Biogeosciences

Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures

C. J. Friedline[1], R. B. Franklin[2], S. L. McCallister[2,3], and M. C. Rivera[1,2]

[1]Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA, USA
[2]Department of Biology, Virginia Commonwealth University, Richmond, VA, USA
[3]Center for Environmental Studies, Virginia Commonwealth University, Richmond, VA, USA

Correspondence to: M. C. Rivera (mcrivera@vcu.edu)

Abstract. Microbial communities are recognized as major drivers of the biogeochemical processes in the oceans. However, the genetic diversity and composition of those communities is poorly understood. The aim of this study is to investigate the composition of bacterial assemblages in three different water layer habitats: surface (2–20 m), deep chlorophyll maximum (DCM; 28–90 m), and deep (100–4600 m) at nine stations along the eastern Atlantic Ocean from 42.8° N to 23.7° S. The sampling of three discrete, predefined habitat types from different depths, Longhurstian provinces, and geographical locations allowed us to investigate whether marine bacterial assemblages show spatial variation and to determine if the observed spatial variation is influenced by current environmental conditions, historical/geographical contingencies, or both. The PCR amplicons of the V6 region of the 16S rRNA from 16 microbial assemblages were pyrosequenced, generating a total of 352 029 sequences; after quality filtering and processing, 257 260 sequences were clustered into 2871 normalized operational taxonomic units (OTU) using a definition of 97 % sequence identity. Community ecology statistical analyses demonstrate that the eastern Atlantic Ocean bacterial assemblages are vertically stratified and associated with water layers characterized by unique environmental signals (e.g., temperature, salinity, and nutrients). Genetic compositions of bacterial assemblages from the same water layer are more similar to each other than to assemblages from different water layers. The observed clustering of samples by water layer allows us to conclude that contemporary environments are influencing the observed biogeographic patterns. Moreover, the implementation of a novel Bayesian inference approach that allows a more efficient and explicit use of all the OTU abundance data shows a distance effect suggesting the influence of historical contingencies on the composition of bacterial assemblages. Surface bacterial communities displayed a general congruency with the ecological provinces as defined by Longhurst with modest exceptions usually associated with unique hydrographic and biogeochemical features. Collectively, our findings suggest that vertical (habitat) and latitudinal (distance) biogeographic signatures are present and that both environmental parameters and ecological provinces drive the composition of bacterial assemblages in the eastern Atlantic Ocean.

1 Introduction

Prokaryotes are responsible for cycling the organic and inorganic compounds essential for life, and are the main drivers of global biogeochemistry. Our knowledge of the microbial loop and its significance in the carbon and nutrient cycles of the ocean has grown exponentially over the last 35 years. The introduction of culture-free techniques for enumerating bacteria (Hobbie et al., 1977), coupled with the increased sample throughput provided by flow cytometry (Gasol and del Giorgio, 2000), have revealed some significant roles prokaryotes play, accounting for the majority of total biomass and respiration in the ocean as well as the majority of primary production on earth (Fuhrman and Azam, 1982; Azam et al., 1983; Fuhrman and Campbell, 1998; del Giorgio and Duarte, 2002). Although prokaryotes are ubiquitous and abundant in marine ecosystems, relatively little is known about the diversity and composition of the complex

microbial communities found there. In particular, it remains unclear what the principle drivers are that control the distribution of marine microbes, or how this diversity may influence the biogeochemical functioning of marine ecosystems (Anderson et al., 2010). Likewise, our understanding of how microbial diversity varies over time and space is lacking, hampering our ability to forecast potential alterations in microbial community structure and ecological function that may result from global climate change.

Methodological constraints have historically limited our ability to study microbial diversity, and any biogeochemical implications it may have, but these restrictions have lessened with the development of genomic technologies. In particular, the coupling of polymerase chain reaction (PCR) amplification and high-throughput tag sequencing has become a popular tool to survey microbial communities (Huse et al., 2008; Giovannoni and Stingl, 2005; Massana et al., 2008). More recently, next-generation sequencing technology, such as 454 pyrosequencing, has been applied, yielding amounts of data in orders of magnitude greater than conventional sequencing approaches (Huse et al., 2008; Kirchman et al., 2010; Agogue et al., 2011). These molecular techniques often make use of the gene that encodes the small subunit ribosomal RNA (16S rRNA) as a phylogenetic marker. However, there are currently no clear rules for the depth of phylogenetic relationships that define taxonomic ranks with this system; as a result, efforts to study microbial diversity are further complicated by the lack of an acceptable classification scheme for defining diversity units (Cases and de Lorenzo, 2002). Microbiologists proceed by using operational taxonomic units (OTUs) in which a pre-defined level of sequence identity is necessary for organisms to be classified as distinct taxa.

During the past decade, the application of molecular techniques to survey ocean microbial communities has become quite popular, with most research efforts focused on cataloging microbial diversity (Breitbart et al., 2002; Venter et al., 2004; Rusch et al., 2007) or documenting how specific environmental conditions influence the distribution of selected taxa (Galand et al., 2009; Hewson et al., 2009; Agogue et al., 2011). As a result, we know that prokaryotic diversity is high in the ocean, that these communities tend to be dominated by a few abundant taxa, and that the communities show high richness in rare species (Breitbart et al., 2002; Venter et al., 2004; Sogin et al., 2006; Huber et al., 2007; Gilbert et al., 2009; Roesch et al., 2007). As our techniques improve, and our inventory of microbial taxa grows, the next challenge for ocean microbial ecologists is to place this diversity in broader ecological contexts. Very little is known about the distribution of bacteria in the ocean as it relates to physicochemical parameters or ocean biogeochemistry. Moreover, what little information we do have has been largely region-specific (Hewson et al., 2006; Galand et al., 2009; Yokokawa et al., 2010), limiting both our understanding of the factors that structure ocean bacterial communities across biomes as well as our ability to study biogeographic patterns.

The purpose of the research presented here was to utilize high-throughput pyrosequencing technology to explore the diversity of bacterial assemblages in the eastern Atlantic Ocean, while simultaneously considering whether these communities exhibited biogeographic patterns at large spatial scales. Ecologists recognize the observed biogeographical patterns of community composition can be the result of at least three different not mutually exclusive mechanisms: (1) environmentally-defined features or niche-based community processes, (2) stochastic process like ecological drift, dispersal and speciation, and (3) the effects of historical contingencies (Ricklefs and Schluter, 1994; Nekola and White, 1999; Hubbell, 2001; Chave, 2004). Historical events, including dispersal limitations and ancestral environmental conditions, define different biological provinces. This biological provincialism is manifested by the observation of distance decay in community similarity after decoupling the distance from current environmental effects (Martiny et al., 2006a). Samples were collected along a 7700 km meridional transect, and the relative importance of environmental conditions versus spatial separation was evaluated using both ecological statistics and a novel application of Bayesian hypothesis testing. The sampling of three discrete environments across several geographic locations or provinces, combined with the finer resolution provided by the Bayesian approach, allowed us to uncover the potential influence of both contemporary environmental features and historical effects (provincialism) on the biogeographic patterns of the bacterial communities of the eastern Atlantic Ocean.

2 Materials and methods

2.1 Study area and water sample collection

Sampling was conducted in November 2008 on a meridional transect from 50.2° N to 31.4° S during the cruise ANT XXV/1 on the RV *Polarstern* as it traveled from Bremerhaven (Germany) to Cape Town (South Africa). Considering Longhurst's conceptual model, which partitioned the ocean into biogeochemical provinces based on thermohaline properties, remotely-sensed data of chlorophyll concentrations, nutrient fields, and seasonal changes in the mixed-layer depth (Longhurst, 1998), we collected samples from six major provinces: North Atlantic Subtropical East (NASE), North Atlantic Tropical Gyral (NATR), Western Tropical Atlantic (WTRA), Eastern Tropical Atlantic (ETRA), South Atlantic Gyral Province (SATL), and Benguela Current Coastal (BENG). Water samples (∼50 l) were collected for bacterial community analysis from nine different stations along the transect; at selected stations, water was collected from multiple depths yielding a total of sixteen samples for pyrosequencing (Fig. 1). Surface water samples were collected from a Teflon "Fish" sampler fixed alongside the ship; water
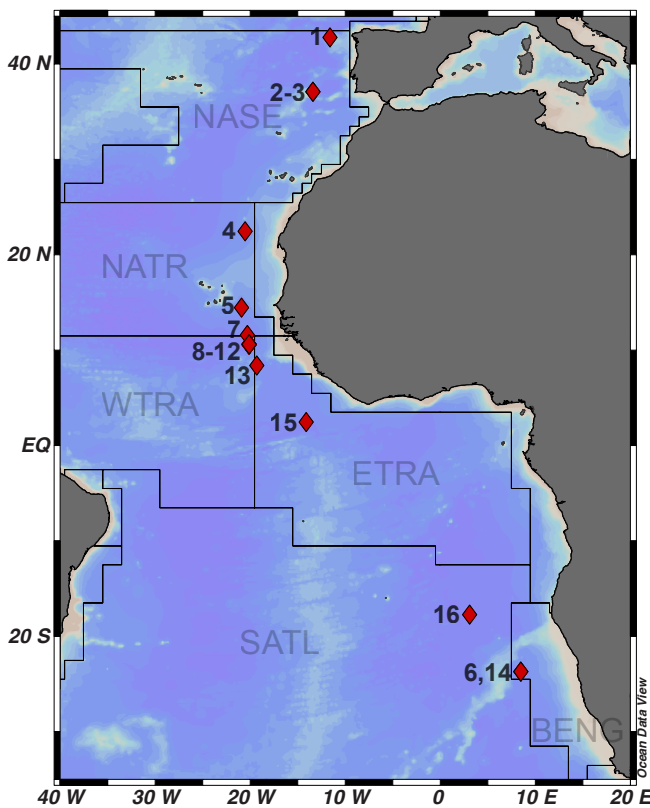
Fig. 1. Collection locations of 16 water samples analysed in this study. Figure created with Ocean Data View (Schlitzer, 2011) with shape file overlay from VLIZ (2009).

from depth was obtained from a rosette sampler connected to Conductivity/Temperature/Depth (CTD) instrumentation.

