

ORIGINAL ARTICLE

Seasonality and vertical structure of microbial communities in an ocean gyre

Alexander H Treusch¹, Kevin L Vergin¹, Liam A Finlay¹, Michael G Donatz², Robert M Burton², Craig A Carlson³ and Stephen J Giovannoni¹

¹Department of Microbiology, Oregon State University, Corvallis, OR, USA; ²Department of Mathematics, Oregon State University, Corvallis, OR, USA and ³Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, USA

Vertical, seasonal and geographical patterns in ocean microbial communities have been observed in many studies, but the resolution of community dynamics has been limited by the scope of data sets, which are seldom up to the task of illuminating the highly structured and rhythmic patterns of change found in ocean ecosystems. We studied vertical and temporal patterns in the microbial community composition in a set of 412 samples collected from the upper 300 m of the water column in the northwestern Sargasso Sea, on cruises between 1991 and 2004. The region sampled spans the extent of deep winter mixing and the transition between the euphotic and the upper mesopelagic zones, where most carbon fixation and reoxidation occurs. A bioinformatic pipeline was developed to de-noise, normalize and align terminal restriction fragment length polymorphism (T-RFLP) data from three restriction enzymes and link T-RFLP peaks to microbial clades. Non-metric multi-dimensional scaling statistics resolved three microbial communities with distinctive composition during seasonal stratification: a surface community in the region of lowest nutrients, a deep chlorophyll maximum community and an upper mesopelagic community. A fourth microbial community was associated with annual spring blooms of eukaryotic phytoplankton that occur in the northwestern Sargasso Sea as a consequence of winter convective mixing that entrains nutrients to the surface. Many bacterial clades bloomed in seasonal patterns that shifted with the progression of stratification. These richly detailed patterns of community change suggest that highly specialized adaptations and interactions govern the success of microbial populations in the oligotrophic ocean.

The ISME Journal advance online publication, 4 June 2009; doi:10.1038/ismej.2009.60

Subject Category: microbial population and community ecology

Keywords: marine bacteria; community dynamics; ocean stratification

Introduction

The global impact of ocean biology is heavily determined by its most populous members, the microbial plankton. Study of these organisms has advanced rapidly, but there are still large gaps in basic ecological information about these organisms. A particularly fundamental issue is their seasonal, vertical and biogeographical distributions. Advances in sequencing technology have made it possible to gather very large amounts of sequence data from a few samples, but the effectiveness of these data for resolving ecological processes depends on the coverage of molecular diversity within each sample and the mesh of the sampling schemes, which inevitably are compromises that

subsample and underdetermine detailed patterns of seasonality or vertical variation.

The vertical stratification of microbial communities in the oceans was first apparent in the obvious case of phytoplankton (Olson *et al.*, 1990a, b; Moore *et al.*, 1998), but later studies that used molecular probes as genetic markers for uncultivated microbial taxa uncovered many more examples. Three new bacterial clades were reported from the aphotic zone—the SAR406/Marine Group A lineage, which is related to the *Fibrobacteres* and *Chlorobi* phyla, the SAR202 clades, which are in the *Chloroflexi* phylum, and SAR324, in the *Deltaproteobacteria* (Giovannoni *et al.*, 1996; Gordon and Giovannoni, 1996; Fuhrman and Davis, 1997; Wright *et al.*, 1997). All three were observed to peak in abundance in the upper mesopelagic zone, below the deep chlorophyll maximum (DCM) in both the Pacific and Atlantic Oceans. Later studies corroborated these reports, and SAR202 emerged as one of the most abundant aphotic zone bacterioplankton groups, accounting for about 10% of the cells in the integrated water column community (Morris *et al.*,

Correspondence: SJ Giovannoni, Department of Microbiology, Oregon State University, Nash Hall, Corvallis, OR 97331, USA.

E-mail: steve.giovannoni@oregonstate.edu

Received 22 January 2009; revised 21 April 2009; accepted 21 April 2009

2004; Varela *et al.*, 2008). Similarly, probe studies showed that marine *Crenarchaeota* are largely confined to the aphotic zone, where they are the single most abundant class of cells (Fuhrman and Davis, 1997). These findings were confirmed by many subsequent studies (for example, DeLong *et al.*, 1999; Karner *et al.*, 2001). Recently, a metagenomic study at the Hawaii Ocean Time-series confirmed some of the earlier findings (DeLong *et al.*, 2006), but also reported anomalous observations—almost no SAR11 16S rRNA genes, and evidence of phototrophs well below the photic zone.

Seasonal variation in ocean microbial surface communities has been reported in several studies, although the logistical problems of periodically collecting samples makes ‘time-series’ studies less common. Time-series have proven to be powerful tools in oceanography (for a review see Ducklow *et al.*, 2009) and have recently been applied to marine microbial ecology. Murray *et al.* used DNA hybridization methods to show seasonal changes of *Archaea* in the Antarctic Ocean (Murray *et al.*, 1998) and the Santa Barbara Channel (Murray *et al.*, 1999). DuRand *et al.* (2001) used flow cytometry to measure seasonal and spatial patterns of *Prochlorococcus* and *Synechococcus* at the Bermuda Atlantic Time-series Study (BATS) site, and Heywood *et al.* (2006) showed that the standing stocks of both groups undergo little interannual change. Morris *et al.* (2002) showed annual peaks in the relative abundance of SAR11 by rRNA probe hybridization. A combination of fluorescence *in situ* hybridization probing and terminal restriction fragment length polymorphism (T-RFLP) time-series data showed regular seasonal patterns of three ecotypes of SAR11 in the surface 300 m at BATS (Carlson *et al.*, 2009). To correlate marine archaeal abundance with nitrification, Wuchter *et al.* (2006) analyzed a one-year time-series from coastal North Sea water samples. In a more global approach, Moeseneder *et al.* (1999) used spring and summer samples to study the bacterial community in the Aegean and Adriatic seas by T-RFLP. Bano and Hollibaugh (2002) analyzed the bacterioplankton assemblages in the Arctic Ocean over three seasons using denaturing gradient gel electrophoresis and clone libraries. Studying the microbial community at BATS over several years using T-RFLP, Morris *et al.* (2005) demonstrated a response by specific bacterial lineages to the annual spring deep mixing events within the mesopelagic zone at BATS. In a long-term study of the microbial community at the San Pedro Ocean Time-Series site, Fuhrman *et al.* (2006) used automated rRNA intergenic spacer analysis and were able to link several of the operational taxonomic units to environmental factors such as nutrients. Recently, Nelson *et al.* (2008) reported a coastal two-year time-series from a 34-km transect in the Mid-Atlantic Bight using T-RFLP. They showed distinct temporal microbial community patterns,

mainly the separation of the terminal restriction fragments observed into a summer and a winter group. Although many of these time-series were short-termed, those by Morris *et al.* (2005), Fuhrman *et al.* (2006) and Carlson *et al.* (2009) relied on samples collected over several years.

Here, we report a study that focuses on the 16S rRNA gene diversity in the surface layer (0–300 m) at BATS with vertical and relatively high temporal resolution, with the objective of defining ecological patterns. About one-third of the CO₂ that is photosynthetically reduced on Earth is fixed in the oceans by photoautotrophs, but most of it is rapidly reoxidized in the marine surface layer by heterotrophic microorganisms (Field *et al.*, 1998). Dynamic interactions between the organisms detected in this study lie at the heart of the carbon cycle, although progress understanding these interactions has been slow to emerge because of the complexity of microbial community ecology.

The BATS program initiated in 1988, provides a rich oceanographic and ecological context, including reports of seasonal patterns of biogeochemistry (Michaels *et al.*, 1996; Steinberg *et al.*, 2001), diatom abundance (Nelson and Brzezinski, 1997), photosynthetic picoplankton abundance (Olson *et al.*, 1990a), microbial size spectra (Gin *et al.*, 1999) and heterotrophic bacterioplankton dynamics (Carlson and Ducklow, 1996). We studied microbial community composition at BATS with the high-throughput community profiling technique, T-RFLP, targeting the small subunit rRNA gene (16S rRNA) (Liu *et al.*, 1997). This approach enabled us to process large numbers of samples reliably and reproducibly (Nocker *et al.*, 2007), and relate the data to large 16S rRNA gene sequence databases. The findings span a timeframe of 12 years, and show in rich detail the highly structured patterns of microbial communities within the surface 300 m of the oligotrophic Sargasso Sea.

Materials and methods

Sample collection and nucleic acid extraction

Monthly samples (1 m and 200 m) were collected at the BATS site (31° 40'N, 64° 10'W) from August 1991 to February 1994 and from September 1997 to January 2004. In addition, depth profile samples (1, 40, 80, 120, 160, 200, 250 and 300 m) were collected at 35 sampling dates (see overview in Table 1 and full table in Supplementary Material S1). Altogether, 412 samples were included in this study. To collect the microbial biomass, seawater was filtered through 0.2-µm polyethersulfone membranes (Supor, Pall, East Hills, NY, USA) and was stored at –80 °C until further use. Nucleic acids were extracted and processed as described earlier (Giovannoni *et al.*, 1996; Morris *et al.*, 2005). During the extractions, nucleic acids were separated isopycally by buoyant density centrifugation in

Table 1 Number of samples collected at specified depths for each month spanning the years 1991–2003

Depth (m)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	9	9	8	8	7	7	7	6	10	8	8	10
40	3	5	4	1	3	0	3	3	2	0	5	0
80	3	6	5	1	3	1	3	3	2	0	4	0
100	0	0	0	0	0	1	0	0	1	0	0	0
120	3	7	5	1	3	1	3	4	2	0	5	0
160	3	6	5	0	3	1	2	4	2	0	5	0
200	8	9	8	7	7	7	6	8	9	9	8	9
250	2	4	4	1	3	1	1	4	2	0	5	0
300	2	4	3	0	3	1	1	3	2	0	4	0

cesium salts and the yields were measured by spectrophotometry.

