70.
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548.
- Fuhrman, J. A., Hewson, I., Schwalbach, M. S., Steele, J. A., Brown, M. V., and Naam, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13104–13109.
- Fuhrman, J. A., and Steele, J. A. (2008). Community structure of marine bacterioplankton, Patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* 53, 69–81.
- Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P. J., Huse, S., and Joint, I. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* 11, 3132–3139.

- Giovannoni, S. J., Rappé, M. S., Vergin, K. L., and Adair, N. L. (1996). 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7979–7984.
- Gotelli, N. J. (2001). Research frontiers in null model analysis. *Glob. Ecol. Biogeogr.* 10, 337–343.
- Gotelli, N. J., and McCabe, D. J. (2002). Species co-occurrence, a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* 83, 2091–2096.
- Green, J., and Bohannan, B. J. M. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol. Evol. (Amst.)* 21, 501–507.
- Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). PAST, paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* 4, 9.
- Herndl, G. J., Agogue, H., Baltar, F., Reinthaler, T., Sintes, E., and Varela, M. M. (2008). Regulation of aquatic microbial processes, the “microbial loop” of the sunlit surface waters and the dark ocean dissected. *Aquat. Microb. Ecol.* 53, 59–68.
- Hewson, I., Steele, J. A., Capone, D. G., and Fuhrman, J. A. (2006). Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar. Ecol. Prog. Ser.* 311, 67–77.
- Horner-Devine, M. C., Carney, K. M., and Bohannan, B. J. M. (2004). An ecological perspective on bacterial biodiversity. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 113–122.
- Horner-Devine, M. C., Silver, J. M., Leibold, M. A., Bohannan, B. J. M., Colwell, R. K., Fuhrman, J. A., Green, J. L., Kuske, C. R., Martiny, J. B., Muyzer, G., Ovreås, L., Reysenbach, A. L., and Smith, V. H. (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88, 1345–1353.
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311, 1737–1740.
- Karl, D. M. (1999). A sea of change, biogeochemical variability in the North Pacific subtropical gyre. *Ecosystems* 2, 181–214.
- Karl, D. M. (2002). Nutrient dynamics in the deep blue sea. *Trends Microbiol.* 10, 410–418.
- Karl, D. M., Bjorkman, K. M., Dore, J. E., Fujieki, L., Hebel, D. V., Houlihan, T., Letelier, R. M., and Tupasa, L. M. (2001). Ecological nitrogen-to phosphorus stoichiometry at station ALOHA. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 48, 1529–1566.
- Karl, D. M., and Lukas, R. (1996). The Hawaii Ocean Timeseries (HOT) program, background, rationale and field implementation. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 129–156.
- Keddy, P., and Weiher, E. (2001). “Introduction, the scope and goals of research on assembly rules,” in *Ecological Assembly Rules – Perspectives, Advances, Retreats*, eds E. Weiher and P. Keddy (Cambridge: Cambridge University Press), 1–22.
- Lane, D. J. (1991). “16S/23S rRNA sequencing,” in *Nucleic Acid Techniques in Bacterial Systematics*, eds E. Stackebrandt and M. Goodfellow (New York: John Wiley and Sons), 115–175.
- Letelier, R. M., Dore, J. E., Winn, C. D., and Karl, D. M. (1996). Seasonal and interannual variations in photosynthetic carbon assimilation at station ALOHA. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 467–490.
- Little, A. E. F., Robinson, C. J., Peterson, S. B., Raffa, K. E., and Handelsman, J. (2008). Rules of engagement, interspecies interactions that regulate microbial communities. *Annu. Rev. Microbiol.* 62, 375–401.
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krums, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Ovreås, L., Reysenbach, A. L., Smith, V. H., and Staley, J. T. (2006). Microbial biogeography, putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112.
- McCune, B., and Grace, J. B. (2002). *Analysis of Ecological Communities*. Gleneden Beach, OR: MjM Software Design.
- Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., and Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806–810.
- Morris, R. M., Vergin, K. L., Cho, J.-C., Rappé, M. S., Carlson, C. A., and Giovannoni, S. J. (2005). Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic time-series study site. *Limnol. Oceanogr.* 50, 1687–1696.
- Osborne, C. A., Rees, G. N., Bernstein, Y., and Janssen, P. H. (2006). New threshold and confidence estimates for terminal restriction fragment length polymorphism analysis of complex bacterial communities. *Appl. Environ. Microbiol.* 72, 1270–1278.
- Paerl, H. W., and Pinckney, J. L. (1996). A mini-review of microbial consortia, their roles in aquatic production and biogeochemical cycling. *Microb. Ecol.* 31, 225–247.
- Pham, V. D., Konstantinidis, K. T., Paiden, T., and DeLong, E. F. (2008). Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific subtropical gyre. *Environ. Microbiol.* 10, 2313–2330.
- Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Feckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, A. M., Solan, M., van der Gast, C. J., and Young, J. P. (2007). Essay – The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 384–392.
- Roughgarden, J. (1983). Competition and theory in community ecology. *Am. Nat.* 122, 583–601.
- Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., Chow, C. E. T., Sachdeva, R., Jones, A. C., Schwalbach, M. S., Rose, J. M., Hewson, I., Patel, A., Sun, E., Caron, D. A., Fuhrman, J. A. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* doi: 10.1038/ismej.2011.24
- Strom, S. L. (2008). Microbial ecology of ocean biogeochemistry, a community perspective. *Science* 320, 1043–1045.
- Thingstad, T. F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* 13, 19–27.
- Torsvik, V., Ovreås, L., and Thingstad, T. F. (2002). Prokaryotic diversity – magnitude, dynamics and controlling factors. *Science* 296, 1064–1066.
- Treusch, A. H., Vergin, K. L., Finlay, L. A., Donatz, M. G., Burton, R. M., Carlson, C. A., and Giovannoni, S. J. (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J.* 3, 1148–1163.
- Tripp, H. J., Kitner, J. B., Schwalbach, M. S., Dacey, J. W. H., Wilhelm, L. J., and Giovannoni, S. J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* 452, 741–744.
- Vuilleumier, F., and Simberloff, D. (1980). Ecology versus history as determinants of patchy and insular distribution in high Andean birds. *Evol. Biol.* 12, 325–379.
- Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., and Boon, N. (2009). Initial community evenness favours functionality under selective stress. *Nature* 458, 623–626.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 April 2011; accepted: 14 June 2011; published online: 01 July 2011.