2.2 Oceanographic variables and nutrient concentrations

A Seabird 911 Plus CTD with a WET Labs ECO-FL fluorometer was used to record the water temperature, salinity, and chlorophyll *a* fluorescence (Chl *a*), associated with each bacterial sample. In addition, samples for determination of dissolved nutrients were collected concomitantly and processed according to standard methods for seawater analysis (see also Koch and Kattner, 2012). Available data included dissolved organic carbon and nitrogen (DOC, DON), nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), phosphate (PO_4^{3-}), and dissolved silicate (Si) concentrations.

2.3 Preparation of bacterial community samples

Samples for molecular analysis of bacterial community composition were obtained by sequential filtration using 142 mm diameter Isopore™ polycarbonate membranes (Millipore, Billerica, MA, USA). Water was first passed through a 3 μm pore-size filter (Millipore TSTP 14250), to remove the eu-

karyotic fraction of the community, and then through a 0.2 μm pore-size filter (Millipore GTTP 14250) to concentrate the prokaryotic biomass. Each filter was immediately placed in a sterile polyethylene sample bag with 10 ml of filter-sterilized TENS buffer (50 mM Tris-HCL (pH 8.0), 20 mM EDTA, 400 mM NaCl, and 0.75 M sucrose) and stored at −80 °C as per Rusch et al. (2007). Frozen samples were subsequently shipped in a dry shipper filled with liquid nitrogen and then returned to −80 °C freezer until DNA extraction could be performed (within 1 month). Total DNA was extracted from the 0.2 μm filters using the Mo-Bio UltraClean® Water DNA Isolation Kit (Cat #14800-NF, Carlsbad, CA, USA). DNA concentrations were determined using a Nanodrop 8000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). DNA extracts were dried and shipped to the International Census of Marine Microbes (ICoMM) program at Marine Biological Laboratory (MBL) in Woods Hole, MA (USA) for 454 pyrosequencing.

2.4 PCR amplification and sequencing of the 16S rRNA gene fragments

To analyze community diversity, bacteria-specific primers complementary to the hypervariable region 6 (V6) of the 16S rRNA gene were used to generate PCR amplicons using a combination of five forward and four reverse primers (Huber et al., 2007; Huse et al., 2008, 2010). Three independent PCR reactions were performed for each sample; the products were combined and analyzed using standard MBL protocols on 454 GS-FLX sequencer (454 Life Sciences, Branford CT, USA). The sequence data generated from the present study are available via the Visualization and Analysis of Microbial Population Structures (VAMPS) web interface (http://vamps.mbl.edu), identified as ICM_AOT_Bv6.

2.5 Sequence data processing

Trimmed Fasta sequences containing neither sequencing primer nor multiplexing barcode tag were downloaded from the VAMPS website. The initial trimming, performed at MBL (Huse et al., 2007), removed suspected low-quality reads containing unexpected sequencing tags, primers, or ambiguous bases, and any sequences that were less than 50 nucleotides in length after trimming. We further filtered the reads which contained suspected homopolymers ($n > 4$) and were longer than 75 bases, to increase the sample classification accuracy by reducing the effect of sequencing errors and PCR-generated chimeras. The sequence reads were clustered into OTUs using the tools implemented in mothur (Schloss et al., 2009).

The sequencing reads were first classified using the mothur implementation of the Ribosomal Database Project (RDP) classifier (Wang et al., 2007), and all reads that were not classified as Bacteria were excluded. The remaining sequence reads were reduced to a unique set of sequences by collapsing

to one sequence all the sets of identical sequences. The unique sequences were aligned using mothur (Schloss, 2009) to a V6-specific, curated, pre-aligned database derived from the SILVA alignment. The full database (50 000 sites) is distributed from www.mothur.org and contains 14 956 sequences while our database (2985 sites) contains 13 275 bacterial sequences, following removal of any sequence containing ambiguous bases. Briefly, BLAST (Altschul et al., 1990) was used to determine the boundaries of the V6 region within the full 16S rRNA SILVA alignment, and the region was extracted to generate a multiple sequence alignment containing all the unique sequences with the gap pattern of the original SILVA alignment. To further reduce sequencing noise, the reads were pre-clustered such that any set of sequences with a single nucleotide change were considered as equivalent (Huse et al., 2010). As a result of the pre-clustering step, reads that were found only once in the sample set (singletons) were removed from the analysis. From this non-singleton alignment, a pairwise distance matrix was created (treating multiple gaps as one) and used to construct OTUs using average-linkage clustering at 97 % identity, which loosely translates as species-level separation. Because initially every sequenced read had been assigned a taxonomic classification, each OTU taxonomic classification is the consensus derived from the classification of the individual reads comprising the OTU (50 % bootstrap support over 1000 iterations, similar to Claesson et al., 2009). The abundances of the resulting OTUs were normalized using the smallest sample number ($n = 6687$). Those OTUs that, following normalization, had zero abundance in all samples ($n = 555$), were removed from further analysis; the total number of reads left after normalization was 106 613, clustered into 2871 OTUs.

2.6 Ecological statistics

Richness estimates for each community were obtained using CatchAll version 3.0 (Bunge, 2011). The CatchAll parametric estimates ("Best Model") are reported along with traditional non-parametric species richness estimates, ACE (Chao and Lee, 1992) and Chao1 (Chao, 1984). The CatchAll parametric estimator is particularly attractive in this setting because it tends to avoid underestimation common with ACE and Chao1 when diversity is high, and is robust to outliers (Bunge, 2011). Rarefaction curves were also generated using mothur to assess the degree to which sampling effort was saturated.

Sequencing results were analyzed using principal coordinate analysis. PCoA was conducted using the Bray-Curtis index of similarity to group samples based on the normalized abundance data. The first two coordinates were plotted as a means of visualizing the relative similarity in community composition across samples. Analysis of similarity (ANOSIM) (Clarke, 1993) was used to test whether groups were different using the Bray-Curtis index and 10 000 permutations. Spearman correlation analysis was used to com-

pare the PCoA coordinates to OTU abundance to determine taxonomic drivers of the PCoA separation. Similarly, a correlation analysis was performed to determine the environmental parameters best linked to community separation in PCoA space. The environmental variables tested were depth, salinity, temperature, Chl *a*, DOC, DON, NO_3^-, NO_2^-, NH_4^+, PO_4^{3-}, and Si. A series of Mantel (Rossi, 1996) and partial Mantel tests (Smouse et al., 1986) were used to examine the relationship between bacterial community structure, environmental conditions, and geographic separation. Specifically, we compared the following matrices: community composition as Bray-Curtis similarity based on OTU abundance, spatial separation as surface distance in km (Stott, 2011), and environmental similarity calculated using Gower's coefficient (Gower, 1971). The environmental matrix included depth (log transformed), salinity, temperature, Chl *a*, and dissolved nutrients. When necessary, similarity matrices were transformed to dissimilarity matrices as: Dissimilarity = 1 − Similarity. Because environmental measurements were not made for sample 8, it was excluded in any data manipulations that included environmental properties. The significance of the Mantel and partial Mantel test results was determined via permutation using 10 000 iterations. All ecological statistics were performed using the PAST software package version 2.10 (Hammer and Harper, 2001).

2.7 Bayesian inference

In addition to the ecological statistics described above, we also employed a Bayesian inference statistical framework to examine the bacterial community dataset. The input used for Bayesian inference, as in the classical community ecology approach, was the highly-dimensional contingency table containing the distribution of read abundances, in which the columns correspond to each one of the 2871 OTUs identified and the rows to the 16 sampling locations. In order to correct for unequal sampling in both approaches, the OTU abundances were normalized to the smallest sample size as described by mothur (Schloss et al., 2009). Experimental measurements, like OTU abundance, that generate an array of named characters for each sample studied, can be classified as character-type data and represented as a data matrix; when analyzed in a multivariate framework, each one of the OTUs is essentially treated as a phenotypic character (i.e., trait) of the community.

The classical approach to the analysis of this type of contingency table for community comparisons requires the computation of a measure of resemblance and the generation of a distance matrix (e.g., Bray-Curtis dissimilarity between all pairs of samples) from the data, followed by the application of numerical methods such as cluster analysis and representation as a dendrogram. This methodology is one of the many numerical techniques and concepts developed by Sokal and Sneath (1963, 1973) for classification based on phenetics principles. More recently, Legendre and Legendre

(1998) apply the same techniques and principles to ecological groupings.

