Vol. 79: 165–175, 2017
https://doi.org/10.3354/ame01824

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online May 22

Marine bacterioplankton consortia follow deterministic, non-neutral community assembly rules

Kevin L. Vergin[1], Nicholas Jhirad[1], Jonathan Dodge[2], Craig A. Carlson[3], Stephen J. Giovannoni[1,*]

[1]Department of Microbiology, and [2]Department of Computer Sciences, Oregon State University, Corvallis, Oregon 97331, USA
[3]Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93106, USA

ABSTRACT: The Bermuda Atlantic Time-series Study site provided an opportunity to study bacterial community assembly processes at 2 different depths, the surface and 200 m, in the upper mesopelagic, just below the euphotic zone. Over 100 monthly bacterioplankton DNA samples, from each depth, were analyzed using 16S rRNA gene sequences parsed with the custom software package PhyloAssigner. Co-occurrence networks, filtered for potential autocorrelation artifacts, were constructed for each depth. Network characteristics for the 2 depths were remarkably similar, and network parameters, such as connectance, were in the same range as previously published for ecological networks. Spectral clustering applied to similarity matrices based on exact connections revealed clusters of nodal taxonomic units (NTUs) that peaked at similar times, supporting deterministic, niche-based assembly. An algorithm that used hierarchical Dirichlet processes (HDPs) to model neutral communities based on learned parameters indicated that community assembly processes fit niche-based models at the metacommunity level for both depths. However, HDP analyses restricted to SAR11, SAR86, or SAR202 NTUs supported the neutral assembly hypothesis, suggesting that neutral process models may apply within some phylogenetic domains. To understand whether phylogenetically related taxa can substitute for one another in networks, we created a new metric, phylogenetically weighted connectivity, which considered the similarity of connections among near phylogenetic neighbors. This analysis suggested that phylogenetically similar lineages share similar network connections. Overall, our findings show that niche-based community assembly models are the best fit at both depths but that the neutral model may apply at some phylogenetic scales.

KEY WORDS: Community assembly · Network analysis · Marine bacterioplankton · Bermuda Atlantic Time-series Study

INTRODUCTION

Samples from the Bermuda Atlantic Time-series Study (BATS) site in the Sargasso Sea provide an ideal opportunity to study bacterial community assembly processes. Monthly microbial samples have been collected for over a decade, yielding a rich dataset covering the upper 300 m of the water column (Steinberg et al. 2001, Lomas et al. 2013). The Sargasso Sea is extremely oligotrophic and is characterized by annual deep mixing events in late winter and early spring that extend frequently to 200 m. Summer waters are stably stratified as the surface layers warm. Sampling effort was most complete for surface and 200 m depths, making comparisons of major environmental drivers such as temperature, photic exposure, and annual mixing possible.

High throughput pyrosequencing was used to deeply probe 16S rRNA genes from over 384 samples from a 9 yr time series collected at the BATS site

*Corresponding author: steve.giovannoni@oregonstate.edu

(Vergin et al. 2013b). These sequences were aligned and binned with a custom pipeline, PhyloAssigner, that uses pplacer (Matsen et al. 2010) to assign sequences to nodes of a reference phylogenetic tree (Vergin et al. 2013b). These assignments, which we call nodal taxonomic units (NTUs), accomplish 2 goals: (1) they uniquely identify binned sequences by the reference sequence, and (2) they are embedded in a framework that allows quantification of the phylogenetic relationship between units. A previous study showed that phylogenetically grouped sequences in the SAR11 clade had characteristic ecological distributions and could be defined as ecotypes (Vergin et al. 2013b).

Networking analyses of this BATS dataset could reveal ecological processes. Other researchers have pioneered the use of association and co-occurrence networks to discover potential relationships within bacterial communities (Langfelder & Horvath 2008, Steele et al. 2011, Xia et al. 2011, Faust et al. 2012). This approach has revealed and confirmed many ecological processes such as top-down control of bacterial communities (Chow et al. 2014), nitrogen fixation in corals (Rodriguez-Lanetty et al. 2013), and community succession in a salt marsh (Dini-Andreote et al. 2014). Cluster analyses of microbial networks, as has been applied in yeast gene metabolic networks (Said et al. 2004), may reveal additional insights.

The ultimate goal of this project was to address community assembly processes in the 1991–2004 BATS dataset. Hubbell's unified neutral theory of biodiversity and biogeography (Hubbell 2001) provides a useful model for neutral processes. According to this theory, individuals are equal with respect to demographic rates of birth and death. At the opposite end of the conceptual spectrum is niche assembly theory, in which the community is determined by niches (the position or status of an organism within its community and ecosystem resulting from the organism's structural adaptations, physiological responses, and specific behavior [Odum 1953]) that filter species adapted to current sets of environmental parameters (Keddy 1992). More recent studies suggest that community assembly is not limited to the dichotomy of these 2 models but more likely falls within a continuum between these extremes (Gravel et al. 2006, Wennekes et al. 2012). A recent study developed a machine-learning algorithm to estimate parameters for the neutral model and compare the likelihood of the observed species abundance distributions with modeled communities at both the local and metacommunity level (Harris et al. 2017).