In silico predictions of microbial diversity

To determine the number of enzymes needed to provide good resolution for the T-RFLP analysis, we created a data set of 16S rDNA sequences from Genbank (Benson *et al.*, 2008) and unpublished sequence data representing all major groups of bacteria currently known to be present in marine environments. Altogether, more than 6500 sequences were analyzed with *ARB* (Ludwig *et al.*, 2004) and *TRF-CUT* (Ricke *et al.*, 2005). The enzymes *BsuRI*, *Bsh1236I* and *MspI* were predicted to provide the highest resolution of marine microbial communities. Each rRNA sequence or clade of sequences gave a specific fingerprint of three fragment sizes for the three enzymes, which we refer to below as 'triplets'. During the T-RFLP analysis, we often observed a difference between the apparent sizes of fragments estimated from their electrophoretic mobility and their *in silico* predicted sizes. This phenomenon (called T-RF drift) is documented in the literature and could be correlated with the purine content of DNA (Kaplan and Kitts, 2003).

PCR, restriction digestion and T-RFLP analysis

DNA was diluted to 5 ng μl^{-1} and 1 μl was amplified with the primer pair 8F-FAM (5-carboxy-fluorescein labeled) and 519R (both Morris *et al.*, 2005) in 50 μl reaction volume using the following conditions: 30 cycles, denaturation at 94 °C for 15 s, annealing at 55 °C for 30 s and elongation at 72 °C for 1 min. Acetamide was added at a final concentration of 5% to reduce G + C content-induced amplification bias (Reysenbach *et al.*, 1992). PCR products were analyzed on gels and restricted in parallel with *BsuRI*, *Bsh1236I* and *MspI* (Fermentas, Hanover, MD, USA). To avoid overestimation of diversity because of partially digested fragments (Nocker *et al.*, 2007), the restriction reactions were incubated overnight with an excess of enzyme. Restriction products were cleaned using a MultiScreen 96-well

filtration plate (Millipore, Billerica, MA, USA) containing Amersham Sephadex G-50 superfine (Amersham, Piscataway, NJ, USA). Terminal restriction fragments were resolved on an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA). To size the fragments based on the internal standard (Bioventures MapMarker 1000, Cambio, Cambridge, UK), the ABI Genotyper software was used. The samples that failed to amplify consistently because of low DNA concentrations or that had incomplete T-RFLP data were not considered for further analysis.

Clone libraries

To correlate observed peaks with species and to verify predicted triplets, clone libraries were constructed from the samples BATS136, 160 m; BATS136, 250 m (both January 2000); BATS138, 160 m (March 2000); BATS146, 250 m (November 2000), BATS174A, 300 m (March 2003) and BATS180, 1 m (September 2003) as described in Morris *et al.* (2004). Samples were chosen on the basis of the likelihood that they contained T-RFLP peaks of interest. Ninety-six clones from each library were screened as described earlier (Vergin *et al.*, 2001), except that clones were amplified and processed for T-RFLP analysis as described above. Plasmids from 175 clones exhibiting T-RFLP peaks of interest were isolated and sequenced using the primer 519R. After initial phylogenetic analysis of the short sequences, seven full-length sequences were generated from clones and their phylogenetic origins were reconstructed (see Supplementary Table S2 for details). The sequences were deposited in GenBank under the accession numbers FJ960798 to FJ960822.

De-noising, alignment and triplet determination

The raw data sets for all three enzymes were standardized by the method of Osborne *et al.* using a variable percentage threshold (Osborne *et al.*, 2006). To cope with minimal differences between individual T-RFLP runs, we applied a clustering by reduction of intervals algorithm that was originally developed for mass spectrometer data (Finlay *et al.*, 2009). The resulting data set was analyzed by different methods to (i) find triplets of correlated fragments for the three different enzymes used, (ii) find seasonal and spatial patterns and (iii) make predictions about the species represented by the respective triplets.

To find triplets, we developed a new algorithm (Burton *et al.*, 2009). In brief, patterns for the presence and absence of fragments were analyzed to find the most likely triplets. These predictions were verified by comparing them to the database generated during the initial computational analysis of diversity and clone libraries. Species were assigned on the basis of comparisons of the

predicted triplets to the virtual T-RFLP results obtained from the *TRF-CUT* analysis of the marine 16S rRNA gene database. These bioinformatic assignments were supported by comparisons to clones from the already published (Giovannoni *et al.*, 1996; Morris *et al.*, 2005) and newly generated clone libraries. For this, clones were pre-screened by T-RFLP, and those producing triplets observed in the environmental data set were sequenced and their phylogenetic origins were reconstructed. To plot the microbial distributions shown in the figures, a 'triplet minimum' was calculated for each data point. For each environmental sample, a triplet yields three relative abundance values, one for each of the three restriction digestions. The lowest of these three values is the 'triplet minimum'. For example, if the relative abundances of fragments for a sample were 20% for *BsuRI*, 16% for *Bsh1236I* and 18% for *MspI* for a given triplet, the triplet minimum would be 16%. This approach minimizes overestimation of abundance caused by interference between fragments originating from different organisms.

Statistical analysis of the T-RFLP data set

After standardization with a variable percentage threshold (Osborne *et al.*, 2006), a non-metric multidimensional scaling (NMS) analysis was carried out to reveal non-linear trends in community structure. For this purpose, the data from all three enzymes were concatenated to build a master matrix including every sample. Before carrying out the statistical analysis, fragments that were present in fewer than five samples were removed. In addition, samples with an average distance more than three standard deviations from the mean distance were excluded from the analysis. The resulting matrix (360 samples \times 590 T-RFLP markers) was analyzed using the *PC-ORD 5* software package (McCune and Grace, 2002). Sørensen distance measurement and the 'slow and thorough' autopilot setting (250 ordinations with real data, 250 randomized Monte Carlo runs, 500 iterations) were chosen. Final stress of the stable, two-dimensional solution was 15.8 with an instability of 0.00428. Probability of a similar stress in the randomized data set was 0.004. The plot was rotated to maximize overlap between axis 2 and the depth from which the samples were collected. By overlaying the NMS plot with physical, chemical and biological measurements from BATS (depth, month of the year, temperature, salinity, several nitrogen species, dissolved organic carbon, bacterial cell counts and pigments, including chlorophyll), correlations between the T-RFLP data and these parameters were determined.

The analysis of similarities (ANOSIM) was carried out using PRIMER v6 (Clarke, 1993). Groups were defined as upper euphotic zone (UEZ), with surface, 40 and 80 m samples, DCM, with samples from 120

and 160 m, and upper mesopelagic (UMP), with samples from 200, 250 and 300 m. A Sørensen similarity matrix of the complete data set was used for the one-way ANOSIM with the number of permutations set to 999.

Annual deep convective mixing extends the mixed layer depth into the mesopelagic zone every year at BATS and affects the surface ocean biogeochemistry at this site (see Steinberg *et al.*, 2001 and citations within). To examine the microbial community variability relative to the mixing event, we aligned each year's data to the timing (month) of the deep mixing event, and collapsed the decade of triplet minima data to a single composite year. These data were plotted for the final figures using *Ocean Data View* (Schlitzer, 2000). Seasonal and spatial (depth distribution) patterns were identified by checking for the presence/absence of the fragments at every sampling date and depth using a Microsoft Excel spreadsheet.

Richness, mixed layer depth, and primary productivity

Richness was calculated as the average number of the observed T-RFLP markers for each enzyme and depth. The results for the different enzymes were then combined and averaged. Mixed layer depth was calculated as the depth where sigma-*t* was equal to sea surface sigma-*t*, plus an increment in sigma-*t* equivalent to a 0.2 °C temperature decrease (Sprintall and Tomczak, 1992). Primary production data were extracted from the public website of the Bermuda Atlantic Time-series Study (<http://bats.bios.edu/index.html>).

Results

Sample collection and T-RFLP measurements

The analysis we discuss here included 412 samples collected over a period of 12 years and 9 depths ranging from the surface to 300 m (see Table 1 and Supplementary Material Table S1). We chose T-RFLP techniques to profile microbial community structure from a decade of nucleic acid samples because they are high-throughput and T-RFLP fragments can be predicted *in silico* from 16S rRNA sequence databases. An initial computational analysis of a marine microbial 16S rRNA gene data set using *ARB* (Ludwig *et al.*, 2004) and *TRF-CUT* (Ricke *et al.*, 2005) was carried out to identify the number and types of restriction enzymes to be used. We settled for *BsuRI*, *Bsh1236I* and *MspI* because they were predicted to have the highest resolution and the least redundancy. Most of the sequences in the marine data set were cut by all three enzymes, whereas only a small percentage was cut by one or two of the enzymes. Using the three enzymes in independent digests increased phylogenetic specificity, and partially overcame the problem of multiple organisms contributing to individual fragment

lengths (Marsh *et al.*, 2000; Kent *et al.*, 2003). Adopting this approach, we developed the triplet minimum method to reduce interference between phylogenetic groups in the estimation of microbial abundances. Of the 412 samples, 369 produced T-RFLP profiles for all the three restriction enzymes and were thus used in the analysis. Higher resolution of the microbial community was obtained with the enzymes *Bsu*RI and *Msp*I than with *Bsh*1236I, which was not predicted from the initial computational analysis of the marine sequences from the database. Altogether, 302 different fragment lengths were detected for *Bsu*RI, 248 for *Bsh*1236I and 289 for *Msp*I. On average, 42 and 38 fragments were observed in the surface samples and 55 and 56 fragments in the 200 m samples, for *Bsu*RI and *Msp*I, respectively. In contrast, *Bsh*1236I digestions yielded an average of only 30 fragments from surface samples and 35 fragments from 200 m samples.