Our approach conceptually follows the general principles of numerical classification, but applies a Bayesian hypothesis-testing framework instead of a distance-based clustering algorithm to generate the dendrogram (i.e., tree) best summarizing the relationships among the samples. Instead of reducing the contingency table to a distance matrix prior to the generation of the dendrogram, the Bayesian approach makes an explicit and more efficient use of all OTU abundance data by operating directly on each one of the columns of the contingency table. In practice, Bayesian inference methods directly consider all of the OTU data when generating a tree topology because they use the product of the likelihoods of all the OTUs and the summation over all possible tree topologies. By using standard statistical methodology within a probabilistic model of changes, Bayesian inference requires the following assumptions: (1) the properties of a community can be modeled as a collection of independent and identically distributed (i.i.d.) OTU abundances, (2) the abundance pattern for any OTU is independent of the pattern for any other OTU, and (3) the distribution and abundance pattern of all OTUs contain information about the relationships of the sampled communities. In this analysis, the null hypothesis is that the ocean contains a single random bacterial assemblage and that, in the absence of environmental disturbing forces, all samples are assumed to have the same community structure. The effect of the environmental disturbing forces on this random bacterial assemblage will result in the differentiation into distinct bacterial assemblages from the common, ancestral bacterial assemblage. Bayesian inference was then used to estimate the unrooted tree topology that best describes the relationships among the bacterial assemblages. In our case, the relationship between 16 samples is described by one of 2.13×10^{14} alternative tree topology hypotheses. The relationships of the samples illustrated by the tree topology reflect only similarity in the composition and abundance of OTUs and no direct molecular evolutionary relationships are assumed, because no molecular sequence evolution model was employed. Bayesian tree inference, as implemented in MrBayes, can reconstruct tree topologies using four different types of data: nucleotides, proteins, restriction enzymes, and standard morphology character data. Any type of analysis using OTU abundance data to determine relationships between communities implicitly or explicitly assumes that the OTU distribution data provides information that uniquely characterize the community, and we assume that this can be equated with phenotypic or morphological character data. We have modeled the OTU abundance data using the model for standard morphology data type in MrBayes, which allows one to describe the characters utilizing up to ten discrete character states (0–9). Character data is defined as information about the attributes of the objects under study and those characters can be visualized as a set of independent variables existing in a set of mutually exclusive discrete character states. In our case, each OTU is considered a character and its abundance represents one of the many different character states the data can assume. The OTU abundance values across our 16 samples ranged from zero to thousands, similar to morphology or character data with a large variance. Because MrBayes input can accommodate up to 10 character states, the OTU abundance is converted to a score between 0 and 9 by range-standardizing the normalized abundances according to Eq. (1) as in Thiele (1993) and Schols et al. (2004).

$$A_{OS_{\text{coded}}} = \text{round}\left[\frac{9 \times \left(A_{OS_{\text{norm}}} - \min[A_{O_{\text{norm}}}]\right)}{\text{range}[A_{O_{\text{norm}}}]} \right]. \quad (1)$$

In this formula, $A_{OS_{\text{coded}}}$ represents the normalized abundance value of OTU O in sample S range-standardized to a discrete value in the range [0, 9]. Min[$A_{O_{\text{norm}}}$] denotes the minimum normalized OTU O abundance value in the matrix across all 16 samples and range[$A_{O_{\text{norm}}}$] represents the value of the range of the normalized OTU abundances across all 16 samples. The 10 transformed values represent character states that are compatible with the standard model for discrete morphology data as implemented by MrBayes 3 (Huelsenbeck and Ronquist, 2001). However, because OTUs that are distributed only among a few states across all samples will be treated equally, we need to account for variations in character state changes of different magnitude. For example, an OTU distributed among samples with all 0 and 1 will be treated equally if that same OTU was distributed with those 1 changed to 9. To do this, we converted each integer value to its four-bit binary equivalent, in effect quadrupling the number of matrix columns (e.g., 9 was converted to 1001, and 1 to 0001). Using MrBayes, topologies were reconstructed from the range-standardized/binary-encoded matrix using the default parameters for the standard 0–9 character states morphology model with the following assumptions: (1) equal state frequencies, (2) across-sites rate variation following a gamma distribution, (3) all sites are informative and unordered, and (4) 10 million iterations (Huelsenbeck and Ronquist, 2001). Trace plots runs from the MrBayes runs were analyzed for mixing using Tracer 1.5 (http://tree.bio.ed.ac.uk/software/tracer) and the model fit of gamma for across-site rate variation was evaluated using a Bayes factors (Kass and Raftery, 1995). Finally, considerable testing, simulations, and data randomizations were performed prior to the application of Bayesian inference to this data set, and a formal proof and extension to the approach will be discussed elsewhere (Friedline and Rivera, 2012).

3 Results and discussion

In this study, we report the vertical and latitudinal distribution and abundance of bacteria along an eastern Atlantic Ocean transect (Fig. 1). Specifically, our goals were to determine: (1) if the bacterial communities from the surface,
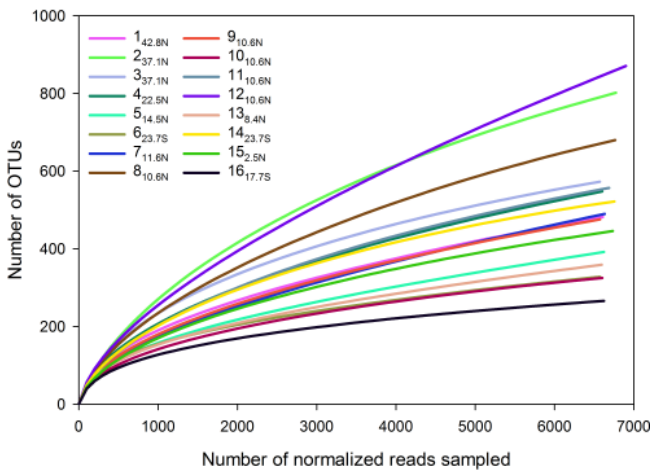
Fig. 2. Rarefaction curves for the 16 samples from the eastern Atlantic Ocean. The curves were obtained by resampling the normalized NS sequence reads using the *rarefaction.single* function from mothur, and plotted as the number of operational taxonomic units (OTUs) versus the number of normalized sequence reads sampled. An OTU is composed of a set of sequences with a similarity equal to or greater than 97 %. Each curve corresponds to a different sample (1–16) with the latitude of the sampling location as a subscript in the legend. Samples assignments to the DCM, photic, and deep zones are shown in Fig. 5.

DCM, and deep water layers were significantly different, (2) if the composition of those communities show a pattern of decreasing similarity with geographical distance, and (3) if the numerous physical processes defining the different ecological provinces may also influence the structure of the bacterial communities. In addition, because the seaward boundaries of Longhurst provinces were set at approximately 320–640 km from shore, and our cruise track was 480–720 km off shore, we anticipated an effect from the coastal waters of the Canary Current upwelling system (Taylor et al., 2011). The Canary Current ecosystem (12–43° N) borders NASE, NATR, WTRA and ETRA, and consists of complex hydrographic features that contribute to unique biogeographic subregions (Aristegui et al., 2009a). The association and potential entrainment of bacterial assemblages within these distinct hydrographic features may also have an impact on community composition and diversity. Further, we used the depth of the pycnocline, the euphotic zone, and the DCM as indicators of the physical state of the water column and considered whether communities from equivalent vertical zones were influenced by similar physical processes. Our study takes advantage of the latest sequencing technologies to understand the phylogenetic composition and structure of those bacterial communities, and then places the information in a biogeographic as well as a biogeochemical context.

3.1 Sequencing statistics and diversity estimates

Pyrosequencing generated 352 029 raw sequence reads ($22\,002 \pm 9959$ reads/sample); summary statistics are shown in Table 1. After filtering out low-quality reads and removing sequence reads present only once (singletons), a total of 257 260 non-singleton (NS) reads remained. It has been observed that sequencing and base-calling errors are potentially responsible for singletons, which can artificially inflate diversity estimates (Quince et al., 2008; Reeder and Knight, 2009; Kunin et al., 2010). After average-linkage clustering, 3426 OTUs were identified, using an average 97 % sequence identity per OTU.

In order to correct for the broad range in sample sizes, the OTU abundances were normalized to the smallest sample size, resulting in a total of 2871 OTUs. Overall, we find an average of $16\,079 \pm 7374$ NS reads per sample, with a range between 6687 and 36 569 reads (samples 11 and 9, respectively). The reads have an average length of 62 ± 3 bases and GC content of $47\,\% \pm 5\,\%$ When all filtered reads are considered, we find that of the 628 ± 270 unique reads/sample, 408 ± 163 reads/sample are found only once per sample (i.e., singletons). The estimates of sample richness vary across a wide range (Table 1), the highest of which come from CatchAll (1979 ± 1737). Within the CatchAll richness estimates, sample 16 had the lowest estimate (521 OTUs) while sample 12 had the highest (7295 OTUs). Comparison of the rarefaction curves (Fig. 2) shows that some of the bacterial communities have reached an asymptote, indicating they were completely sampled. Even given normalization, our rarefaction curves indicate that there are still several communities for which more sampling may be needed in order to adequately assess community composition. Analysis of the rarefaction curves suggests that some of the deep-water communities (e.g., 2, 8, and 12) are more diverse than the rest of the communities, as previously shown for the bathypelagic bacterial communities of the North Atlantic (Agogue et al., 2011).