In this study, we used a time series dataset to test the hypothesis of neutral pelagic bacterial community assembly processes at BATS. We compared characteristics from networks constructed for both the surface and 200 m samples and probed for deeper relationships within the networks by applying clustering analyses. We also applied new machine-learning algorithms to the sequence abundance data to directly assess the degree of neutral community assembly at each depth. The comparison of the 2 depths enables us to contrast surface samples subject to the longer-duration annual disturbance with 200 m samples subject to shorter-duration disturbances of the deep mixing event (Lomas et al. 2013), providing a unique opportunity to study factors that drive community assembly processes for bacterioplankton.

MATERIALS AND METHODS

Sample collection, nucleic acid isolation, PCR amplification, pyrosequencing procedures, and pyrosequencing data processing

Details for sample collection and processing were described previously (Vergin et al. 2013a). Briefly, in this study, 454 pyrosequencing data amplified from the V1 and V2 regions of the 16S rRNA gene from 209 monthly samples at the BATS site (approximately 9 yr of samples) were used to construct surface (106 samples) and 200 m networks (103 samples; Table S1 in the Supplement at www.int-res.com/articles/suppl/a079p165_supp.pdf). Additional depth profile samples from the entire dataset (384 total samples) were used for some analyses (Vergin et al. 2013a). For the entire dataset, means of 6684 sequences with a mean of 257 bp length were generated. PhyloAssigner, a custom-designed Perl pipeline (Vergin et al. 2013b) which incorporates the phylogenetic placement algorithm pplacer (Matsen et al. 2010), grouped sequences into well-defined phylogenetic groups, and a comprehensive backbone phylogenetic tree (http://aforge.awi.de/gf/project/phyloassigner) was used to sort the sequence data into NTUs. NTUs are therefore similar to more familiar operational taxonomic units (OTUs), except that NTUs are defined phylogenetically by reference sequences and clade structure rather than by cutoff thresholds. Although OTU binning using cutoff thresholds is computationally expedient, it has been demonstrated that phylogenetic placement using this method results in a loss of fine-scale phylogenetic information that can be important for resolving ecotypes (Koeppel & Wu 2013, Vergin et al.

2013b). NTUs were classified by phylum (or class in the case of *Proteobacteria*) and clade (SAR11, SAR86, SAR116, SAR202, SAR276, SAR324, SAR406, plastids, and *Roseobacter*) based on their position in the reference phylogenetic tree.

Calculating networks from Spearman and Pearson correlations

Amplicon sequence abundance was normalized for each sample. These percentages were then applied to the total cell count for each sample to obtain a relative abundance less biased by differences in cell counts. Because of other well-known biases, these abundances are not treated as proxies for cell counts but are still relative abundances. Samples with missing cell counts were estimated from the means of all samples from the same depth and month (relative to the month of deepest mixing). Time series of relative abundance for pairs of NTUs were compared using Spearman and Pearson correlations, as implemented in local similarity analysis (LSA) (Ruan et al. 2006), with the significance level adjusted for multiple testing using the Benjamini-Hochberg correction (all algorithms were written for R [R Core Team 2013] and are available at https://github.com/kevinvergin). Both positive and negative correlations were detected, but time lags were not considered due to the high number of gaps in the time series data.

Network pruning using diagnostic filtering to remove potentially spurious correlations

Potentially significant correlations were filtered in 3 steps. First, linear regression analysis was used to test for a non-zero slope of the linear model comparing the relative abundances of the 2 NTUs (*F*-statistic p-value < 0.05). Second, since seasonality was a strong component in the data, samples were seasonally differenced, linear regressions were re-modeled, and residuals were examined using autocorrelation functions to detect non-normal distributions. Since seasonal differencing is sensitive to gaps in the time series, the >100 samples were trimmed to 6 complete years, and partial years were removed (Table S1 in the Supplement). Data for missing months (surface: September 1993, August 1998, and October–November 1998; 200 m: December 1992–January 1993, September 1993, October–November 1998, July 2000, and July 2002) were extrapolated by averaging through the 1 mo prior and 1 mo after the missing sample(s). Simulations demonstrate that non-normality in the residuals of only 1 NTU does not affect the type I error rate (see Supplementary methods section in the Supplement), so the full time series was used for subsequent analyses. Finally, assumptions about the residuals in the linear regression analysis (independence and normal distribution) were examined using the autocorrelation function to detect non-random, sinusoidal patterns indicative of seasonality. Initially, the residuals for potential linear correlations were tested and scored for non-normality. A diagnostic test for correlations strongly influenced by seasonality was not readily apparent, so we conservatively eliminated correlations with high potential for a spurious relationship by eliminating correlations between NTUs with strong seasonal signals. Ninety-five percent of the NTUs had no or a few non-random autocorrelation function results, so correlations between NTUs in the top 5% of total non-random autocorrelation functions were eliminated to yield the final screened matrix of correlations, hereafter referred to as connections.