Richness and RNA/DNA ratios

The average richness of markers detected by the three different restriction enzymes increased below the euphotic zone (Figure 1). Interestingly, the RNA/DNA ratios of the entire planktonic community, an indicator of metabolic activity, decreased with

increasing depth and were inversely related to the observed richness of bacterial species (Figure 1). The greatest transition in the trends of the RNA/DNA ratio and richness occurred at ~ 140 m (Figure 1), the average depth of the 0.1% light level and the beginning of the mesopelagic zone. As with species richness, seasonality in RNA/DNA ratios was much less pronounced than vertical variation (data not shown).

Patterns in microbial community composition

The northwestern Sargasso Sea undergoes annual cycles of productivity with spring phytoplankton blooms corresponding to periods of convective mixing, and productivity lows coincident with periods marked by extensive oligotrophy in summer and autumn, when the water column becomes highly stratified (Steinberg *et al.*, 2001). NMS ordination of the complete data set resolved four microbial communities (Figure 2). During seasonal stratification, surface (UEZ; 0–80 m), deep chlorophyll maximum (DCM; 120–160 m) and deep samples (UMP; 200–300 m) clustered separately. The ‘spring bloom’ community resolved separately in the ordination, propelled in particular by eukaryotic phytoplankton, which emerged in the analysis because plastid 16S rRNA genes co-amplified with bacterial 16S rRNA genes.

We used the ANOSIM algorithm to provide statistical support for the microbial communities we report here. Defining ANOSIM analyses on the basis of ordination results (Figure 2) would not provide a reliable test because of circular reasoning. Therefore, we defined the communities for ANOSIM analyses based on depth horizons (Supplementary Figure S3). The *R*-value received for the pair-wise comparison UEZ–UMP was 0.78, showing that they are fairly distinct communities. The DCM community had similarities with both the UEZ and UMP communities (DCM–UEZ: *R* = 0.39, DCM–UMP: *R* = 0.36) (see inset Supplementary Figure S3). The unique characteristics of the spring bloom community have been discussed earlier by Morris *et al.* (2005).

Although the trends identified by ordination are the dominant patterns of microbial community variation, many individual examples of spatial and temporal patterns were discerned by examining the distributions of individual fragments (Figure 3). Approximately half of the fragments showed seasonal patterns and half showed vertical patterns, with some degree of overlap between the groups. When we excluded fragments that showed little variation throughout the year, 39% of all fragments exhibited distinct seasonal bloom patterns, whereas others varied with no discernable pattern. Similarly, after excluding fragments that were present at all depths, 43% of all fragments showed stratified distributions. For a better comparison, the abundance of fragments with specific patterns were normalized to the

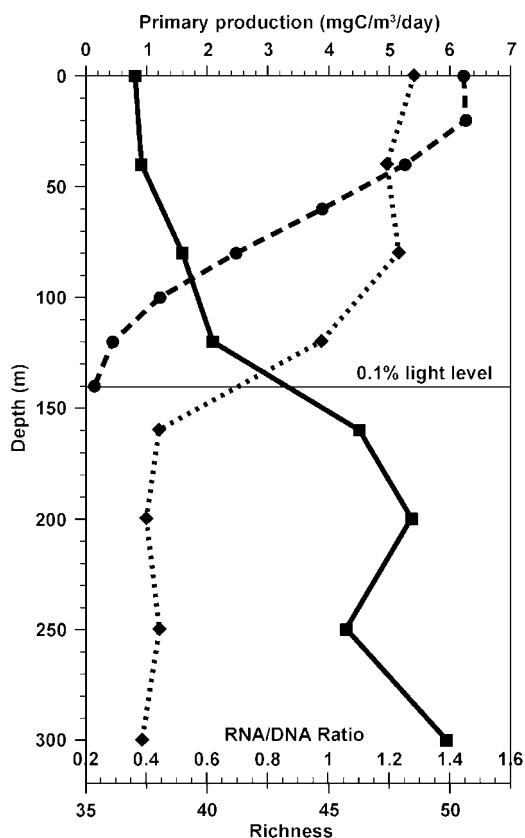


Figure 1 Average species richness, primary production and RNA/DNA ratios in the surface layer. Richness is indicated by the solid line, primary production by the dashed line and RNA/DNA ratios by the dotted line. The beginning of the mesopelagic zone (0.1% light level at 140 m) is marked.

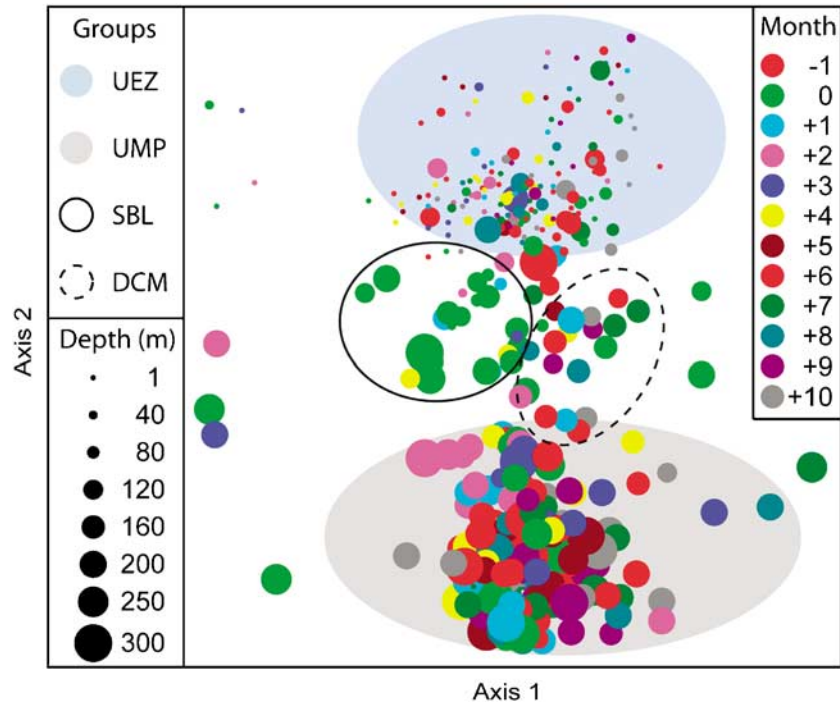


Figure 2 Non-metric multidimensional scaling ordination of 360 samples in terminal restriction fragment length polymorphism marker space consisting of 590 unique fragment lengths. Samples sharing similar markers cluster in this analysis, which takes into account the magnitudes of the markers. Sizes of dots are proportional to the depths from which the samples were taken (1, 40, 80, 120, 160, 200, 250 and 300 m). Color shows the sampling month, indicating that seasonality is much less conspicuous than vertical stratification. The ordination was rotated to maximize overlap of axis 2 with depth. DCM, deep chlorophyll maximum; SBL, spring bloom; UEZ, upper euphotic zone; UMP, upper mesopelagic.

number of all fragments that showed non-random patterns, including those present in all seasons or at all depths, and then averaged. This was possible because all three enzymes exhibited the same trends (r^2 between enzymes 0.59–0.84 for seasonal and 0.9–0.92 for spatial patterns). Disregarding fragments that were present year-round, the most common pattern for fragments that displayed seasonal trends (Figure 3a) was high abundance in the spring (14.3%), as suggested by the NMS analysis, followed by those showing a summer/fall/winter maximum (12.4%), a spring/summer maximum (9.1%) and a summer only maximum (8.1%). Similarly, disregarding fragments that were present throughout the water column, most of the fragments with vertical trends (Figure 3b) showed a 40–300 m distribution (25.8%), followed by fragments that were present in the surface waters down to 160 m (20.1%) and fragments that were only observed in waters 200 m and deeper (15.6%).

We took a conservative approach to assign fragments to taxa (Figure 3). Generally, if a fragment could not be assigned to a triplet, we considered the phylogenetic origin of the fragment uncertain and did not analyze the fragment further. A summary of patterns analyzed in this study or associated with the spring bloom at BATS is shown in Figure 3c.

The relative abundance of fragments from the SAR11 clade of *Alphaproteobacteria* averaged 19%,

overshadowing all other organisms. A detailed analysis that combined these T-RFLP data with SAR11 fluorescence *in situ* hybridization cell counts to model temporal and vertical distribution patterns in SAR11 ecotype populations has been published (Carlson *et al.*, 2009).

The alphaproteobacterial clade OCS116 bloomed with phytoplankton in the spring (Figure 4a). Morris *et al.* (2005) reported this earlier; however, our analysis showed a previously undetected population of OCS116 in the region of the DCM during the period of water column stratification. An additional group of organisms that as stratified in the DCM included members of the phylum *Actinobacteria* which are closely related to uncultured *Microthrix* species (Supplementary Figure S4). They were present in the DCM year-round, with occasional blooms during the summer stratification period.

Another well-studied group of microorganisms at BATS is the marine picocyanobacteria clade. Our results with these organisms (Figure 4b) corroborate earlier findings, showing the validity of our approach. *Synechococcus* and *Prochlorococcus*, which are about 97% similar in 16S rRNA sequence, displayed the same triplet (*Bsu*RI 291 bp, *Bsh*1236I 95 bp, *Msp*I 471 bp), a finding that the clone libraries confirmed. Composite plots of the marine picocyanobacteria clade triplet minima confirmed a high relative abundance (up to 35% of the total fragment fluorescence) of these organisms in the summer

months in the euphotic zone (Figure 4b). The prominent T-RFLP pattern of the marine cyanobacteria resembles *Prochlorococcus* distributions reported earlier in the literature (Figure 4b) (Durand *et al.*, 2001; Heywood *et al.*, 2006). Although our analysis could not resolve *Synechococcus* from *Prochlorococcus*, those studies showed that *Synechococcus* is most abundant during spring deep mixing and *Prochlorococcus* populations reach their maximum abundances during summer and fall. In

addition, Durand *et al.* observed that *Prochlorococcus* was ten times more abundant and had a deeper population maximum than *Synechococcus* (Durand *et al.*, 2001). They also showed that, during the winter months, these organisms are closer to the surface, and when the maximum MLD extends below the euphotic zone in late winter and early spring, their relative abundance is reduced (Durand *et al.*, 2001; Steinberg *et al.*, 2001).