3.2 Bacterial community composition

Analysis of the normalized abundance of OTUs across all communities reveals that nearly half of the reads are distributed among only the top 25 OTUs (Table 2). These 25 OTUs represent less than 1 % of the total richness. This suggests that our communities are structured around a relatively small number of very abundant OTUs and, possibly, a large number of sparsely represented OTUs. This conclusion is further supported by the fact that 35 % of the total read abundance is represented by the top 10 most abundant OTUs (0.3 %), 75 % by the top 86 OTUs (3 %), and 95 % by the top 589 OTUs (20 %). A total of 35 OTUs are common to all 16 samples, representing 52 % of the total read abundance. Of these ubiquitous OTUs, many are prominent marine organisms, including six OTUs affiliated
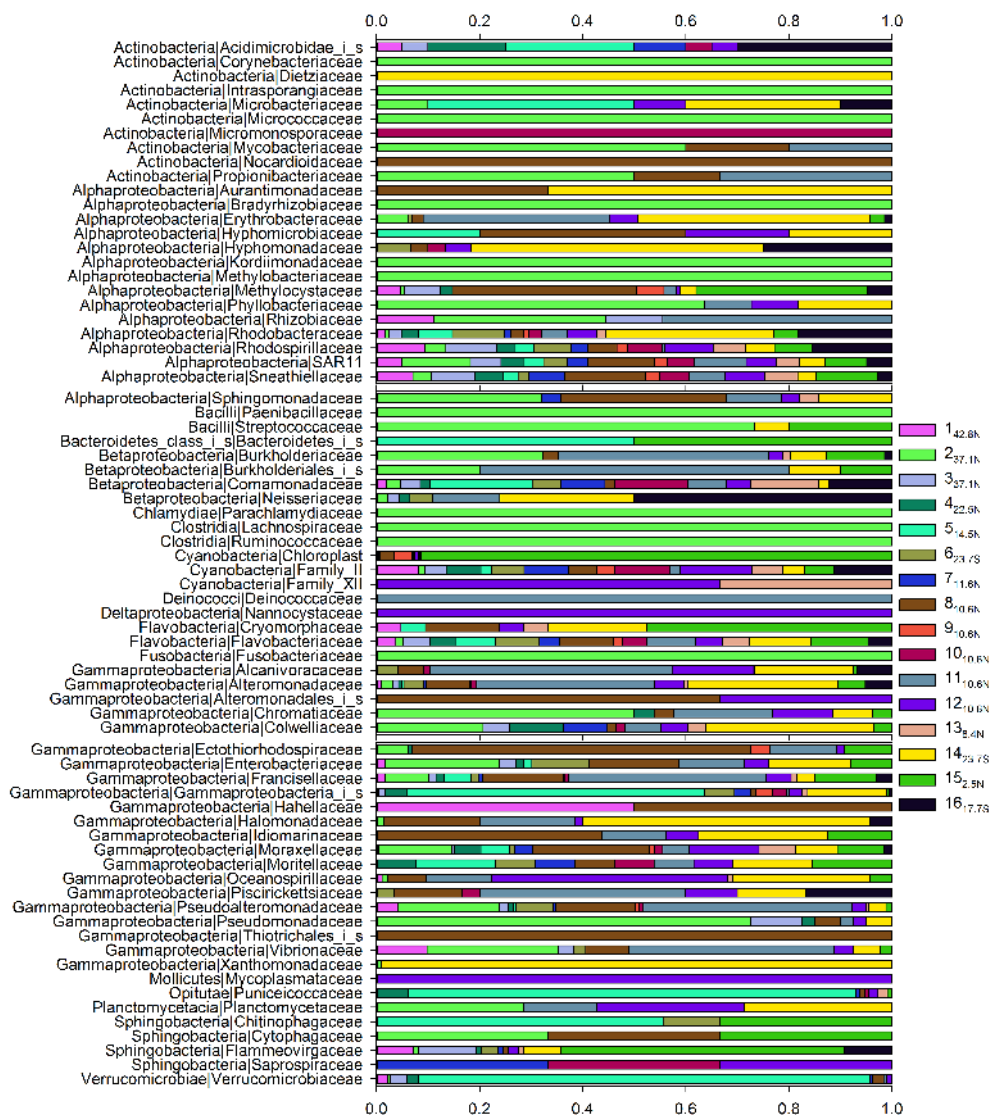
Fig. 3. Relative abundance and affiliation of the 72 bacterial taxonomic families identified in this study with >50 % RDP classifier bootstrap support. A total of 962 OTUs (34 %) were assigned taxonomy to the family level, representing 58 % of the total read abundance. The size of the bars represents the proportion of the particular family in a sample. *Incertae sedis* is denoted as "i_s". Each color corresponds to a different sample (1–16) with the latitude of the sampling location as a subscript in the legend.

with the Alphaproteobacteria SAR11 clade (22 % of the total read abundance), four from the Cyanobacteria Family II (8 %), and one identified as *Alteromonas* (3 %). Previous global ocean surveys have identified *Pelagibacter* (SAR11) and *Prochlorococcus* (Family II) as the dominant clades in surface waters (Giovannoni and Stingl, 2005; Giovannoni et al., 2005; Rusch et al., 2007; Gilbert et al., 2012).

Analysis of the distribution range of the 25 most abundant OTUs immediately suggests a significant difference in OTU abundance across the three different habitats (Table 2). Differences in the abundance of OTUs associated with the same taxonomic level across the three different habitats appear to support the ecotype concept (Cohan and Perry, 2007; Cohan,

2002), which predicts the existence of multiple OTUs from the same taxonomic groups adapted to local ecological conditions. For example OTUs 31 and 24, both taxonomically affiliated to *Pelagibacter*, appear to show different habitat preferences, as shown in the top two rows of Table 2. These differences appear even more dramatic when we consider the relative, per-sample contribution of the 72 bacterial taxonomic families identified, thus revealing taxa-specific distributions across sampling locations (Fig. 3). Although taxonomic families like the Alphaproteobacteria SAR11 and Cyanobacteria Family II clades are found across all samples, others like some of the Gammaproteobacteria families, were identified only in the samples from deep waters. It is

Table 1. Sequencing statistics and OTU richness estimates for 16 bacterial communities along the eastern Atlantic Ocean. The statistical estimates of the total OTU richness of each sampled community were determined from the observed OTUs using the CatchAll parametric estimator and the non-parametric estimators Chao1 and ACE; all implemented in CatchAll V3.0. NS = Non-singletons, the remaining reads after the removal of reads present only once (singletons). SD = standard deviation. SE = standard error.

	Mean/sample	SD	Range
Total NS reads	16 079	7374	6687–36 569
NS read length	62	3	51–74
NS GC %	47	5	18–67
Total unique reads	628	270	225–1271
Total singleton reads	408	163	182–864
Total OTUs at 3 %	721	193	446–1157
	Mean/sample (SD)	SE (SD)	Range
CatchAll	1979 (1737)	508 (738)	521–7295
Chao1	994 (503)	91 (62)	432–2619
ACE	1400 (933)	155 (117)	466–4507

Table 2. Taxonomic affiliation of the 25 most abundant operational taxonomic units (OTUs) across all 16 Atlantic Ocean bacterial communities sampled. The table shows the total normalized OTU abundance (n) as well as the mean and standard deviation according to habitat. The number of samples in each zone are 4, 7, and 5 for surface, DCM, and deep, respectively. The numerical identifier of each OTU is shown in parentheses in the first column to differentiate OTUs with identical taxonomic labels.

Taxa (OTU)	n	Surface	DCM	Deep
Pelagibacter (31)	9615	907 (133)	574 (130)	393 (213)
Pelagibacter (24)	9188	201 (11)	676 (170)	730 (381)
GpIIa (282)	4606	632 (423)	187 (283)	153 (295)
Alteromonas (2602)	2987	58 (32)	55 (69)	474 (415)
Proteobacteria (3151)	2505	196 (49)	205 (117)	58 (83)
GpIIa (2949)	2256	2 (2)	281 (274)	56 (48)
Proteobacteria (928)	1621	0 (1)	81 (159)	210 (119)
Bacteria (273)	1460	0 (1)	193 (208)	21 (26)
Pelagibacter (63)	1439	5 (2)	169 (65)	48 (52)
Gammaproteobacteria (367)	1417	54 (17)	145 (103)	36 (16)
Marinobacter (467)	1368	1 (1)	69 (139)	177 (175)
Proteobacteria (3026)	1357	138 (92)	87 (72)	39 (73)
Pelagibacter (127)	1335	263 (47)	26 (19)	20 (38)
Pelagibacter (45)	1297	66 (22)	130 (82)	25 (23)
Chlorophyta (1015)	1239	0 (0)	173 (380)	6 (10)
Rhodobacteraceae (110)	1187	82 (57)	105 (102)	25 (28)
GpIIa (286)	1164	37 (11)	125 (259)	28 (25)
Flavobacteria (3178)	1118	78 (33)	67 (24)	67 (46)
Actinobacteria (496)	1043	0 (0)	22 (21)	178 (122)
Alphaproteobacteria (41)	1041	70 (17)	90 (44)	26 (24)
Alcanivorax (206)	978	8 (12)	27 (44)	152 (134)
Pseudoalteromonas (1556)	976	14 (4)	37 (46)	133 (101)
Thalassobius (113)	974	54 (12)	97 (132)	17 (23)
Bacteria (179)	971	155 (88)	26 (14)	34 (60)
Alphaproteobacteria (168)	964	116 (27)	54 (16)	24 (36)
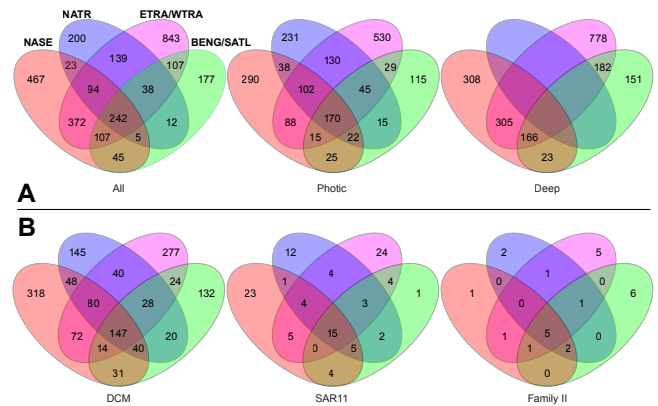
Fig. 4. Venn diagrams comparing the OTU richness identified in the Longhurst provinces sampled in this study. Panel A compares OTU richness identified in all the samples (All, $n = 2871$), samples from the photic zone (Photic, $n = 1845$), and samples from deep waters (Deep, $n = 1913$). Panel B compares OTU richness identified in all the deep chlorophyll maximum zone samples (DCM, $n = 1416$). Also panel B compares the OTU richness identified in the DCM zone for two of the most abundant taxonomic families, SAR11/*Pelagibacter* ($n = 107$) and Family II/Cyanobacteria ($n = 25$). The studied provinces are grouped and labelled in the diagram for all samples (ALL) as follows: North Atlantic Subtropical East (NASE, 3 samples), North Atlantic Tropical Gyral (NATR, 2 samples), Eastern and Western Tropical Atlantic (ETRA/WTRA; ETRA, 2 samples; WTRA, 8 samples), and South Atlantic Gyral and Benguela Current Coastal (BENG/SATL; BENG, 3 samples; SATL, 1 sample). The photic diagram compares two samples each for NASE, NATR, and BENG/SATL provinces and five samples for the ETRA/WTRA provinces. The deep diagram compares one sample each for NASE and BENG/SATL provinces and three samples for ETRA/WTRA provinces. The DCM, SAR11 and Family II diagrams compare one sample for the NATR province and two samples for each one of the other provinces. Samples from the DCM, photic, and deep zones were defined as in Fig. 5.

some of those unique taxa are endemic to the Mediterranean communities and were transported into the Atlantic by the Mediterranean outflow waters.