Calculating phylogenetically weighted connectivity

A new similarity coefficient, called phylogenetically weighted connectivity (PWC, ∂), was developed to calculate similarities for all pairwise comparisons of NTUs (e.g. NTU A and NTU B) based on their shared connections to other NTUs. Since PhyloAssigner provides evolutionary distances between NTUs, a weighting system was developed that considered near phylogenetic neighbors in the calculation of this similarity metric. For example, in the comparison of NTU A to NTU B, sliding windows considered the 5 nearest phylogenetic neighbors of all NTUs connected to A and B, in both directions within the tree. To facilitate this, profiles consisting of all connected NTUs plus or minus 5 nodes from the NTU being profiled were constructed for all NTUs. Each connected node in the profile was then scored as plus or minus 1, depending on the sign of the correlation, and then neighboring NTUs were weighted inversely to phylogenetic distance from the connected NTU. Thus, we created a metric of the similarity between the connections of any 2 nodes, taking into account the network relationships of nodes closely related to the nodes being compared. This weighting is different from the strength-of-correlation weighting used by others (Langfelder & Horvath 2008). Primer 6.0 software (Clarke & Gorley 2006) was then used to construct a non-metric multidimen-

sional scaling plot and to conduct permutational multivariate analysis of variance (PERMANOVA) after categorizing by phylogeny.

Spectral clustering

A kernel *K*-means approach was implemented in the R package specc to define clusters of NTUs in a similarity matrix. Similarities between NTUs based on direct connections to other NTUs were again calculated, except that phylogenetic relatedness was not considered, since a more direct similarity measure was desired. The number of clusters, *K*, was estimated using the pamk function in the R package fpc and the default asw criterion. A *K*-value of 7 was used for both the surface and 200 m networks. Three different kernel functions (linear, Gaussian, and polynomial) were applied to the matrix to identify clusters. Analyses for each kernel were repeated 100 times with 90% of the data points for a bootstrap-like confirmation of cluster assignment. Combined results for all kernels and replicates were used to generate a new similarity matrix where NTU assignments on a pairwise basis were calculated as the percentage of instances where both NTUs were present in a given replicate analysis and assigned to the same cluster. Final, new clusters, consisting of lists of NTUs assigned to the same bootstrap analysis cluster a majority of times, were then generated.

Unified neutral theory of biodiversity ecological parameters

The unified neutral theory of biodiversity (Hubbell 2001) uses 3 parameters to characterize sample population distributions: (1) theta, a dimensionless fundamental biodiversity constant that measures speciation–drift in the metacommunity; (2) *I*, a fundamental immigration number that measures migration–drift in the local community; and (3) *m*, the migration rate. These parameters were estimated using the untb package in R (Hankin 2007). Cell counts at BATS vary within an order of magnitude, but there is a systematic bias towards greater numbers at the surface. Therefore, theta was not assumed to be uniform and, consequently, the G_{ST} method in the untb package was used to estimate *m* and *I* values. Relationships between these estimates, as well as temperature and cell counts, for each sample were graphed in contour plots using Ocean Data View software (Schlitzer 2014).

Individual samples (local communities) from both the surface and 200 m and separately combined surface and 200 m samples (metacommunity) were assessed for fit to a neutral model using published software (Harris et al. 2017). The parameters for the neutral assembly model were estimated from the data by modeling the data as a hierarchical Dirichlet process (HDP, Harris et al. 2017). A Gibbs sampler, a type of Bayesian Markov chain Monte Carlo algorithm, then generated samples from the posterior distribution of the parameters given the data. Thus, the likelihood of the data was compared against a distribution of likelihoods from an artificial data matrix generated using the learned parameters. Data were trimmed to exclude NTUs with zero abundance at the depth under consideration (1031 and 960 NTUs for the surface and 200 m, respectively). Surface and 200 m datasets parsed to include only the SAR11 clade (surface), SAR202 clade (200 m), and SAR86 clade (both surface and 200 m) were also analyzed.

RESULTS

Correlation networks were generated for 2 depths, the surface and 200 m, from monthly samples collected at the BATS site over a discontinuous 9 yr time span. An algorithm based on LSA (Ruan et al. 2006) detected 45 980 and 43 153 putative significant correlations for the surface and 200 m networks, respectively (Table 1). Three diagnostic algorithms were subsequently applied to the networks to reduce the likelihood of type I errors (reject the null hypothesis of no significant correlation when the null hypothesis is true). The first diagnostic algorithm tested for a non-zero slope in the relationship between the relative abundances of each pair of NTUs. The second

Table 1. Number of significant correlations at each step of diagnostic testing and final network characteristics. ACF: autocorrelation function

	Surface	200 m
Network correlations		
Spearman and Pearson	45 980	43 153
Linear regression slope	40 435	39 762
Seasonal differencing check	35 749	38 150
Dual ACF check	35 571	38 150
Network characteristics		
Connectance	0.029	0.031
Average path length	2.55	2.55
Average edge betweenness	47.2	43.2
Clustering coefficient	0.51	0.50

diagnostic algorithm applied seasonal differencing to the time series and eliminated correlations with underlying non-normality in the linear regression residuals. The final diagnostic algorithm eliminated potentially spurious correlations due to underlying non-normality of linear regression residuals in non-differenced data. After these diagnostics were used to remove correlations that were identified as potential type I errors, the networks had 35 571 and 38 150 correlations for the surface and 200 m networks, respectively. Correlations that passed these diagnostic tests are referred to as connections throughout.

Network characteristics were remarkably similar for the 2 networks and comparable to many other reported microbial and food web networks (Dunne et al. 2002, Steele et al. 2011, Barberán et al. 2012; Table 1). The small connectance values and short average path lengths indicate small-world characteristics for these networks. A cumulative distribution of node connectivity had a complex shape but could be approximated by a truncated power law function. As was the case for the other characteristics, coefficients for the power law function were similar for the 2 networks (a = 0.087 and 0.063, b = 0.026 and 0.026 for the surface and 200 m, respectively; Fig. S1 in the Supplement at www.int-res.com/articles/suppl/a079p165_supp.pdf).