Similar to earlier reports, we observed that the SAR202 clade of *Chloroflexi* was mainly found in the aphotic zone (Morris *et al.*, 2004). However, the better resolution afforded by this analysis allowed us to observe an earlier unseen summer bloom of this clade in the upper mesopelagic zone (Figure 5a). Although nearly absent from the euphotic zone, SAR202 blooms in the summer in the upper mesopelagic zone reached a relative abundance of up to 2% in May through June (3 and 4 months after deep mixing), compared with an average yearly upper mesopelagic value of 0.5–1%.

The deltaproteobacterial clade, SAR324/Marine Group B, which was initially described by Fuhrman and Davis (1997) and Wright *et al.* (1997), was also prevalent in our data set (Figure 5b and Supplementary Figure S5). The clade's depth distribution increased from 1.9% of the relative fragment abundance at the surface to 14.4% at 300 m. Recently, Brown and Donachie (2007) delineated three species-level subclades (named I, II and SAR276) and estimated their abundances in different geographical locations on the basis of clone libraries and database entries. We were able to distinguish between the subclades on the basis of their triplets, with subclade I contributing the greatest average relative abundance (9%) in the upper mesopelagic zone. Subclade II, which was only detected below 160 m, reached 2% relative abundance at 300 m. The SAR276 subclade was present throughout the upper 300 m of the water column, with a mean relative abundance that increased from 1.6% at the surface to over 3% at

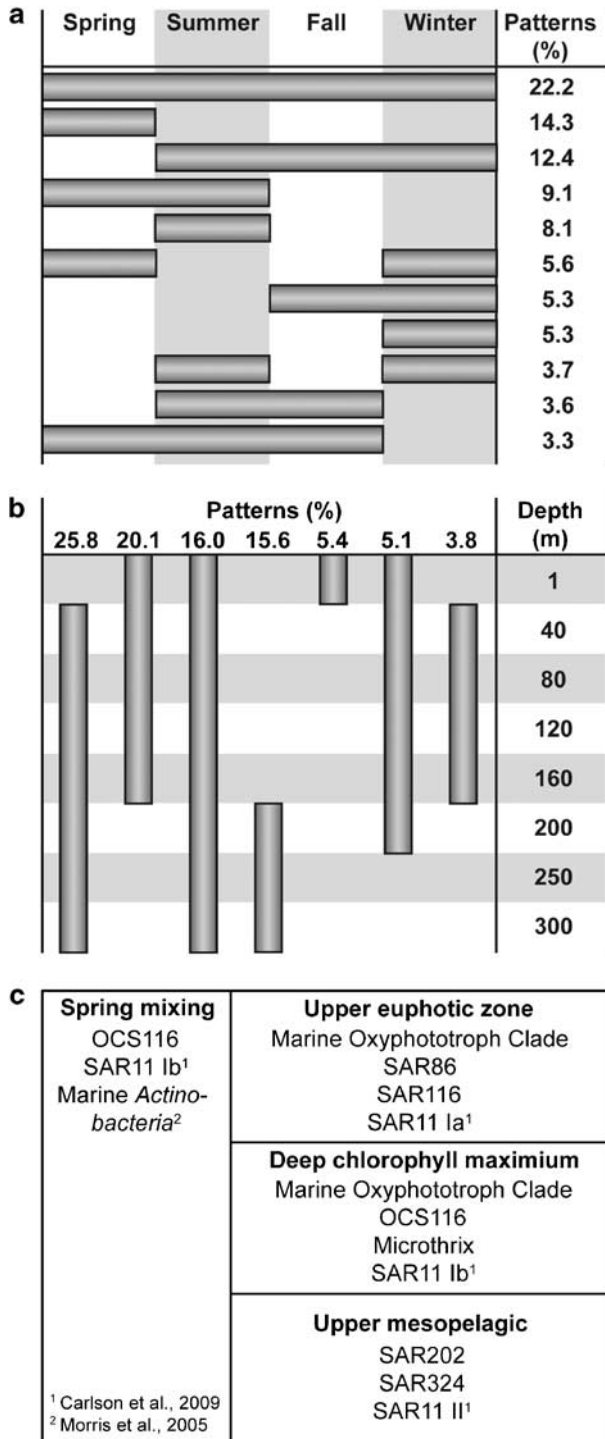


Figure 3 Seasonal (a) and vertical (b) patterns inferred from individual terminal restriction fragment length polymorphism peaks and (c) organisms that contribute significantly to the distinctness of the upper euphotic zone, deep chlorophyll maximum, upper mesopelagic and spring bloom microbial communities. (a) The most common pattern was occurrence only in the spring, followed by a summer/fall/winter pattern. Seasons were defined as follows: December–February, winter; March–May, spring; June–Aug, summer; September–November, fall. Rare patterns (those that were represented by <3% of all fragments) are not shown. Altogether, 6.9% of all fragments were in the ‘rare pattern’ category. (b) Fragments showing vertical distribution patterns did not follow the euphotic/mesopelagic zone separation seen in the non-metric multidimensional scaling analysis, but rather had natural borders at 40 m and between 160 and 200 m. Rare patterns are not shown. Altogether, 8.1% of all fragments were in the ‘rare’ pattern category. (c) Microbial clades observed during different times and in the different zones of the water column.

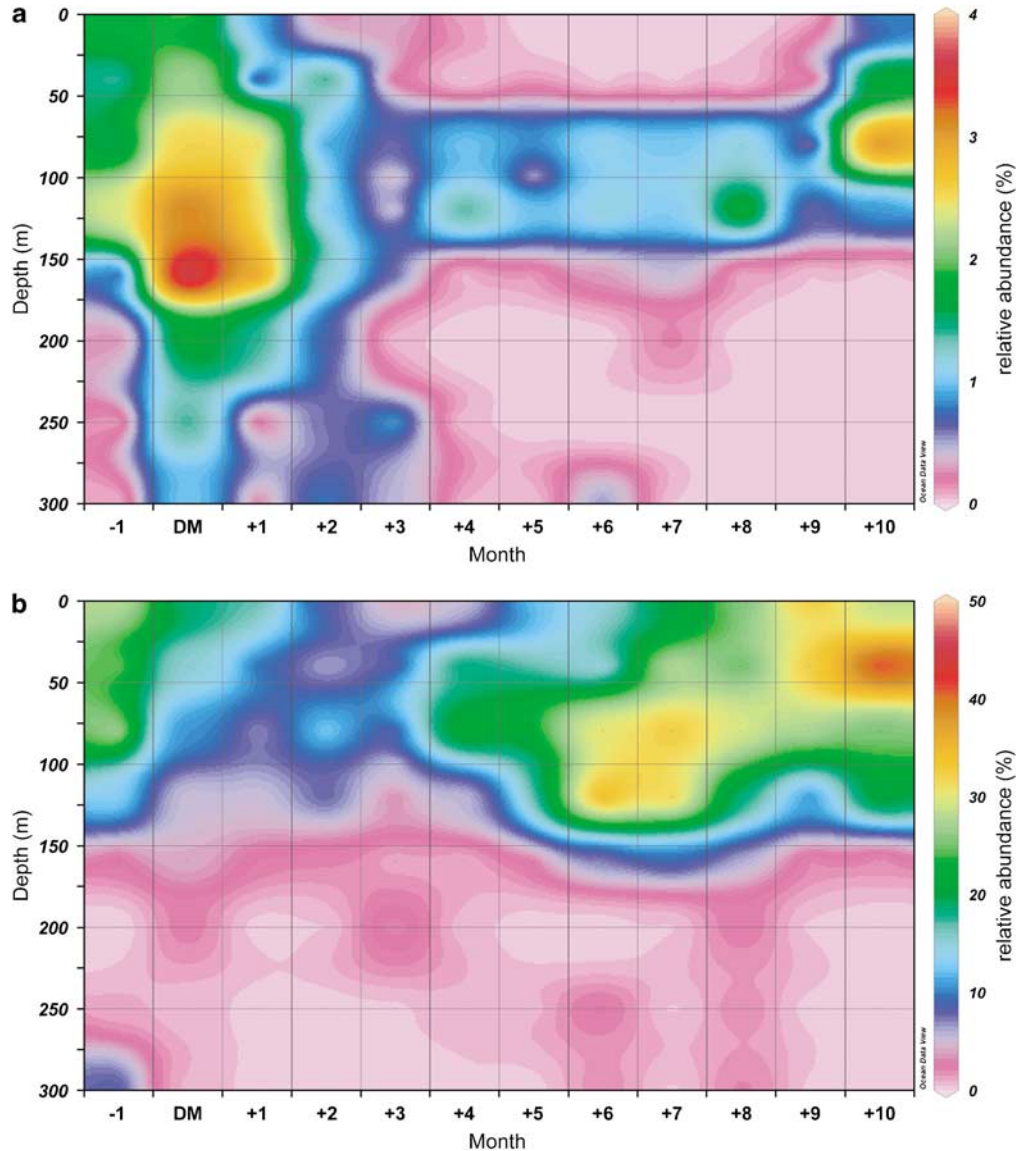


Figure 4 Averaged seasonal composites of (a) OCS116 and (b) the marine picocyanobacteria clade distributions in the surface 300 m at Bermuda Atlantic Time-series Study. (a) The plot shows the spring bloom and summer deep chlorophyll maximum (DCM) occurrence of OCS116 clade bacteria. After the nutrient intake by deep mixing and a subsequent spring bloom of OCS116 bacteria, these organisms persist in a depth horizon of 80–120 m, the area of the DCM, during the period of summer stratification. This might indicate an association of this clade with photosynthetic organisms. (b) The sum of the relative abundances of *Prochlorococcus* and *Synechococcus* are shown. DM, the month of the deep mixing event.

300 m. No seasonal patterns were detected among the SAR324 subclades (data not shown).