These observations suggest a differential composition of the bacterial communities, dependent not only on depth (habitat) but also on geographical location. To further illustrate these observations, we used Venn diagrams to compare the distribution of OTUs across the three water layers (surface, DCM, and deep) and the four ecological provinces (Fig. 4). Of the 2871 OTUs identified, 1942 are found in the WTRA/ETRA provinces, 1355 in the NASE province, 753 in the NATR province, and 733 in the BENG/SATL provinces. Samples from the photic zone (surface and DCM) have a total richness of 1845 OTUs, while the deep-water samples have a total richness of 1913. Consistently, the highest richness is found in WTRA and ETRA provinces, which agrees with previous research that has demonstrated that marine prokaryotes exhibit a latitudinal gradient of increasing

particularly interesting that there are several taxonomic families that are found exclusively in sample 2 (NASE province and Eurafrican Mediterranean water mass). It is possible that

diversity toward the equator (Fuhrman et al., 2008). Overall, the Venn diagrams demonstrate that there are province- and depth-specific OTUs, revealing both habitat and geographical (spatial) signal in the genetic composition of the communities.

3.2.1 Surface communities

The surface layer of the ocean has high exposure to solar radiation, and is very low in inorganic nutrients such as phosphorous and nitrogen; some researchers have even classified this environment as extreme because it is so oligotrophic (Treusch et al., 2009). Further, surface waters influenced by particular wind bands may experience significant wind-induced turbulence and atmospheric deposition of aerosols, pollutants and Aeolian dust (Pohl et al., 2011; Xie et al., 2011). Analysis of the surface samples (i.e., those with a depth of 20 m or less) shows that over 75 % of the abundance is dominated by only five taxonomic groups: the Alphaproteobacteria clade SAR 11 (26 % ± 3 %), the Cyanobacteria clade GpIIa (10 % ± 5 %), the Gammaproteobacteria (10 % ± 2 %), other Proteobacteria (14 % ± 1 %), the Bacteroidetes (9 % ± 2 %) and the Verrucomicrobia (3 % ± 5 %). Similar patterns of dominant, surface-water taxa and community composition shifts related to diatom blooms were recently observed during a six-year time series in the English Channel (Gilbert et al., 2012). In sample 5, from a bloom encountered along the transect, the Cyanobacteria comprised only 2 % of the abundance, while the Verrucomicrobia, the Gammaproteobacteria, the Bacteroidetes and the Actinobacteria are over-represented and comprise 37 % of the total abundance. The altered community composition of this diatom-dominated bloom agrees with previous observations (Carlson et al., 2002; West et al., 2008; Gilbert et al., 2012) suggesting that the phytoplankton bloom may play an important role in determining the structure and composition of the bacterial community.

3.2.2 DCM communities

The DCM is the region in the ocean water column with the highest concentration of Chl a, which forms when the simultaneous availability of high concentrations of inorganic nutrients and appropriate intensity and wavelength of light generate the optimal conditions for phytoplankton development. Seven communities from the DCM layer were analyzed in this study, representing all six ecological provinces along the transect. The depth of the DCM layer varied from 28 m in the BENG province to 90 m in the NATR province. Most of the sequence diversity present in the DCM communities is dominated by the following groups: Alphaproteobacteria clade SAR 11 (25 % ± 4 %), the Cyanobacteria clade GpIIa (10 % ± 4 %), the Gammaproteobacteria (13 % ± 5 %) and the Bacteroidetes (7 % ± 2 %).

Table 3. Spearman correlations between environmental variables and principal coordinate analysis (PCoA) ordination. Table shows number of available data points (n), Spearman correlations (r_s), and significance value (p). Only significant relationships are shown.

Coordinate	Env. variable	n	r_s	p
1				
	PO_4^{3-}	15	0.82	0.000
	NO_3^-	15	0.81	0.000
	Depth$_{(log)}$	16	0.73	0.001
	Si	15	0.59	0.021
	DOC	14	−0.85	0.000
	Temperature	16	−0.82	0.000
	DON	14	−0.67	0.009
2				
	NO_2^-	15	−0.76	0.001
	Chl-a	15	−0.68	0.005
	Salinity	16	−0.55	0.027

3.2.3 Deep communities

Recent studies have demonstrated a rich diversity within the bacterial communities residing in the deep waters of the ocean, despite the high pressure, low temperatures, and an absence of light that might otherwise be expected to restrict bacterial distributions (Nagata et al., 2000; Hewson et al., 2006; Sogin et al., 2006; Reinthaler et al., 2010; Agogue et al., 2011). Five of the communities we studied were obtained from this water layer (i.e., below the pycnocline, ranging in depth from 100–4600 m). Two of the samples (8 and 11) were obtained from the South Atlantic Central Water (SACW) within the WTRA province at depths of 100 and 200 m, respectively. The three other samples, from depths greater than 1000 m (Sample 2: 1100 m, 12: 1300 m, and 14: 4604 m), were obtained from distinct geographical locations (NASE, WTRA, and BENG provinces, respectively) and different water masses with variable physical parameters (mainly temperature and salinity). The water masses included are Eurafrican Mediterranean water (EMW), Antarctic Intermediate Water (AAIW), and Antarctic Bottom Water (AABW). The deep-zone communities are dominated by the Proteobacteria (73 % abundance), with the SAR11 clade representing 28 % of the total abundance. The phyla Cyanobacteria, Bacteroidetes, and Actinobacteria represent 7 %, 4 %, and 4 % of the deep zone communities' abundance, respectively. Gammaproteobacteria are enriched in the deep-zone communities, representing 24 % of the abundance. Among the Gammaproteobacteria, the Alteromonadales and the Oceanospirillales families represent 11 % and 3 % of the abundance, respectively. In comparison with the communities from the surface and the DCM zone, the deep zone communities showed a higher relative abundance of the Gammaproteobacteria, Actinobacteria, Firmicutes, and Bacteroidetes. Relative enrichment, in deep water samples,

Table 4. Top 10 correlated (Spearman, r_s) OTUs with principal coordinate analysis (PCoA) ordination. Counts (n) are normalized read abundance and an OTU numerical identifier is shown in parentheses to differentiate OTUs with identical taxonomic labels.

Coordinate	Taxa (OTU)	n	r_s (PCoA)
1			
	Francisella (509)	533	0.90
	Actinobacteria (496)	1043	0.86
	Proteobacteria (934)	55	0.83
	Vibrionaceae (1166)	93	0.81
	Alphaproteobacteria (1301)	138	0.81
	Pelagibacter (24)	9188	0.79
	Bacteria (1293)	56	0.78
	Proteobacteria (1204)	301	0.77
	Actinobacteria (493)	474	0.76
	Proteobacteria (928)	1621	0.75
2			
	Porphyrobacter (1678)	33	0.79
	Bacteria (2272)	76	0.75
	Actinobacteria (2111)	29	0.74
	Gammaproteobacteria (2241)	86	0.74
	Marichromatium (2244)	25	0.74
	Bacteria (2269)	18	0.73
	Proteobacteria (2532)	6	0.73
	Pelagibacter (2607)	6	0.73
	Colwellia (2192)	30	0.73
	Actinomycetales (2215)	6	0.73
3			
	GpIIa (287)	247	0.87
	GpIIa (286)	1164	0.74
	Bacteria (464)	54	0.74
	Rhodobacteraceae (97)	5	0.72
	Proteobacteria (452)	23	0.71
	Bacteria (390)	10	0.70
	Maricaulis (380)	21	0.70
	Shimia (93)	34	0.68
	Rhodobacteraceae (95)	17	0.68
	Proteobacteria (141)	3	0.68

for the same phyla has been observed by others (Martín-Cuadrado et al., 2007; Sogin et al., 2006; Agogue et al., 2011).