The NTUs with the greatest number of connections for each network were identified (Table S2 in the Supplement). Interestingly, nearly all of these NTUs, for both networks, are in the class *Gammaproteobacteria*, many of which belong to the SAR86 clade, a ubiquitous and diverse phylogenetic group of uncultivated marine bacteria. These NTUs are commonly present throughout the time series and were classified as Frequent Abundant in a previous analysis (Vergin et al. 2013a; Fig. S2 in the Supplement).

PWC, described in 'Materials and methods', revealed that phylogenetically similar NTUs tend to have similar connections to other NTUs. PWC matrices were constructed for the networks, NTUs were sorted into phylogenetic categories (phylum/class or family), and PERMANOVAs were used to test the hypothesis that variance within a phylogenetic category was not significantly different from a random expectation. Within the PWC matrix, pairwise comparisons of phylogenetic categories showed that NTUs within a phylogenetic category were grouped more closely together than randomly assigned NTUs in most cases (Table S3 in the Supplement).

It was reasoned that 2 NTUs from a co-occurring community should have more similar connections to other NTUs in the same community than 2 NTUs from different communities. Thus, these co-occurring communities should be revealed as clusters in a similarity matrix of direct NTU connections. As opposed to the PWC similarity measure, phylogenetic relationships are not used as weights when determining similarities of connections in this analysis. Spectral cluster assignments were determined using 3 different kernel functions and 100 bootstrap-like replicates for each of the functions, resulting in 35 and 50 clusters for the surface and 200 m networks, respectively. Distributions of the NTUs in each cluster revealed that members of a cluster tended to have peak relative sequence abundances at the same time(s) during the time series (Fig. 1).

Neutral community parameters were estimated from relative abundance data for the entire dataset (384 samples) using the untb package in R (Hankin 2007). In addition, surface and 200 m depth strata were each tested for neutral community assembly processes using a published algorithm (Harris et al. 2017). For the entire dataset, the migration rate parameter, m, increased in surface waters during winter months corresponding to the deep mixing period and was much lower in the summer months when the water column was stratified (Fig. 2A). At 200 m, m varied less throughout the year. The fundamental dispersal number, I, followed a similar pattern (Fig. 2B), with higher values at the surface during deep mixing and lower values in surface stratified waters. There was strong evidence for non-neutral assembly processes for both surface and 200 m metacommunities (pseudo-p = 0, test to reject neutral assembly) and marginal evidence for non-neutral assembly processes for surface local samples (pseudo-p = 0.038). However, 200 m local samples were consistent with a model of neutral assembly processes (pseudo-p = 0.996). When subsets of the data consisting of a single phylogenetic clade at a specific depth were considered (surface SAR11, 200 m SAR202, surface SAR86, and 200 m SAR86), neutral assembly processes were supported at both the metacommunity and local levels in all cases (pseudo-p > 0.70).

DISCUSSION

We chose to modify an existing algorithm, LSA (Ruan et al. 2006), to study time series data from BATS. The modified approach we took retained the Spearman and Pearson correlations from LSA but filtered out potentially spurious correlations by linear modeling and examination of residuals from linear regressions for non-normal behavior. Since time