One striking pattern that emerged from this analysis was a seasonal bloom of the SAR86 clade in the mixed layer during the height of summer stratification (Figure 6a). The SAR86 clade belongs to the *Gammaproteobacteria* and is one of the most common groups observed in clone libraries made from surface water samples. They attracted attention when proteorhodopsin was first discovered in the fragment of a SAR86 genome (Béjà *et al.*, 2000a). We used phylogenetic reconstruction, with confirmation by gene cloning and sequencing, to resolve the subclades of SAR86 (Suzuki *et al.*, 2004) and show that they can be distinguished by the T-RFLP

methods we used, but not without some overlap (Figure 7; Supplementary Figures S6a and b). The summed relative abundances of all SAR86 fragments showed that the SAR86 clade is present throughout the year, but increases in relative abundance from April to October (months 1–7 after DM), reaching a peak in relative abundance of 3% in June and July (months 3 and 4 after DM; Figure 6a). They were nearly absent from waters below 100 m, except in May after spring convective overturn, when they were observed as deep as 150 m. Surprisingly, all the SAR86 subclades exhibited similar distribution patterns when analyzed separately (data not shown). In contrast to other studies (Suzuki *et al.*, 2004), we did not include the clades CHAB-I-7 and San Pedro

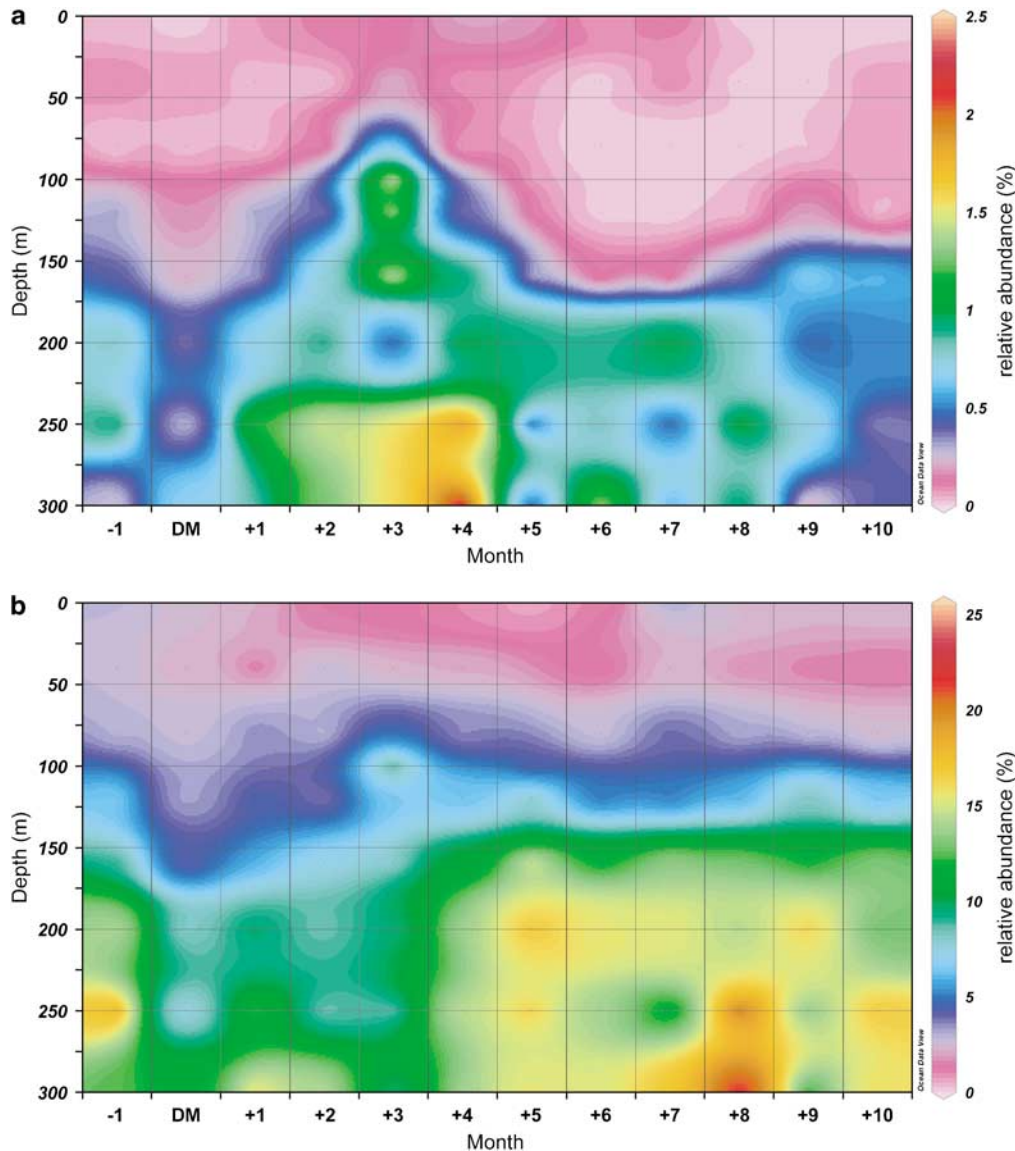


Figure 5 Averaged seasonal composite plots showing (a) the summer, upper mesopelagic bloom of SAR202 clade and (b) the year-round upper mesopelagic occurrence of the SAR324 clade. (a) To represent a composite year in the context of deep mixing, each year of the 12-year data record was first aligned to the month of each year's maximal mixing and expressed as the month before and 10 months after deep mixing. The values plotted are triplet minima. (b) Similar to SAR202, the SAR324 clade is found in the mesopelagic zone. SAR324 populations decrease in abundance during spring mixing, and generally extend to shallower depths than SAR202. DM indicates the month of the deep mixing event.

Ocean Time-Series (SPOTS) in the SAR86 clade because they are phylogenetically distant.

The alphaproteobacterial clade SAR116 was also prevalent in the data set. These organisms had a distribution similar to that observed for SAR86; during June, July and August (months 3–5 after DM), a strong bloom was observed in the upper 50 m of the water column, reaching relative abundances of about 20% (Figure 6b). The SAR86 bloom begins with the onset of water column stratification after spring convective overturn. In a phylogenetic reconstruction of selected full-length SAR116 sequences from the databases, different subclades could be delineated (Supplementary Figures S7 and S8). Suzuki *et al.* (2001) described SAR116

subclades I, II and III and our phylogenetic analysis, supported by triplets, showed a further subgrouping of these subclades. Subgroups of SAR116 subclade I showed distinct triplets (Ia 192/96/178 bp; Ib 189/93/175 bp for *BsuRI/Bsh1236I/MspI*). Similarly, SAR116 subgroup II also exhibited distinct triplet patterns. Subgroups IIa (226/196/149 bp) and IIc (289/93/439 bp) were predicted to have consistent triplets; however, subgroup IIb, did not produce a consistent triplet across all analyzed sequences. Therefore, subgroup IIb could not be traced in the T-RFLP data set. SAR116 subclade III resolved into two different phylogenetic subgroups (IIIa and IIIb) that did not have distinguishable triplets. We detected two additional subclades that have not