3.3 Ecological drivers

3.3.1 Spatial and environmental controls on bacteria assemblages

In addition to increasing our understanding of the genetic diversity and distribution of bacteria in the Atlantic Ocean, the pyrosequencing data allowed us to explore the relationship between bacterial assemblage structure and several environmental factors. This was accomplished using multivariate statistical tools common in community ecology to examine overall patterns in community composition in both biogeo-
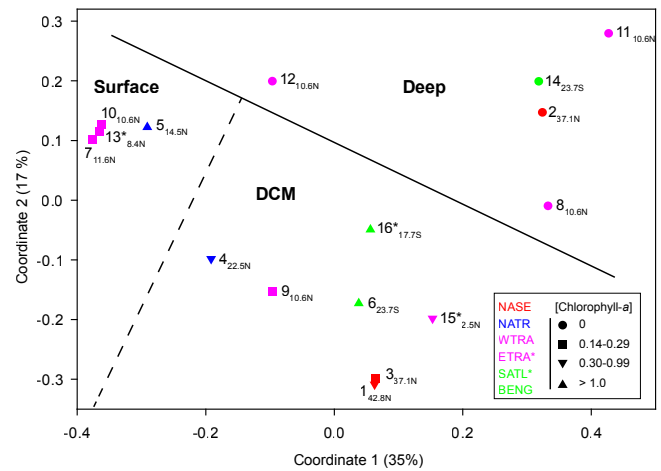

Fig. 5. Principal coordinate analysis (PCoA) of the normalized OTU abundances based on Bray-Curtis dissimilarity. Each individual sample is identified by its respective biogeographic province (symbol color) and Chl a concentration in $mg\,l^{-1}$ (symbol shape). Each numerical identifier corresponds to a different sample (1–16) with the latitude of the sampling location as a subscript in the legend. The first three coordinates of the PCoA explained 65 % of the variance in the community data (coordinate 1: 35 %, 2: 17 %, and 3: 13 %). The solid line represents the separation above and below the pycnocline, and the dashed line separates photic zone samples into surface and DCM samples.

graphic and biogeochemical contexts. First, a principal coordinate analysis (PCoA) was used to reduce the OTU abundance data into a small number of derived variables (coordinates) which, when plotted in a graphical space, arranged our samples along a gradient in overall community similarity. Using this approach, we see a clear clustering of the bacterial communities according to their location in the water column (Fig. 5). The samples contained in the top triangle of the graph, above the solid line (Deep), are from the aphotic zone, with depths ranging from 100 to 4600 m. These samples were all from beneath the pycnocline, whereas the samples below the solid line were all above the pycnocline and within the photic zone. The photic-zone samples were further differentiated into two groups, separated by the dashed line: Surface samples (2–20 m) and DCM samples (28–90 m). ANOSIM confirmed these three groups were significantly different ($r = 0.65$, $p = 0.0001$; all pairwise comparisons $p < 0.02$), and that classification by Longhurst province was not ($r = 0.15$, $p = 0.14$). The lack of a significant relationship between all samples and Longhurst province is not necessarily unexpected as province designations are primarily restricted to the epipelagic zone and do not consider deeper water masses (Longhurst et al., 1995). The observation that samples from similar habitats cluster together on the PCoA allows one to conclude that the current environmental factors influence the biogeographic patterns exhibited by the eastern Atlantic Ocean bacterial assemblages.

Correlation analysis was used to identify the OTUs contributing to the separation of samples via PCoA. Only four of the 25 most abundant OTUs (Table 4) identified in Table 2) are important contributors to separation of samples in the PCoA; those four OTUs are affiliated to taxonomic clades *Pelagibacter*, Proteobacteria, GpIIa, and *Actinobacter* (OTUs 24, 928, 286, and 496, respectively). Previous evidence also find the same clades showing vertical stratification (Giovannoni and Stingl, 2005; DeLong et al., 2006; Martiny et al., 2006a; Aristegui et al., 2009b; Zinger et al., 2011).

As one would expect, we see high correlations with some of the most abundant OTUs, but not exclusively. Of the less abundant OTUs contributing to the separation of the samples in the PCoA, some belong to taxonomic families showing differential patterns of abundance across the samples (Fig. 3), including members of the Erythrobacteraceae, Hyphomonadaceae, Rhodobacteraceae, Chromatiaceae, Colwelliaceae, Franscisellaceae and Vibrionaceae families (OTUs 1678, 380, 93, 2244, 2192, 509, and 1166, respectively). The first axis of the PCoA distinguishes the surface samples from the deeper ones, with DCM samples being intermediate between the two. The second axis demonstrates a separation of the DCM samples, which appears to be driven primarily by many non-abundant OTUs. Overall, it is clear that the driving force behind separation in ordination space is not necessarily the most abundant community member, but rather the less abundant OTUs presumably adapted to environmental and biogeochemical conditions associated with geography, depth, or both. We tested this hypothesis using the Multivariate Cutoff Level Analysis (MultiCoIA) developed by Gobet et al. (2010) and found that even after removing 80 % of the most abundant OTUs, a non-metric multidimensional scaling (NMDS) ordination of the truncated data set resulted in a Procrustes correlation of approximately 0.8 (data not shown). However, because the Bray-Curtis distance metric used in our ordination and embedded in MultiCoLA can be sensitive to the influence of highly abundant community members, we are cautious to discount the influence of less abundant OTUs. In fact, this highlights an important distinction between traditional distance-based methods of ordination and the Bayesian analysis presented here: we are able to recover the major depth gradient using distance-based ordination, but the tail of the distribution is what allows us to hone in the biogeographic signal that is presumably lost when the OTU abundances are reduced to a simple distance.

Given the broad geographic range of these samples, the distinct bacterial communities found in each depth category are likely not due to location in the water column per se, but a response of the community to environmental factors that co-vary with depth and define distinct habitats. For example, the deep waters have colder temperatures, lower concentrations of DOC, and higher amounts of nitrate; each of these parameters has been shown to influence the composition of the ocean microbial community (Schattenhofer et al., 2009; Wietz et al., 2010; Agogue et al., 2011). Similarly, phytoplank-

ton abundance and community structure can be an important factor in determining bacterial community composition (Kerkhof et al., 1999; Pinhassi et al., 2003, 2004), and Chl *a* values from our samples varied dramatically with depth; concentrations are below detection in the deeper samples and significantly higher in the DCM. In the surface samples, Chl *a* concentration was low (0.14–0.29 mg l^{-1}) with the exception of sample 5 (2.8 mg l^{-1}). The anomalously high Chl *a* value for sample 5 is most likely attributable to a phytoplankton bloom stimulated by Aeolian dust deposit delivered from northeast trades winds blowing across the Saharan desert (Pohl et al., 2011; Taylor et al., 2011). It is notable that pigment analyses suggest that diatoms dominate blooms in these surface waters whereas the phytoplankton community at depth consists of a different assemblage and appears completely disconnected from the surface bloom (Taylor et al., 2011).

There has been some limited research prior to our own demonstrating unique microbial assemblages associated with depth. For example, Treusch et al. (2009) identified distinct bacterial communities found in the low nutrient surface waters, the DCM, and the upper pelagic zone from vertical profiles at the Bermuda Atlantic Times series (BATS). Their findings echo previous work noting disparate bacterial assemblages in the euphotic zone compared to the mesopelagic (Giovannoni et al., 1996; Gordon and Giovannoni, 1996; Fuhrman and Davis, 1997; Wright et al., 1997). A novel outcome from the work at BATS was the resolution of a distinct microbial community in the surface, DCM, and deep layers, which our data also illustrate (Fig. 5). These community shifts may partially be a response to changes in the physical state of the water column thereby inducing selective forces based on light availability, destructive potential of UV radiation, pressure, and temperature. Further, community differences may be a response to changes in resource availability that develop as a consequence of these physical conditions. For example, we know that the distribution of heterotrophic bacteria may vary with depth due to (1) changes in the amount of DOC (Eiler et al., 2003), (2) variations in the molecular composition of DOC derived from different phytoplankton communities (Van Hannen et al., 1999), and (3) alterations in the diagenetic state of DOC reflected in its size, molecular composition, and age (Covert and Moran, 2001). It is worth noting that our observed differences in microbial community composition with depth also correspond with molecular changes in DOC composition and its associated age, as observed by Flerus et al. (2012). Additionally, microbial community diversity in the bathypelagic ocean has been assumed to be low given such stable physical conditions. However, our samples show higher diversity with depth, partially dispelling this notion. One suggested mechanism for enhanced diversity at depth is that the microorganisms respond to episodic delivery of resources, such as particulate organic matter from surface waters, which leads to higher

community richness (Baltar et al., 2009; Bochdansky et al., 2010; Agogue et al., 2011).

To explicitly examine the biogeochemical conditions influencing the separation of the bacterial communities, we compared the environmental variables with each coordinate from the PCoA (Table 3). In addition to depth, the first coordinate from the PCoA was significantly correlated with temperature and the concentration of several dissolved organic and inorganic nutrients. In contrast, separation of the community on the second coordinate primarily relates to differences in the concentration of Chl *a* and salinity. These findings are consistent with the research summarized above, which indicates that the community changes we observed are the result of a complex coupling of multiple environmental parameters and biotic variables that may co-vary with location in the water column. In addition, we found evidence that the distribution of OTUs in our dataset reflected distance decay in community similarity. Significant results were obtained when a Mantel test was applied to compare community similarity to geographic distance (km separation calculated from sample latitude and longitude, $r_M = 0.25$, $p = 0.04$). When a partial Mantel test was conducted to remove the influence of depth and local environmental conditions, this spatial relationship becomes even stronger ($r_M = 0.35$, $p = 0.01$). Although heterogeneous environmental conditions and geographic separation have a strong influence on the biogeographic distribution of species, only recently have we begun to understand how these conditions may define distinct microbial communities in marine habitats (Giovannoni and Stingl, 2005; Martiny et al., 2006b; Pommier et al., 2006; DeLong, 2009; Fuhrman, 2009). For example, ocean water masses are frequently associated with unique microbial communities (Yokokawa et al., 2010; Varela et al., 2007; Galand et al., 2010; Hewson et al., 2009; Agogue et al., 2011). However, it has been much more challenging to assess differences in microbial communities across adjacent oceanographic biomes and, in particular, across those with complex physical hydrographic features. Prior surveys have focused on the center regions of ecologically discrete provinces (Pommier et al., 2006; Martiny et al., 2009; Schattenhofer et al., 2009; Wietz et al., 2010), and there is little work, such as ours, that considers the distribution pattern associated with the fluid boundaries between provinces (Ducklow, 2003) or in transition zones.