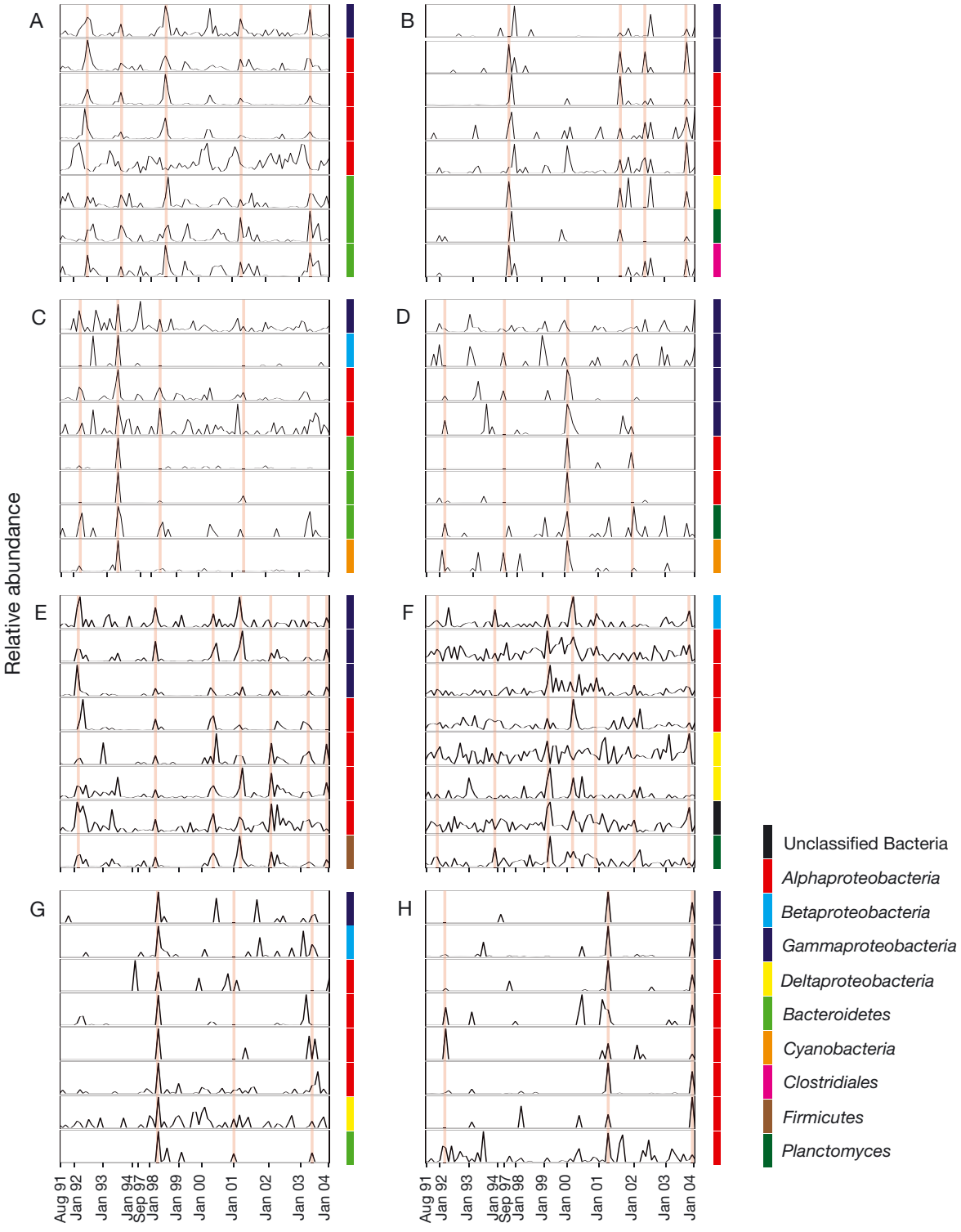

Fig. 1. Relative abundance plots for 8 representative nodal taxonomic units (NTUs) in each of 4 spectral clusters for the (A–D) surface network and (E–H) 200 m network. Relative abundance of amplicon sequences (*y*-axis) for individual NTUs over the entire concatenated time series (*x*-axis) is shown. The first months for each calendar year are marked on the *x*-axis. Colored bars to the right of the plot indicate the phylum, or class in the case of *Proteobacteria*, of the NTU as detailed in the legend. NTU numbers and expanded phylogenetic designations are available in Table S4 in the Supplement at www.int-res.com/articles/suppl/a079p165_supp.pdf. Faded red lines for each cluster indicate periods during which a majority of NTUs within the cluster had relatively greater abundances, suggesting a coordinated response to environmental conditions

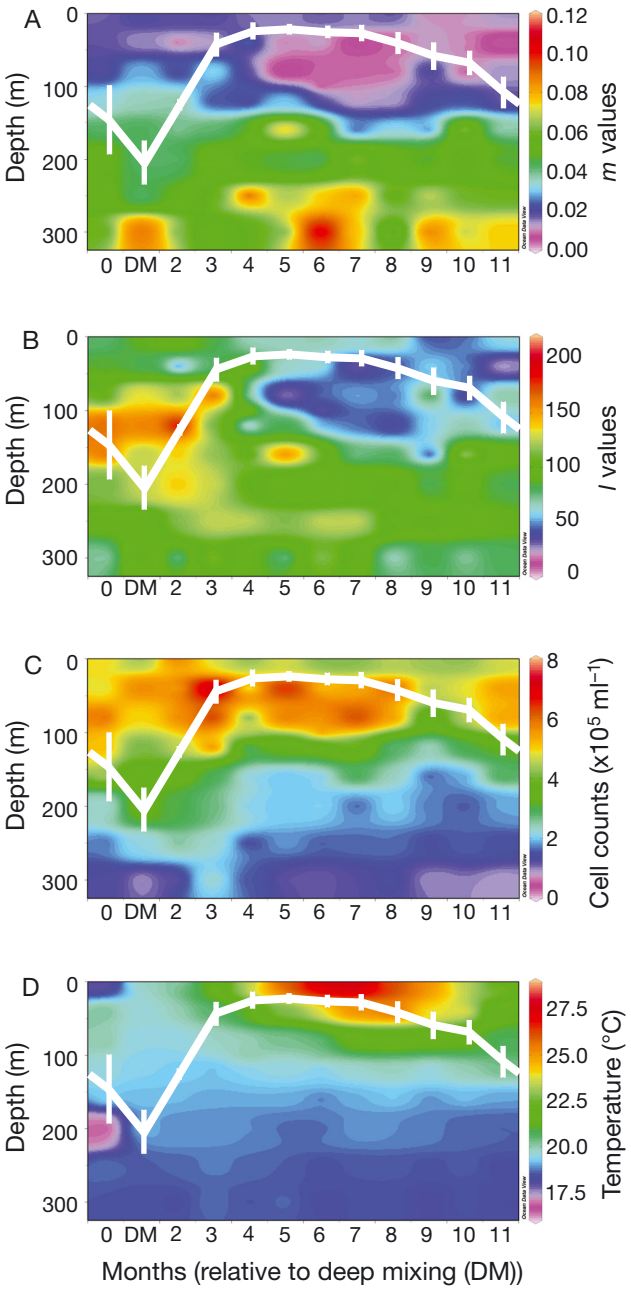

Fig. 2. Contour plots for (A) migration rate, *m*; (B) fundamental dispersal number, *I*; (C) cell count; and (D) temperature for all time series samples plotted as an average over a 1 yr time span. The plots are indexed to the month of deepest mixing (DM) for each year, designated as month 1 (*x*-axis). Sample depth is indicated on the *y*-axis. The white line represents the average mixed layer depth, and bars indicate standard deviations. Heat map scales are indicated to the right of the plot