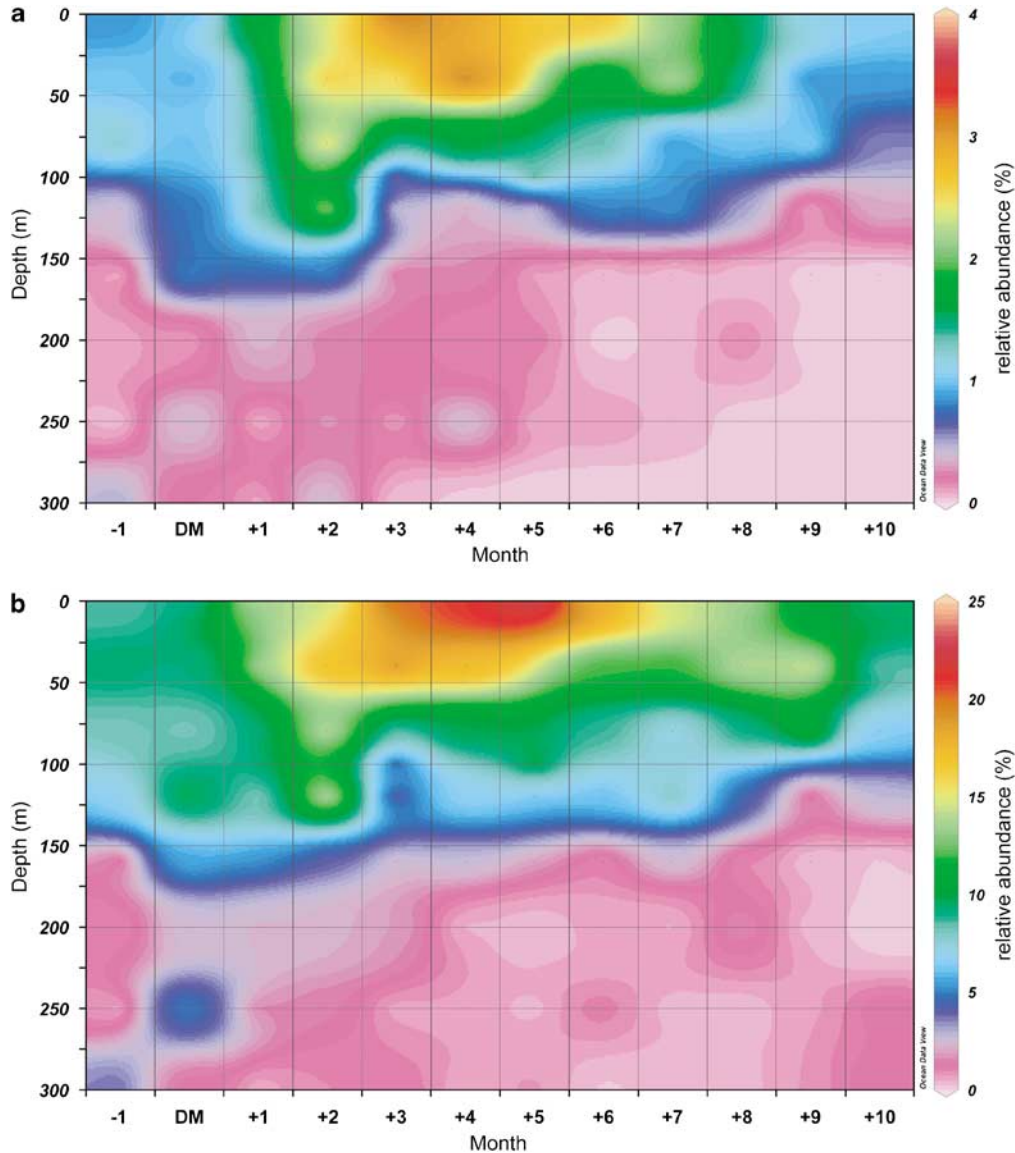


Figure 6 Averaged composite plot showing annual summer blooms in the upper euphotic zone of (a) the SAR86 clade and (b) the SAR116 clade at the Bermuda Atlantic Time-series Study site. (a) SAR86 relative abundance is highest in the summer mixed layer, which extends to 10–20 m. This plot sums all of the SAR86 subclades shown in Figure 7. Identical seasonal and spatial distributions were observed for the subclades, despite their forming distinct phylogenetic groups that suggest ecotypes. These data indicate that both clades are most abundant in the summer mixed layer; however, the scales are different, suggesting that either the SAR116 clade is relatively more abundant than SAR86 throughout the year, or that it has more copies of the 16S rRNA gene. DM indicates the month of the deep mixing event.

been reported earlier, IV and V, with triplets of 193/97/439 and 193/197/163 bp, respectively. Owing to overlap between fragments for subclades III, IV and V, we were not able to distinguish between these groups.

Discussion

The Bermuda Atlantic Time-series Study provided an oceanographic context to study the patterns of microbial community change in an ocean gyre. The surface layer of the BATS site differs from some

other oligotrophic gyres, for example, the Hawaii Ocean Time-series site, in that it is seasonally oligotrophic marked by periods of deep mixing followed by long periods of highly stratified low nutrient conditions. Each winter/spring, there are deep convective overturn events at BATS where the MLD extends to as deep as 350 m, entrain nutrients to the surface, and trigger annual phytoplankton blooms that result in rapid nutrient drawdown and increased organic particle flux (Steinberg *et al.*, 2001). As the water column begins to stratify, dissolved organic matter accumulates in the euphotic zone (Carlson *et al.*, 1994; Hansell and

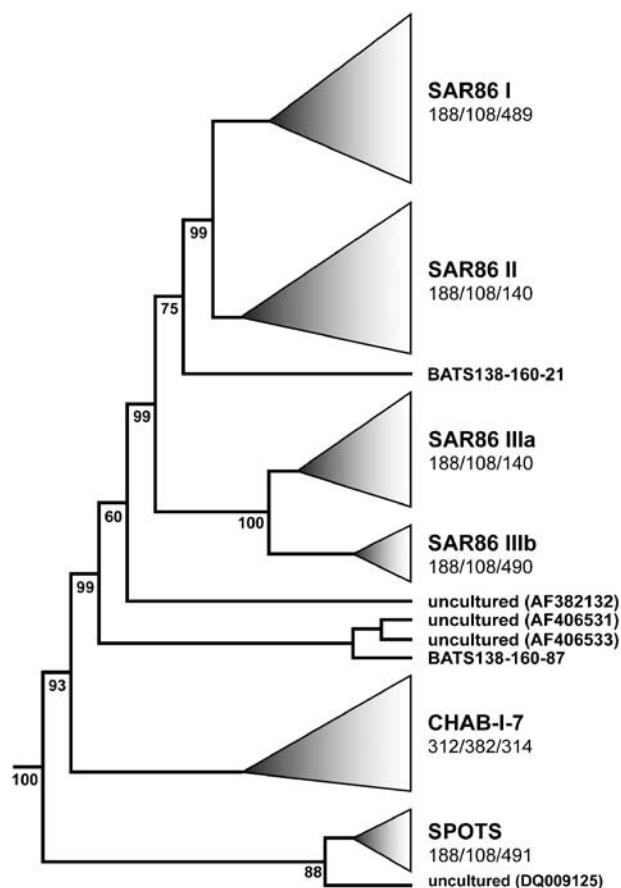


Figure 7 16S rRNA evolutionary tree of the SAR86 clade of *Gammaproteobacteria*, showing the phylogenetic distribution of triplets (*BsuRI/Bsh1236I/MspI*). Naming of the subclades follows Suzuki *et al.* (2001); however, we observed two subgroups within SAR86 subclade III (IIIa and IIIb) that were not described earlier, but are consistent with terminal restriction fragment length polymorphism triplets. Further, two of the subclades shown (CHAB-I-7 and San Pedro Ocean Time-Series (SPOTS)) are not included in the SAR86 clade in some studies. The sequence of *Alcanivorax borkumensis* (AF062642) was used to set the root of the tree, which was calculated using the Neighbor-Joining algorithm and is based on 1423 nucleotides of sequence. Numbers indicate the percentage of bootstrap replicates that support the nearby branch. Dominant triplets are indicated for all subclades. Some sequences, including clones from this study, did not group with the major clades and might represent additional, less abundant clades.

Carlson, 2001). A broader assessment of the BATS biogeochemical data is presented elsewhere (see Deep Sea Research II volumes 43 No. 2–3 and 48: No. 8–9 for relevant studies).

Richness

Species richness—the total number of species in a sample—is perhaps the broadest measure of community diversity. Unexpectedly, we found that richness increased significantly below the euphotic zone (Figure 1). Although richness does not necessarily translate to phylogenetic diversity, a factor in

the lower diversity of the euphotic zone may be that it is largely the domain of only two phyla, the *Cyanobacteria* and *Proteobacteria*, whereas the upper mesopelagic zone harbors large populations of bacteria from the *Chloroflexi* and *Fibrobacteres/Chlorobi* phyla, as well as *Deltaproteobacteria* that are not common in the euphotic zone. The upper mesopelagic zone's microbial community was also considerably more uniform in richness throughout the year (data not shown), and occupies a region where fluctuations in light, temperature and nutrients are much less extreme. Variability in export flux of both particulate and dissolved organic matter (Steinberg *et al.*, 2001) in combination with solubilization of sinking particles (Azam and Long, 2001; Azam and Malfatti, 2007) within the mesopelagic zone potentially introduce heterogeneity of available organic substrates in an otherwise nutrient-rich thermally stable environment. All of these factors may contribute to the greater diversity we observed in the upper mesopelagic microbial community. Interestingly, Hewson *et al.* (2006) found higher richness in 500 and 3000 m samples than in samples taken at 1000 m, indicating additional factors controlling richness in the deeper meso- and bathypelagic. Other studies assessing the richness of marine microbial communities focused on surface samples and showed a decrease in richness with an increase in latitude (Pommier *et al.*, 2007; Fuhrman *et al.*, 2008). Richness, latitude and temperature were highly correlated, indicating that higher temperatures support a richer community structure on the geographical scale. However, in our data set, temperature and richness were inversely related, suggesting that factors other than temperature control richness on a vertical scale in the oceans.

RNA/DNA ratios

We found that RNA/DNA ratios were inversely related to species richness, declining below the euphotic zone (Figure 1). High RNA/DNA ratios generally indicate greater metabolic activity because of the increased protein synthesis in rapidly growing cells (Dortch *et al.*, 1983). However, across species, RNA/DNA ratios tend to increase with increasing cell size and decrease with increasing genome size (Kemp *et al.*, 1993). Thus, although the high RNA/DNA ratios observed in the euphotic zone may reflect high relative productivity and turnover rates, we suspect that other factors are at work. In particular, relatively small genome sizes have been reported for some euphotic zone bacteria (Giovannoni *et al.*, 2005b, 2008), as well as an increase in percentage of cells showing a higher DNA content with depth (Reinthalder *et al.*, 2006). However, from the information available, it is not possible to identify trends in genome size and cell size that might affect RNA/DNA ratios along the vertical axis in ocean gyres.

Ordination indicates modes of community organization

Four modes of microbial community composition emerged from the ordination analysis: a post-mixing spring bloom community, a stratified surface community, a DCM community and a stratified upper mesopelagic community (Figures 2 and 3c). Apart from mixing events, aphotic and euphotic zones' communities at BATS separated very clearly in the NMS analysis. Some of these findings echo earlier reports, particularly those of Morris *et al.* (2005) and Carlson *et al.* (2009), who observed that spring mixing events at BATS have a significant effect on microbial community composition, and the numerous reports which have shown euphotic zone and mesopelagic microbial communities are distinct (Giovannoni *et al.*, 1996; Gordon and Giovannoni, 1996; Fuhrman and Davis, 1997; Wright *et al.*, 1997). However, the DCM community that emerged in the ordination (Figure 2) has not been recognized previously as a separate community with a distinctive composition, as we report here. As evident in the ordination, the DCM community is related to the spring bloom community in its composition. This is discussed further below.

Inspection of individual T-RFLP fragment plots supported the ordination results, but also underscored that the four communities revealed by ordination are generalizations from many individual distribution patterns that are unique to organisms whose natural histories do not necessarily follow the general trends. For example, less than half (43%) of all fragments having stratified depth distributions, and one-fourth of all fragments showing non-random patterns were present from 40 m down to 300 m (Figure 3b). This perspective, illustrated in Figures 3a and b, can be reconciled with the ordination results by considering the importance of 'keystone' microbial groups, which affect ordinations in proportion to their abundance and thereby contribute to the distinctness of communities. Some T-RFLP fragments could not be identified. Possible explanations, for example, episodic blooms and pseudo T-RFs (Egert and Friedrich, 2005) are discussed in the Supplementary Material S9.