3.3.2 Bayesian inference of bacterial assemblage relationships

We find a clear separation of communities based on depth which reveals that samples from the same water layer are more similar to each other than they are to geographically-proximal samples obtained from different water layers. Communities separated by thousands of kilometers (e.g., samples 2 and 14; Fig. 5) in the deep ocean are more similar in composition than communities separated by just a few
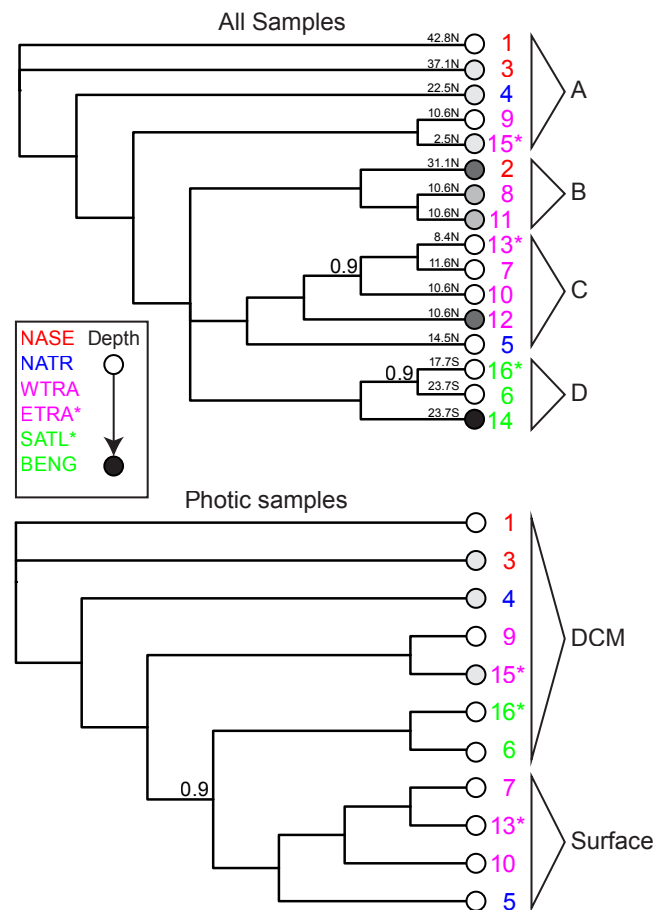

Fig. 6. Tree topology estimated by Bayesian inference operating on the abundance of the OTUs identified at 97 % identity. Province is shown by color and depth is indicated by circles of increasing darkness (2–4600 m). Only branch posterior probabilities <1.0 are shown on the tree. The latitude of the sampling location is indicated at the tip of the branches of the "All Samples" topology.

meters in depth but residing in different water layers (e.g., samples 9 and 10; Fig. 5). These findings support a growing understanding that the major discontinuities in the ocean are related to the physicochemical properties that form the different water masses. These previous results were obtained using PCoA and Mantel tests, which are both based on an initial calculation of a similarity (or distance) matrix that fundamentally reduces a dataset into a single numerical value relating each pair of samples. In our study, the original OTU data exist as a matrix of 16 samples with 2871 values per sample. The first step for both the PCoA and the Mantel test is to reduce this information to a mere 136 values (a single number for each pair of samples). Though valuable in their power to distill large and complex datasets, these approaches simply cannot maintain the full information potential of the original data, and ecologically valuable information may be lost. The Spearman correlation analysis of the OTUs with the PCoA ordination (Table 4) revealed that both abundant

and less abundant OTUs were associated with the ordination-space separation of the bacterial communities; this suggested to us that using the full information content of the data matrix, instead of just the pairwise distance between the samples, might provide enhanced discriminating power and allow us to explore ecologically-relevant patterns nested within our depth-defined habitats. With this goal in mind, we designed an analysis strategy not previously used to our knowledge in this type of study. By applying Bayesian inference of tree topology to the full matrix of OTU abundance, we were able to identify the optimal tree topology that best explains the relationships based on overall patterns of OTU abundance (Fig. 6). Although Bayesian inference is more computationally expensive than distance-based clustering algorithms, it uses the information contained at each column of the data matrix to test not only if a tree structure is the best representation of the relationships of the samples under study but also to determine which tree topology is best supported by the data. The materials and methods section presents a more detail explanation of the justifications and assumptions used to model OTU abundance and bacterial assemblages relationships using a Bayesian inference approach; a formal proof and extensions will be discussed elsewhere (Friedline and Rivera, 2012).

The inferred tree topology resolves four interesting clades or groups of samples, labelled by their common nodes (Fig. 6, A–D). The three samples from the Southern Hemisphere (6, 14, and 16) form a well-supported clade (Fig. 6, node D). This Southern Hemisphere clade includes DCM-zone samples from the BENG and the SATL ecological provinces separated by a surface distance of 868 km (6 and 16) and by a depth of over 4600 m (14 and 16). The samples within the Northern Hemisphere are arranged into three major clades (Fig. 6, nodes A, B, and C); showing a more complex pattern of relationships that seems to be determined by both common habitat and geographic separation. Samples from the Northern Hemisphere DCM-zone form a well-defined group (Fig. 6, node A), to the exclusion of samples from the same geographical region but different depths profiles. Clade B is a well-supported group including three aphotic zone samples (2, 8, and 11). This is an interesting group as it includes samples separated by ∼3000 km and from two different provinces (NASE and WTRA), but with the unifying characteristic of having been collected from within the central water mass. Sample 2 was collected at a depth of 1100 m and a temperature of 10.77 °C from within the NACW with EMW influence west of the Iberian Peninsula; samples 8 and 11 were collected at depths of 100 and 200 m, respectively, from within the SACW mass. This cluster shows that sample 2 has a community composition more similar to samples 8 and 11 with temperatures of 15.3 and 12.8 °C, respectively, than to samples from similar depths, like sample 12 (1300 m, 4.73 °C) but at a colder temperature. Martin-Cuadrado and collaborators (2007) observed similar results when analyzing EMW bathypelagic samples. The

deep Mediterranean communities were similar to deep communities from the Pacific, but they were more closely related to Pacific mesopelagic communities than to other bathypelagic communities, suggesting temperature as the major driving factor.

Finally, clade C is composed of five samples from two different provinces, NATR and WTRA, all collected within a maximum surface distance of 670 km from each other and at depths ranging for 2–1300 m. In this cluster, samples from the same depth (2 m) but different provinces (7 and 5) have less similar community composition than samples from different depths but within the same province (7, 10, 12, and 13). This analysis may suggest the province signal is not restricted to the surface communities, but can be conveyed to the communities of the deep ocean, particularly in the case of samples 12 and 14. Analysis of a larger number of samples is necessary to further explore this possibility, although other studies from the same cruise observed similar trend (Taylor et al., 2011).

To further investigate the biogeography effect without the confounding signal from the deep-water communities, the same analysis was performed using only the samples from the photic zone (Fig. 6, bottom). This tree topology clearly shows the photic zone communities separated into two groups based on the water layer: DCM and surface. Within those two habitats, the communities show a decay in similarity with increasing geographic distance following a north–south gradient; this pattern is better observed among the communities from the DCM zone where a larger and more diverse set of samples was obtained. The dendrogram relating the photic zone samples allows one to clearly assess the contribution of current environmental factors and historical events; suggesting that at the spatial scales of this study the similarity of the bacterial assemblages is determined by multiple habitats and multiple biotic provinces.

Collectively, our results suggests that eastern Atlantic Ocean bacterial assemblages are vertically stratified by similar water layers (habitat). Within the same water layer, the separation of the communities appears to show a significant geographical distance effect, and indication of provincialism (Martiny et al., 2006a). In general, the Bayesian inference approach provides the finer resolution power needed to infer the influence of current environmental features and historical contingencies on the bacterial assemblages.

4   Conclusions

Our study provides a comprehensive picture of the composition of bacterial assemblages along the eastern Atlantic Ocean using high-throughput pyrosequencing of PCR-amplified 16s rRNA. The application of community ecology statistics to OTU data leads us to conclude that bacterial assemblages are not spatially random, showing a biogeographic separation based on their position in the water

column (surface, DCM, and deep). A novel Bayesian inference approach further extracted information from the community composition and suggested both contemporary habitat and historical contingencies, as demonstrated by the evidence of decreasing community composition similarity with geographical distance (provincialism) influence bacterial biogeography. The observed stratification patterns are driven not only by the most abundant OTUs, but also by less abundant OTUs, suggesting that rare taxa contribute to the unique character of the community and are important biogeographic markers. In general, the distribution patterns of the bacterial assemblages were congruent with the Longhurst ecological provinces. However, a more extensive sampling will be required in order to fully assess the impact of the Longhurst ecological provinces on the distribution and diversity of bacterial communities. Extensive studies at different spatial scales and employing novel analytical methodology that takes full advantage of the wealth of information provided by high-throughput sequencing technology will be required to fully understand if the spatial scaling rules and biogeographic patterns observed in plants and animals also apply to bacterial assemblages.

References

Agogue, H., Lamy, D., Neal, P. R., Sogin, M. L., and Herndl, G. J.: Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing, Molecular Ecology, 20, 258–274, 2011.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D.: Basic local alignment search tool, J. Molecular Biology, 215, 403–410, 1990.