Citation: Eiler A, Hayakawa DH and Rappé MS (2011) Non-random assembly of bacterioplankton communities in the subtropical North Pacific Ocean. *Front. Microbio.* 2:140. doi: 10.3389/fmicb.2011.00140

This article was submitted to *Frontiers in Aquatic Microbiology*, a specialty of *Frontiers in Microbiology*.

Copyright © 2011 Eiler, Hayakawa and Rappé. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

Table A1 | Pair-wise comparison of terminal restriction length polymorphism (T-RFLP) patterns from samples ranging from 5 to 4000 m depth using ANOSIM based on Bray-Curtis similarity.

		Depth (m)															
	5	10	25	45	75	100	125	150	175	200	400	600	800	1000	2000	3000	
5	5																
10	NS	10															
25	NS	NS	25														
45	NS	NS	NS	45													
75	NS	NS	NS	NS	75												
100	NS	NS	NS	NS	NS	100											
125	0.67	0.55	0.63	0.64	0.58	0.21	125										
150	1	0.71	1	0.87	0.89	0.66	NS	150									
175	1	0.84	1	0.86	0.90	0.83	0.44	0.41	175								
200	1	1	1	1	1	0.86	0.37	NS	NS	200							
400	1	0.90	1	0.92	0.97	0.98	0.85	0.85	0.83	0.80	400						
600	1	1	1	1	1	0.98	0.93	0.75	0.79	0.89	NS	600					
800	1	1	1	1	1	0.98	0.92	0.91	0.89	0.72	NS	NS	800				
1000	1	1	1	1	1	0.98	0.97	0.90	0.91	1	0.46	NS	NS	1000			
2000	1	1	1	1	1	0.97	0.97	0.71	0.84	1	0.56	44	NS	NS	2000		
3000	1	1	1	1	1	1	0.96	0.74	0.77	NS	0.67	0.67	0.59	0.61	NS	3000	
4000	0.76	0.87	0.80	0.88	0.94	0.93	0.95	NS	0.81	NS	NS	0.85	0.62	NS	NS	NS	4000