The spring bloom and the DCM microbial community

In broad strokes the paragraphs above define key features of the BATS surface layer microbial ecosystem, but its dynamic nature is illustrated nowhere more aptly than in the post-mixing spring bloom and ensuing transition to stratification. Some of these dynamics are shown in Figures 4–6, but also in the Morris *et al.*'s. (2005) study of seasonal mixing and the Carlson *et al.*'s (2009) study of SAR11 seasonal dynamics.

The annual composite plot of OCS116 (Figure 4a), an uncultured alphaproteobacterial clade, revealed a spring bloom of these organisms, as previously reported by Morris *et al.* (2005), but also showed that

these organisms persist during the remainder of the year as a population in the DCM microbial community. Other organisms follow a pattern similar to that shown here for OCS116. For example, we observed that members of a marine clade of *Microthrix* were also abundant in the DCM community (Supplementary Figure S4), and the SAR11 Ib ecotype follows a similar pattern (Carlson *et al.*, 2009).

Despite some common elements of microbial community structure, the retreat of photoautotrophs to the DCM, brought about by declining nutrient concentrations in the UEZ, is also accompanied by changes in community composition. Figure 4b shows one of the strongest patterns to emerge from the data set—the average yearly distribution of marine *Synechococcus* and *Prochlorococcus*. These organisms replace eukaryotic phytoplankton as the dominant phototrophs after the spring bloom.

Although above we draw attention to the most prominent patterns of seasonal change, seasonal patterns were also prominent in plots for many individual T-RFLP fragments. Indeed, 39% of all fragments followed readily discernable seasonal patterns. In accord with the NMS statistical analysis, the most frequently observed seasonal pattern was the transient occurrence of many fragments in the spring. Many of these were identified as fragments originating from plastid 16S rRNA genes, which amplify with the same primers used for bacterial 16S rRNA gene amplification. These will be analyzed in detail in another publication. Others (SAR86 and SAR116), described below, become dominant in the stratified surface waters of the summer.

The stratified surface community

During stratified periods, the euphotic zone and particularly the upper 40 m of the stratified water column has the highest exposure to solar radiation, and is very low in macronutrients such as phosphorous and combined nitrogen (NO_3^- , NO_2^- and NH_4^+), making it a truly extreme environment. Some features of community stratification that we found, particularly those for SAR11 and photoautotrophs, echo findings from earlier studies. However, the data also unveiled new features, notably highly stratified populations of SAR86 and SAR116 in the summer euphotic zone (Figure 6).

The alphaproteobacterial SAR116 clade produced one of the most prominent T-RFLP patterns in the data set, contributing an average of 12.6% of all fragments above 100 m, and consistently blooming to 20–25% of the total 16S rRNA gene amplicon signal during summer stratified periods (Figure 6b). Although SAR116 is a common bacterioplankton group, until now its association with the euphotic zone has been on the basis of clone library data, which are limited to few samples and lack robust statistical support. Thus, this analysis provides the first overwhelming evidence that SAR116 is a

euphotic zone clade. Even during the fall, winter and spring, SAR116 contributed 7–10% of the T-RFLP signal above 100 m. In contrast, in Global Ocean Survey metagenomic data collected in the winter and spring from the Sargasso Sea, only 2–3% of 16S genes were from SAR116.

Like SAR116, the SAR86 clade of *Gammaproteobacteria* emerged as a highly seasonal clade relegated to the extreme environment of the oligotrophic stratified euphotic zone (Figure 6a). This study confirms their euphotic zone distribution, which previously was largely inferred from the frequent appearance of SAR86 clones in surface samples.

SAR86 and SAR116 clades, as well as the SAR11 Ia and Ib ecotypes (Carlson *et al.*, 2009), are most pronounced in the stratified euphotic zone. Proteorhodopsin has now been observed in many marine microbial groups, including the SAR11 (Giovannoni *et al.*, 2005a) and SAR86 clades (Béjà *et al.*, 2000a,b). It is thought to function as a light-dependent proton pump. Its discovery focused attention on the potential importance of photoheterotrophy in marine bacteria. The adaptations that contribute to the success of SAR86, SAR116 and SAR11 Ia and Ib ecotypes in these high-light low nutrient environments are not yet understood, in part because cultures and genomes are not yet available for SAR86 and SAR116. Our findings invite further scrutiny of these organisms to understand the role of proteorhodopsin, and to identify other factors that could contribute to the seasonal surface blooms we observed.

The upper mesopelagic microbial community

The surface and DCM communities described above overlay a community that is relatively high in mineral nutrients but has insufficient light to support a significant population of photosynthetic organisms. However, it is poised in space to interact with sinking particles and exported dissolved organic matter accumulating from primary production in the layers above.

Observations of the SAR202 clade, a diverse group of uncultured organisms from the phylum *Chloroflexi*, have been largely confined to the deep ocean (Giovannoni *et al.*, 1996; Morris *et al.*, 2004). The SAR202 clade has a complex phylogenetic substructure, suggesting that it may represent substantial phenotypic diversity. These organisms are estimated to constitute about 10% of the total bacterial population in the mesopelagic and bathypelagic realms (Morris *et al.*, 2004; Varela *et al.*, 2008). Our analysis confirmed earlier findings, but also revealed annual blooms of SAR202 populations in the upper mesopelagic zone several months after deep mixing (Figure 5a). The factors that cause these large populations to bloom in shallower regions of the mesopelagic during the summer stratification are unknown.

Similar to SAR202, the SAR324 clade of *Delta-proteobacteria* produced a large T-RFLP signal (average 11.3%, see Figure 5b and Supplementary Figure S5) in the upper mesopelagic zone, but we observed no strong seasonal patterns in their distribution (Figure 5b). They were nearly absent from the euphotic zone (average 2.2%), consistent with earlier reports showing that this uncultured microbial clade is distinctly a meso- and bathypelagic organism (Fuhrman and Davis, 1997; Wright *et al.*, 1997; Gonzalez *et al.*, 2000). We found SAR276, a subclade of SAR324, to be more abundant in the upper 120 m of the water column than subclades I and II (Supplementary Figure S5). Below 120 m, subclade I was much more abundant than both of the other two subclades. This finding contrasts with the observations of Brown and Donachie (2007), who suggested that subclade II is the most abundant in the water column. Their study was on the basis of the analysis of 16S rRNA gene and ITS sequences from different sampling locations, including BATS, and did not consider seasonal changes in abundance.

Sparse evidence for ecotypes

Ecotypes are closely related lineages of bacteria that differ slightly in adaptive physiological traits conferring niche specificity (Cohan, 2006). Differing spatial and temporal distribution patterns that correlate with the fine structure of phylogenetic trees are often used to identify ecotypes. In most cases, the traits that confer niche specificity to ecotypes are unknown (Hunt *et al.*, 2008; Koeppel *et al.*, 2008). Two of the best examples of microbial clades that are composed of multiple ecotypes are SAR11 and *Prochlorococcus*. In the case of *Prochlorococcus*, the distributions of ecotypes and some traits that contribute to niche specificity are both known (Moore *et al.*, 1998). Several reports have documented ecotypes of SAR11. Most recently, and thoroughly, Carlson *et al.* (2009) used the same data we use in this study to model the distributions of SAR11 ecotype populations at BATS.

The phylogenetic substructure of the SAR86 and SAR116 clades (Figure 7 and Supplementary Figure S7) suggests that, like SAR11 and *Prochlorococcus*, these microbial groups may consist of multiple ecotypes. However, we were surprised that evidence for neither SAR86 nor SAR116 ecotypes emerged from this analysis. The failure to resolve ecotypes for these abundant clades is unlikely to be related to the resolution of T-RFLP methods. We could resolve the subclade structure of SAR86 reported in the literature (Suzuki *et al.*, 2001), and additionally delineate several new subclades, which were reflected in the triplets (Figure 7 and Supplementary Figure S6a). Although the clade substructure of SAR86 and SAR116 suggests different ecotypes similar to what was shown for SAR11 (Carlson

et al., 2009), there was no evidence to support the ecotype hypothesis. If there are SAR86 and SAR116 ecotypes at BATS, they are either sympatric, or the differing patterns of their distributions were too fine to be resolved by our experimental design.

Conclusion

This large data set, and the comprehensive power of T-RFLP analysis, showed many earlier unseen features of microbial community population dynamics in the northwestern Sargasso Sea. Some of these features, particularly those observed during the period of summer stratification, are likely to be common features of stratified ocean gyres, whereas others may be peculiar to systems that undergo similar deep convective mixing events and ensuing phytoplankton blooms. A diversity of physical, chemical and biological measurements are taken at BATS on a monthly basis, and were available to us for the NMS analysis to identify factors that might control the distribution of microbial populations. Although many correlations between routine measurements at BATS and microbial distributions were analyzed, none of these emerged that overwrote the dominant theme that seasonal mixing, surface warming, and stratification are the key determinants of transitions of microbial community structure.

Many oceanographic processes—and the natural history of microbial plankton is no exception—are veiled by the vastness and complexity of the oceans. Time-series studies are designed to resolve seasonal and long-term patterns, sacrificing a geographical perspective. All else being equal, a study of similar scope on the basis of random sampling would be expected to observe the same variation, but not discern the patterns emerging from a time-series. Nonetheless, monthly time-series studies, such as BATS, will largely miss episodic and vertical features that are too fine to be resolved by this experimental design. In an era of global change, time-series studies are a powerful approach to discern long-term trends that might be obscured by year-to-year variability. Future studies at BATS and other sites may use pyrosequencing, and methods such as quantitative PCR and fluorescent *in situ* hybridization, to estimate the biomass of populations, to quantify distributions of rare species, to identify fine patterns in ecotype distributions and to identify genomic adaptations that permit different microorganisms to subdivide the ocean surface layer into temporal and spatial niches.