series data are not independent (abundance tends to be related to the previous and subsequent time points), we examined the autocorrelation functions of the residuals to detect non-normal patterns. As a result of the diagnostic steps, 17.3% of all correlations were eliminated. Non-normality in the residuals of linear regression analyses is most likely due to seasonality in the time series data, so it is interesting to note that fewer correlations were eliminated from the 200 m network (1612) than the surface network (4864), where seasonality drives community turnover (Gilbert et al. 2012, Giovannoni & Vergin 2012). The networks that emerged had small-world characteristics very similar to those observed previously by other investigators using different construction methods (Dunne et al. 2004, Steele et al. 2011).

Phylogenetic community ecology attempts to model evolutionary processes resulting from the selection pressures of ecological processes. This idea has been applied to other networks and has demonstrated that phylogenetically similar species interact with the same set of species and/or occupy similar positions in the network (Mouquet et al. 2012 and references therein). We sought to apply similar analyses to this bacterial network, so we implemented a new metric, PWC, to examine the relationship between phylogenetic relatedness, which was readily available from PhyloAssigner, and the connections of NTUs. PWC measures the similarity of 2 taxa in terms of shared connections within the network, but it takes into account not just the connected taxa being compared but also their nearest phylogenetic neighbors. By comparing the PWC values for NTU pairs, we reasoned that we could find sets of NTUs with similar roles in the community. The PWC analysis supported the role of ecological drift, a neutral process, in community assembly by showing that phylogenetically related taxa tend to form similar connections. The similarity of connections increased (PWC values were higher) with decreasing phylogenetic distance. We interpreted this evidence as support for the ecological drift hypothesis because it shows that related taxa tend to have similar network relationships, implying that they are interchangeable in communities (Table S3 in the Supplement). Interestingly, some of the most highly connected taxa were closely related phylogenetically, suggesting the possibility of drift among keystone taxa (Steele et al. 2011, Eiler et al. 2012).

A subsequent clustering analysis provided evidence for deterministic, niche-based community assembly. A second similarity matrix based on NTU connection similarities was generated but this time considered exact matches only and no phylogenetic weighting, thus creating a more finely resolved analysis of relationships between taxa. We hypothesized that connection similarities would be higher among

members of a co-occurring community and that these higher similarities would be revealed as clusters in a similarity matrix. Therefore, spectral clustering utilizing 3 different kernel functions was applied to the similarity matrix to define clusters. For increased confidence, a bootstrap-like analysis was repeated 100 times, and NTUs were assigned to clusters based on their shared cluster assignments with other NTUs. The clusters (35 for the surface network and 50 for the 200 m network) tended to share NTUs that peaked in relative sequence abundances at the same times, meaning that NTUs tended to co-occur, sometimes at multiple time points in the time series. This result suggests that NTUs were assembled into a community deterministically when they favorably responded, presumably to the same set of biotic and/or abiotic conditions present at a given time. Similar to previous studies of this dataset (Vergin et al. 2013a,b), no significant correlations with physical or chemical factors routinely measured at BATS were found, suggesting that interactions with environmental factors are subtle and may be missed by current sampling methods.

We modeled community assembly processes at BATS using methods developed to test Hubbell's unified neutral theory of biodiversity (Hubbell 2001), which seeks to understand whether neutral or niche-based processes dominate assembly. The PWC analyses provided support for the ecological drift model, a prediction from Hubbell's neutral theory, by showing that phylogenetically related taxa share similar connections and can presumably substitute for each other in ecological networks. Hubbell theorized that local communities form stochastically from a metacommunity. In his model, population size remains constant, as all resources are used by the individuals in the local community. As individuals die, there is an equal chance that any individual will reproduce or a variable chance that a member of the metacommunity will migrate to replace the expired individual, thus maintaining a constant population. The bacterioplankton community at BATS, in principle, can provide a good test of the neutral model because bacterial populations fluctuate over a small range, especially within depth strata, thus meeting the model's assumption of constant population size (Fig. 2C). In this analysis, the entire time series at each depth is interpreted as a metacommunity with temporal and spatial dimensions, while the individual sample points are interpreted as the local communities. Comparing the data to neutral models, we found no support for the neutral assembly hypothesis at the metacommunity level, for either the surface or 200 m datasets. But for local communities, at the surface, neutral assembly was rejected at the $p = 0.01$ level but not at the $p = 0.05$ level, indicating marginal support for non-neutral assembly. However, the evidence was strong for neutral assembly processes in local communities at 200 m.