Pair-wise comparisons resulting in insignificant separation are indicated by NS. If $0.05 > p > 0.001$ R-values are given in normal font, if $p < 0.001$ R-values are given in bold. A global test revealed sampling statistics of $R=$ and $p < 0.001$. If $p < 0.05$ denotes a significant difference between bacterial community structures of the compared depths.

Table A2 | Complete list of environmental variables used in the BIOENV procedure.

Environmental subgroup	Variables
Physicochemical	Depth, pH, alkalinity, temperature, salinity
Biogeochemical	Phosphate, nitrite/nitrate, dissolved organic carbon, silicate (sil), particulate phosphorous (pp), particulate nitrogen (pn), particulate silica (psi), particulate carbon (pc), primary production, pc:pn, pc:pp, psi:sil, pc:psi
Biological	Non-pigmented prokaryotes, <i>Prochlorococcus</i> , <i>Synechococcus</i> , chlorophyll <i>a</i> , phaeopigments, chlorophyll <i>c</i> , peridinin, 19'-butanoyloxyfucoxanthin, fucoxanthin, 19'-hexanoyloxyfucoxanthin, prasinoxanthin, diadinoxanthin, zeaxanthin, chlorophyll <i>b</i> , alpha-carotene, beta-carotene, carotenes, chlorophyllide <i>a</i> , violaxanthin, monovinyl chlorophyll <i>a</i> , divinyl chlorophyll <i>a</i> , phycoerythrin (0.4, 5, and 10 μm fractions)

Variables are divided into three categories (subgroups).

Table A3 | Spearman rank correlation between environmental variables and depth.

Variable	ρ_s	P
Temperature	-0.888	5.3E-101
pH	-0.704	1.3E-45
Salinity	-0.606	4.9E-31
Fluorometric chl <i>a</i>	-0.491	2.4E-19
chl <i>a</i>	-0.489	3.4E-19
Zeaxanthin	-0.468	1.5E-17
Particulate N	-0.462	4.5E-17
Bear	-0.447	6.1E-16
DOC	-0.408	2.8E-13
chic	-0.406	3.5E-13
Non-pigmented cells	-0.401	6.9E-13
Particulate C	-0.371	4.4E-11
Particulate P	-0.328	7.8E-09
Diadino	-0.323	1.2E-08
chl <i>b</i>	-0.264	4.0E-06
Phaeopigments	-0.244	2.1E-05
chl <i>a</i>	-0.217	1.6E-04
Primary production	-0.208	3.1E-04
<i>Prochlorococcus</i>	-0.203	4.5E-04
<i>Synechococcus</i>	-0.165	4.5E-03
pc:pn	-0.139	0.017
<i>viol</i>	-0.087	0.136
<i>fuco</i>	-0.056	0.341
<i>pc:psi</i>	-0.054	0.356
<i>psi:sil</i>	-0.029	0.617
<i>perid</i>	0.100	0.085
<i>mvchl <i>a</i></i>	0.107	0.066
<i>hplc</i>	0.110	0.059
<i>pe10</i>	0.119	0.041
dvchl <i>a</i>	0.138	0.018
hex 19	0.144	0.013
pc:pp	0.153	8.5E-03
pn:pp	0.176	2.4E-03
Particulate Si	0.185	1.4E-03
pe4	0.187	1.2E-03
Caroten	0.204	4.0E-04
pe5	0.218	1.6E-04
acar	0.220	1.3E-04
chlplus	0.260	5.7E-06
but19	0.370	4.9E-11
chl <i>b</i>	0.376	2.2E-11
Alkalinity	0.493	1.4E-19
Prasino	0.524	2.6E-22
Silicate	0.627	9.2E-34
Nitrite/nitrate	0.720	1.2E-48
Phosphate	0.745	1.0E-53

Italic style indicates insignificant correlations ($p > 0.05$).