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

Rachel Parsons for technical assistance and Robert Morris for his constructive input. This research was funded by a grant from the *Marine Microbiology Initiative* of the *Gordon and Betty Moore Foundation* to SJG and NSF Microbial Observatory grants MCB-0237728 & OCE-0801991 to CAC and SJG. AHT is supported by a *Feodor Lynen Fellowship* of the *Alexander von Humboldt Foundation*.

References

- Azam F, Long RA. (2001). Oceanography—sea snow microcosms. *Nature* **414**: 495–498.
- Azam F, Malfatti F. (2007). Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5**: 782–791.
- Bano N, Hollibaugh JT. (2002). Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microbiol* **68**: 505–518.
- Béjà O, Aravind L, Koonin EV, Suzuki M, Hadd A, Nguyen LP *et al.* (2000a). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP *et al.* (2000b). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. (2008). GenBank. *Nucleic Acids Res* **36**: D25–D30.
- Brown MV, Donachie SP. (2007). Evidence for tropical endemicity in the Deltaproteobacteria Marine Group B/SAR324 bacterioplankton clade. *Aquatic Microbial Ecology* **46**: 107–115.
- Burton RM, Treusch AH, Finlay LA, Donatz M, Giovannoni SJ. (2009). BestBet. <http://www.mcb.oregonstate.edu/giovannoni/>.
- Carlson CA, Ducklow HW. (1996). Growth of bacterioplankton and consumption of dissolved organic carbon in the Sargasso Sea. *Aquat Microb Ecol* **10**: 69–85.
- Carlson CA, Ducklow HW, Michaels AF. (1994). Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature* **371**: 405–408.
- Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3**: 283–295.
- Clarke KR. (1993). Nonparametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117–143.
- Cohan FM. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos Trans R Soc Lond B Biol Sci* **361**: 1985–1996.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- DeLong EF, Taylor LT, Marsh TL, Preston CM. (1999). Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent *in situ* hybridization. *Appl Environ Microbiol* **65**: 5554–5563.

- Dortch Q, Roberts TL, Clayton JR, Ahmed SI. (1983). RNA DNA ratios and DNA concentrations as indicators of growth-rate and biomass in planktonic marine organisms. *Mar Ecol Prog Ser* **13**: 61–71.
- Ducklow HW, Doney SC, Steinberg DK. (2009). Contributions of long-term research and time-series observations to marine ecology and biogeochemistry. *Annu Rev Mar Sci* **1**: 279–302.
- Durand MD, Olson RJ, Chisholm SW. (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res Part II Top Stud Oceanogr* **48**: 1983–2003.
- Egert M, Friedrich MW. (2005). Post-amplification Klenow fragment treatment alleviates PCR bias caused by partially single-stranded amplicons. *J Microbiol Methods* **61**: 69–75.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.
- Finlay LA, Burton RM, Kujawinski EB, Giovannoni SJ. (2009). T-RFLP-adjust. <http://www.mcb.oregonstate.edu/giovannoni/>.
- Fuhrman JA, Davis AA. (1997). Widespread Archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar Ecol Prog Ser* **150**: 275–285.
- Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al.* (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774–7778.
- Gin KYH, Chisholm SW, Olson RJ. (1999). Seasonal and depth variation in microbial size spectra at the Bermuda Atlantic time series station. *Deep Sea Res Part I Oceanogr Res Pap* **46**: 1221–1245.
- Giovannoni S, Rappé M, Vergin K, Adair N. (1996). 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria. *Proc Natl Acad Sci USA* **93**: 7979–7984.
- Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al.* (2005a). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82–85.
- Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho JC *et al.* (2008). The small genome of an abundant coastal ocean methylotroph. *Environ Microbiol* **10**: 1771–1782.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin K, Batista D *et al.* (2005b). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Gonzalez JM, Simo R, Massana R, Covert JS, Casamayor EO, Pedros-Alio C *et al.* (2000). Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* **66**: 4237–4246.
- Gordon DA, Giovannoni SJ. (1996). Stratified microbial populations related to *Chlorobium* and *Fibrobacter* detected in the Atlantic and Pacific Oceans. *Appl Environ Microbiol* **62**: 1171–1177.
- Hansell DA, Carlson CA. (2001). Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: control by convective overturn. *Deep Sea Res Part II Top Stud Oceanogr* **48**: 1649–1667.
- Hewson I, Steele JA, Capone DG, Fuhrman JA. (2006). Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol Oceanogr* **51**: 1274–1283.
- Heywood JL, Zubkov MV, Tarran GA, Fuchs BM, Holligan PM. (2006). Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: Evaluation of inter-annual variability. *Deep Sea Res Part II Top Stud Oceanogr* **53**: 1530–1547.
- Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.
- Kaplan CW, Kitts CL. (2003). Variation between observed and true terminal restriction fragment length is dependent on true TRF length and purine content. *J Microbiol Methods* **54**: 121–125.
- Karner MB, DeLong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507–510.
- Kemp PF, Lee S, LaRoche J. (1993). Estimating the growth rate of slowly growing marine bacteria from RNA content. *Appl Environ Microbiol* **59**: 2594–2601.
- Kent AD, Smith DJ, Benson BJ, Triplett EW. (2003). Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. *Appl Environ Microbiol* **69**: 6768–6776.
- Koeppl A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM *et al.* (2008). Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* **105**: 2504–2509.
- Liu WT, Marsh TL, Cheng H, Forney LJ. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516–4522.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar Buchner A *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Marsh TL, Saxman P, Cole J, Tiedje J. (2000). Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Appl Environ Microbiol* **66**: 3616–3620.
- McCune B, Grace JB. (2002). Analysis of ecological communities. MjM Software.
- Michaels AF, Olson D, Sarmiento JL, Ammerman JW, Fanning K, Jahnke R *et al.* (1996). Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* **35**: 181–226.
- Moeseneder MM, Arrieta JM, Muyzer G, Winter C, Herndl GJ. (1999). Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **65**: 3518–3525.
- Moore LR, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. *Nature* **393**: 464–467.
- Morris RM, Cho JC, Rappé MS, Vergin KL, Carlson CA, Giovannoni SJ. (2005). Bacterioplankton responses to

- deep seasonal mixing in the Sargasso Sea. *Limnol Oceanogr* **50**: 382–391.
- Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.
- Morris RM, Rappe MS, Urbach E, Connon SA, Giovannoni SJ. (2004). Prevalence of the Chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl Environ Microbiol* **70**: 2836–2842.
- Murray AE, Preston CM, Massana R, Taylor LT, Blakis A, Wu K *et al.* (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl Environ Microbiol* **64**: 2585–2595.
- Murray AE, Wu KY, Moyer CL, Karl DM, DeLong EF. (1999). Evidence for circumpolar distribution of planktonic Archaea in the Southern Ocean. *Aquat Microb Ecol* **18**: 263–273.
- Nelson DM, Brzezinski MA. (1997). Diatom growth and productivity in an oligotrophic midocean gyre: A 3-yr record from the Sargasso Sea near Bermuda. *Limnol Oceanogr* **42**: 473–486.
- Nelson JD, Boehme SE, Reimers CE, Sherrell RM, Kerkhof LJ. (2008). Temporal patterns of microbial community structure in the Mid-Atlantic Bight. *FEMS Microbiol Ecol* **65**: 484–493.
- Nocker A, Burr M, Camper AK. (2007). Genotypic microbial community profiling: a critical technical review. *Microb Ecol* **54**: 276–289.
- Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA. (1990a). Spatial and temporal distributions of prochlorophyte picoplankton in the North-Atlantic Ocean. *Deep Sea Res Part a Oceanogr Res Pap* **37**: 1033–1051.
- Olson RJ, Chisholm SW, Zettler ER, Armburst EV. (1990b). Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45–58.
- Osborne CA, Rees GN, Bernstein Y, Janssen PH. (2006). New threshold and confidence estimates for terminal restriction fragment length polymorphism analysis of complex bacterial communities. *Appl Environ Microbiol* **72**: 1270–1278.
- Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al.* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.
- Reinthal T, van Aken H, Veth C, Aristegui J, Robinson C, Williams PJLR *et al.* (2006). Prokaryotic respiration and production in the meso- and bathypelagic realm of the eastern and western North Atlantic basin. *Limnol Oceanogr* **51**: 1262–1273.
- Reysenbach A-L, Giver LJ, Wickham GS, Pace NR. (1992). Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol* **58**: 3417–3418.
- Ricke P, Kolb S, Braker G. (2005). Application of a newly developed ARB software-integrated tool for in silico terminal restriction fragment length polymorphism analysis reveals the dominance of a novel *pmoA* cluster in a forest soil. *Appl Environ Microbiol* **71**: 1671–1673.
- Schlitzer R. (2000). Ocean Data View. <http://www.awi-bremerhaven.de/GEO/ODV>.
- Sprattall J, Tomczak M. (1992). Evidence of the barrier layer in the surface-layer of the Tropics. *J Geophys Res Oceans* **97**: 7305–7316.
- Steinberg DK, Carlson CA, Bates NR, Johnson RH, Michaels AF, Knap AH. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biochemistry. *Deep Sea Res II* **48**: 1405–1447.
- Suzuki MT, Beja O, Taylor LT, DeLong EF. (2001). Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ Microbiol* **3**: 323–331.
- Suzuki MT, Preston CM, Beja O, de la Torre JR, Steward GF, DeLong EF. (2004). Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48**: 473–488.
- Varela MM, van Aken HM, Herndl GJ. (2008). Abundance and activity of Chloroflexi-type SAR202 bacterioplankton in the meso- and bathypelagic waters of the (sub) tropical Atlantic. *Environ Microbiol* **10**: 1903–1911.
- Vergin KL, Rappe MS, Giovannoni SJ. (2001). Streamlined method to analyze 16S rRNA gene clone libraries. *Biotechniques* **30**: 938–940, 943–934.
- Wright TD, Vergin KL, Boyd PW, Giovannoni SJ. (1997). A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63**: 1441–1448.
- Wuchter C, Abbas B, Coolen MJ, Herfort L, van Bleijswijk J, Timmers P *et al.* (2006). Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* **103**: 12317–12322.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)