The difference in local but not metacommunity assembly processes between the surface and 200 m can be explained by the annual deep mixing event in the BATS system. The migration rate parameter, m (Fig. 2A), and fundamental dispersal number, I (Fig. 2B), for the full dataset, including depth profile samples (Vergin et al. 2013b), show a change in dispersal during the deep mixing period in the late winter/early spring. Barriers to dispersal are probably reduced during deep mixing due to evenly distributed water temperatures (Fig. 2D) and physical mixing from storm systems. By summer, water temperatures at the surface increase, resulting in a stable, stratified water column that likely increases barriers to dispersal, resulting in a lower-diversity community at the surface. This observation is consistent with an analysis we reported previously that suggested that transport resulting from disturbances created opportunities for growth and immigration of rare taxa (Vergin et al. 2013a). The seasonal fluctuation between cold, nutrient-rich, mixed waters and warm, nutrient-depleted, stratified waters likely drives habitat filtering that limits the dispersal of individuals from the surface metacommunity. An analogous example of dispersal limitation was described in an unconfined aquifer system (Stegen et al. 2013). BATS deep mixing sometimes extends to 200 m, where there is a demonstrable, but weak, seasonality (Morris et al. 2005, Vergin et al. 2013a). Thus, the fit of the 200 m data to a local community neutral model is likely related to weaker physical forces driving changes in community structure (Hellweger et al. 2014).

Neutral community assembly was supported by modeling within broad phylogenetic catagories, such as the SAR11, SAR86, and SAR202 clades. When the surface and 200 m data were filtered to include clades of NTUs, there was strong evidence for neutral community assembly, despite the influence of deep mixing. This result is similar to human gut microbiome community assembly for intra-taxon groups (Harris et al. 2017) and ammonia-oxidizing bacteria in wastewater treatment microbial communities (Ofiteru et al. 2010). The results presented here are consistent with other studies of natural systems (Manrique & Jones 2017) and suggest that models supporting ecological drift may be a common feature for lower taxonomic groups.

One striking difference between these sites is the extent of the annual deep mixing event, which impacts the surface much more strongly than the deeper 200 m layer. This mixing, along with differing light and temperature effects, creates a larger gradient of abiotic conditions, thus filtering different microbial groups at different times of the year. The enlargement of niche space by seasonality and mixing likely drives the modeling of the system away from the null hypothesis of neutral assembly. The upper mesopelagic (200 m) is a less variable environment than the surface, both in chemical and physical conditions and in the phylogenetic composition of the microbial community (Morris et al. 2005), but nonetheless the impact of environmental filtering was evident in the fit of the data to niche-based models. Spectral clustering supported the results from HDP models that suggest that niche-based processes dominate community turnover at both the surface and 200 m. However, we observed that in the more stable 200 m environment, the data fit a neutral model at the local community level, while at the metacommunity level, deterministic, niche-based processes were a better fit. The appearance of neutrality may be influenced by the frequency of events, such that community members that respond to conditions that are duplicated infrequently may appear to follow neutral models in shorter time series. A study of a wastewater treatment system showed strong niche-based assembly processes indicated by synchronized community changes in 3 replicate anaerobic digestors (Vanwonterghem et al. 2014). Eukaryotic microbial communities in managed vineyards also showed strong evidence for niche-based assembly processes (Morrison-Whittle 2015). These results suggest that niche-based assembly patterns may emerge in duplicated environments or in time series experiments that are long enough to display repeated conditions.

CONCLUSIONS

The long-term BATS data and the application of PhyloAssigner provided us with opportunities to explore microbial community networks from a new perspective. Overall, microbial community networks at BATS have small-world characteristics that are surprisingly like those that have been described in other microbial plankton systems. We found that both at the surface and at 200 m, these communities appeared to be mainly driven by niche-based assembly processes.

Hubbell's unified neutral theory of biodiversity (Hubbell 2001) provides a context for evaluating the relative contributions of habitat filtering and neutral processes to community assembly. Using tools that were developed to assess the fit of data to the neutral theory, we found that metacommunities at both depths follow niche assembly processes, while upper mesopelagic local communities appear to follow neutral assembly processes. A major difference between the surface and mesopelagic is the impact and duration of mixing — a physical process that transports cells and brings nutrients to the surface. The alternating transition from a disturbed, mixed state to a stable, stratified state may create divergent niches that filter potential migrants. Even in the mesopelagic, where mixing is less pronounced, disrupted conditions may allow habitat filtering to influence metacommunity composition but not enough to disrupt fit to a neutral assembly model at local community levels. The analyses also suggest that within some clades, such as SAR11, the neutral model may apply.

We used phylogenetic distances from PhyloAssigner to show that within the networks, phylogenetically related taxa share similar connections. This implies that, at least in some domains of the phylogenetic tree that we used to parse the data, there is functional redundancy from a network perspective in the data we analyzed. PhyloAssigner, which does not parse sequences with uniform similarity cutoffs, assesses 16S rRNA diversity at relatively fine scales. Exploring the exact phylogenetic scales at which network characteristics transition from fitting niche-based assembly may be an avenue for understanding the relationships between niche-driven and neutral processes in determining microbial diversity. The potential importance of neutral processes in community assembly was illustrated by the recent demonstration (Hellweger et al. 2014) that bacterioplankton biogeography can be impacted by neutral drift.

In summary, we found that the 2 pelagic microbial environments, the surface and upper mesopelagic, harbor networks that are very similar in many network characteristics and in both niche-based community assembly processes dominated over neutral processes. However, we also found evidence for neutral processes operating within some large microbial plankton clades and at the local level in the upper mesopelagic. We attribute the difference between the surface and upper mesopelagic to the effects of seasonal forcing, which impose greater environmental influence at the surface.

Acknowledgements. We thank the officers and crew of the RV 'Weatherbird II' for their assistance and support and the BATS chief scientists and technicians for assistance in water collection and accommodating wire time requests. We thank F. Munoz, T. Sharpton, J. Raes, and K. Faust for helpful discussions. Major support was provided by grants from the Marine Microbiology Initiative of the Gordon and Betty Moore Foundation and the NSF Microbial Observatory program (MCB-0237713 and OCE-0802004). This research was supported, in part, by the Simons Foundation International's BIOS-SCOPE program.

LITERATURE CITED

Barberán A, Bates ST, Casamayor EO, Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6:343–351

Chow CE, Kim DY, Sachdeva R, Caron DA, Fuhrman JA (2014) Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J 8:816–829

Clarke KR, Gorley RN (2006) PRIMER v6: user manual/tutorial. PRIMER-E, Plymouth

Dini-Andreote F, de Cassia Pereira ESM, Triado-Margarit X, Casamayor EO, van Elsas JD, Salles JF (2014) Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. ISME J 8:1989–2001

Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecol Lett 5:558–567

Dunne JA, Williams RJ, Martinez ND (2004) Network structure and robustness of marine food webs. Mar Ecol Prog Ser 273:291–302

Eiler A, Heinrich F, Bertilsson S (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J 6:330–342

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C (2012) Microbial co-occurrence relationships in the human microbiome. PLOS Comput Biol 8:e1002606

Gilbert JA, Steele JA, Caporaso JG, Steinbruck L and others (2012) Defining seasonal marine microbial community dynamics. ISME J 6:298–308

Giovannoni SJ, Vergin KL (2012) Seasonality in ocean microbial communities. Science 335:671–676

Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. Ecol Lett 9:399–409

Hankin RKS (2007) Introducing untb, an R package for simulating ecological drift under the unified neutral theory of biodiversity. J Stat Softw 22:1–15

Harris K, Parsons TL, Ijaz UZ, Lahti L, Holmes I, Quince C (2017) Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. Proc IEEE 105(3):516–529

Hellweger FL, van Sebille E, Fredrick ND (2014) Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. Science 345:1346–1349

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ

Keddy PA (1992) Assembly and response rules: two goals for predictive community ecology. J Veg Sci 3:157–164

Koeppel AF, Wu M (2013) Surprisingly extensive mixed phylogenetic and ecological signals among bacterial operational taxonomic units. Nucleic Acids Res 41: 5175–5188

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559–571

Lomas MW, Bates NR, Johnson RJ, Knap AH, Steinberg DK, Carlson CA (2013) Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea. Deep-Sea Res II 93:16–32

Manrique JM, Jones LR (2017) Are ocean currents too slow to counteract SAR11 evolution? A next-generation sequencing, phylogeographic analysis. Mol Phylogenet Evol 107:324–337

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538

Morris RM, Vergin KL, Cho JC, Rappe MS, Carlson CA, Giovannoni SJ (2005) Temporal and spatial response of bacterial lineages to annual convective overturn at the Bermuda Atlantic time-series study site. Limnol Oceanogr 50:1687–1696

Morrison-Whittle P, Goddard MR (2015) Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. ISME J 9:2003–2011

Mouquet N, Devictor V, Meynard CN, Munoz F and others (2012) Ecophylogenetics: advances and perspectives. Biol Rev Camb Philos Soc 87:769–785

Odum EP (1953) Fundamentals of ecology. Saunders, Philadelphia, PA

Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, Sloan WT (2010) Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci USA 107:15345–15350

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rodriguez-Lanetty M, Granados-Cifuentes C, Barberan A, Bellantuono AJ, Bastidas C (2013) Ecological inferences from a deep screening of the complex bacterial consortia associated with the coral, *Porites astreoides*. Mol Ecol 22: 4349–4362

Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F (2006) Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532–2538

Said MR, Begley TJ, Oppenheim AV, Lauffenburger DA, Samson LD (2004) Global network analysis of phenotypic effects: protein networks and toxicity modulation in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 101: 18006–18011

Schlitzer R (2014) Ocean Data View. https://odv.awi.de

Steele JA, Countway PD, Xia L, Vigil PD and others (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J 5:1414–1425

Stegen JC, Lin X, Fredrickson JK, Chen X and others (2013) Quantifying community assembly processes and identifying features that impose them. ISME J 7:2069–2079

Steinberg DK, Carlson CA, Bates NR, Johnson RJ, Michaels AF, Knap AH (2001) Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. Deep-Sea Res II 48:1405–1447

Vanwonterghem I, Jensen PD, Ho DP, Batstone DJ, Tyson GW (2014) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. Curr Opin Biotechnol 27:55–64

Vergin KL, Done B, Carlson CA, Giovannoni SJ (2013a) Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. Aquat Microb Ecol 71:1–13

Vergin KL, Beszteri B, Monier A, Cameron Thrash J and others (2013b) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. ISME J 7: 1322–1332

Wennekes PL, Rosindell J, Etienne RS (2012) The neutral-niche debate: a philosophical perspective. Acta Biotheor 60:257–271

Xia LC, Steele JA, Cram JA, Cardon ZG and others (2011) Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. BMC Syst Biol 5:S15

Editorial responsibility: Jennifer Martiny, Irvine, California, USA

Submitted: November 1, 2016; Accepted: March 13, 2017
Proofs received from author(s): May 15, 2017