Anderson, B., Arrigo, K., Ward, B., UCI, K., and Van Mooy, B.: The Molecular Biology of Biogeochemistry: Using Molecular Methods to Link Ocean Chemistry with Biological Activity, http://www.whoi.edu/sites/molbiogeochem, 2010.

Aristegui, J., Barton, E. D., Alvarez-Salgado, X. A., Santos, A. M. P., Figueiras, F. G., Kifani, S., Hernández-León, S., Mason, E., Machú, E., and Demarcq, H.: Sub-regional ecosystem variability in the Canary Current upwrelling, Prog. Oceanography, 83, 33–48, 2009a.

Aristegui, J., Gasol, J. M., Duarte, C. M., and Herndl, G. J.: Microbial oceanography of the dark ocean's pelagic realm, Limnol. Oceanogr., 54, 1501–1529, 2009b.

Azam, F., Fenchel, T., Field, J., Gray, J., Meyerreil, L., and Thingstad, F.: The ecological role of water-column microbes in the sea, Marine Ecology-Progress Series, 10, 257–263, 1983.

Baltar, F., Aristegui, J., Gasol, J. M., Sintes, E., and Herndl, G. J.: Evidence of prokaryotic metabolism on suspended particulate organic matter in the dark waters of the subtropical North Atlantic, Limnol. Oceanogr., 54, 182–193, 2009.

Bochdansky, A. B., van Aken, H. M., and Herndl, G. J.: Role of macroscopic particles in deep-sea oxygen consumption, Proc. Natl. Acad. Sci. USA, 107, 8287–8291, 2010.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J., Segall, A., Mead, D., Azam, F., and Rohwer, F.: Genomic analysis of uncultured marine viral communities, Proc. Natl. Acad. Sci. USA, 99, 14250–14255, 2002.

Bunge, J.: Estimating the number of species with CatchAll, Pacific Symposium on Biocomputing, 2011.

Carlson, C. A., Giovannoni, S. J., Hansell, D., Goldberg, S., Parsons, R., Otero, M., Vergin, K. L., and Wheeler, B.: Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern Sargasso Sea, Aquatic Microbial Ecology, 30, 19–36, 2002.

Cases, I. and de Lorenzo, V.: The grammar of (micro)biological diversity, Environmental Microbiology, 4, 623–627, 2002.

Chao, A.: Nonparametric-estimation of the number of classes in a population, Scandinavian Journal of Statistics, 11, 265–270, 1984.

Chao, A. and Lee, S.: Estimating the Number of Classes via Sample Coverage, Journal of the American Statistical Association, 87, 210–217, 1992.

Chave, J.: Neutral theory and community ecology, Ecology Letters, 7, 241–253, 2004.

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilae, J., Marchesi, J. R., Smidt, H., Vos, W. M. D., Ross, R. P., and O'Toole, P. W.: Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine, PLoS One, 4, e6669, 2009.

Clarke, K.: Nonparametric multivariate analyses of changes in community structure, Australian Journal of Ecology, 18, 117–143, 1993.

Cohan, F.: What are bacterial species?, Annual Review of Microbiology, 56, 457–487, 2002.

Cohan, F. M. and Perry, E. B.: A systematics for discovering the fundamental units of bacterial diversity, Current Biology, 17, R373–R386, 2007.

Covert, J. and Moran, M.: Molecular characterization of estuarine bacterial communities that use high- and low-molecular weight fractions of dissolved organic carbon, Aquatic Microbial Ecology, 25, 127–139, 2001.

del Giorgio, P. and Duarte, C.: Respiration in the open ocean, Nature, 420, 379–384, 2002.

DeLong, E. F.: The microbial ocean from genomes to biomes, Nature, 459, 200–206, 2009.

DeLong, E. F., Preston, C., Mincer, T., Rich, V., Hallam, S., Frigaard, N., Martinez, A., Sullivan, M., Edwards, R. A., Brito, B. R., Chisholm, S. W., and Karl, D. M.: Community genomics among stratified microbial assemblages in the ocean's interior, Science, 311, 496–503, 2006.

Ducklow, H. W.: Biogeochemical Provinces: Towards a JGOFS Synthesis, in: Ocean Biogeochemistry, edited by: Fasham, M. J. R., Global Change – The IGBP Series, pp. 3–17, Springer Berlin Heidelberg, 2003.

Eiler, A., Langenheder, S., Bertilsson, S., and Tranvik, L. J.: Heterotrophic bacterial growth efficiency and community structure at different natural organic carbon concentrations, Applied and Environmental Microbiology, 69, 3701–3709, 2003.

Flerus, R., Koch, B. P., Lechtenfeld, O. J., McCallister, S. L., Schmitt-Kopplin, P., Benner, R., Kaiser, K., and Kattner, G.: A molecular perspective on the ageing of marine dissolved organic matter, Biogeosciences Discuss., 8, 11453–11488, doi:10.5194/bgd-8-11453-2011, 2011.

Fuhrman, J. A.: Microbial community structure and its functional implications, Nature, 459, 193–199, 2009.

Fuhrman, J. and Azam, F.: Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters – evaluation and field results, Marine Biology, 66, 109–120, 1982.

Fuhrman, J. and Campbell, L.: Marine ecology - microbial microdiversity, Nature, 393, 410–411, 1998.

Fuhrman, J. A. and Davis, A.: Widespread Archaea and novel Bacteria from the deep sea as shown by 16S rRNA gene sequences, Marine Ecology-Progress Series, 150, 275–285, 1997.

Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L., and Brown, J. H.: A latitudinal diversity gradient in planktonic marine bacteria, Proc. Natl. Acad. Sci. USA, 105, 7774–7778, 2008.

Galand, P. E., Casamayor, E. O., Kirchman, D. L., Potvin, M., and Lovejoy, C.: Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing, The ISME Journal, 3, 860–869, 2009.

Galand, P. E., Potvin, M., Casamayor, E. O., and Lovejoy, C.: Hydrography shapes bacterial biogeography of the deep Arctic Ocean, The ISME Journal, 4, 564–576, 2010.

Gasol, J. and del Giorgio, P.: Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities, in: Scientia Marina, pp. 197–224, CSIC, Inst Ciencias Mar, Dept Biol Marina & Oceanog, E-08039 Barcelona, Spain, 2000.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P. J., Huse, S., and Joint, I.: The seasonal structure of microbial communities in the Western English Channel, Environmental Microbiology, 11, 3132–3139, 2009.

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrueck, L., Reeder, J., Temperton, B., Huse, S., Mchardy, A. C., Knight, R., Joint, I.,

Somerfield, P., Fuhrman, J. A., and Field, D.: Defining seasonal marine microbial community dynamics, The ISME Journal, 6, 298–308, 2012.

Giovannoni, S. J. and Stingl, U.: Molecular diversity and ecology of microbial plankton, Nature, 437, 343–348, 2005.

Giovannoni, S. J., Rappé, M. S., Vergin, K. L., and Adair, N.: 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the Green Non-Sulfur bacteria, Proc. Natl. Acad. Sci. USA, 93, 7979–7984, 1996.

Giovannoni, S. J., Bibbs, L., Cho, J.-C., Stapels, M. D., Desiderio, R., Vergin, K. L., Rappé, M. S., Laney, S., Wilhelm, L. J., Tripp, H. J., Mathur, E. J., and Barofsky, D. F.: Proteorhodopsin in the ubiquitous marine bacterium SAR11, Nature, 438, 82–85, 2005.

Gobet, A., Quince, C., and Ramette, A.: Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets, Nucleic Acids Research, 38, e155, 2010.

Gordon, D. A. and Giovannoni, S. J.: Detection of stratified microbial populations related to Chlorobium and Fibrobacter species in the Atlantic and Pacific oceans, Applied and Environmental Microbiology, 62, 1171–1177, 1996.

Gower, J.: General coefficient of similarity and some of its properties, Biometrics, 27, 857–874, 1971.

Hammer, Ø. and Harper, D.: PAST: paleontological statistics software package for education and data analysis, Palaeontologia Electronica, 2001.

Hewson, I., Steele, J., Capone, D., and Fuhrman, J. A.: Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition, Limnol. Oceanogr., 51, 1274–1283, 2006.

Hewson, I., Paerl, R. W., Tripp, H. J., Zehr, J. P., and Karl, D. M.: Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat, Limnol. Oceanogr., 54, 1981–1994, 2009.

Hobbie, J. E., Daley, R. J., and Jasper, S.: Use of nuclepore filters for counting bacteria by fluorescence microscopy, Applied and Environmental Microbiology, 33, 1225–1228, 1977.

Hubbell, S. P.: The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32) (Monographs in Population Biology), Princeton University Press, 2001.

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

Huelsenbeck, J. and Ronquist, F.: MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics, 17, 754–755, 2001.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., and Welch, D. M.: Accuracy and quality of massively parallel DNA pyrosequencing, Genome Biology, 8, R143, 2007.

Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A., and Sogin, M. L.: Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing, PLoS Genetics, 4, e1000 255, 2008.

Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L.: Ironing out the wrinkles in the rare biosphere through improved OTU clustering, Environmental Microbiology, 12, 1889–1898, 2010.

Kass, R. and Raftery, A.: Bayes Factors, Journal of the American Statistical Association, 90, 773–795, 1995.

Kerkhof, L., Voytek, M., Sherrell, R., Millie, D., and Schofield, O.: Variability in bacterial community structure during upwelling in the coastal ocean, Hydrobiologia, 401, 139–148, 1999.

Kirchman, D. L., Cottrell, M. T., and Lovejoy, C.: The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes, Environmental Microbiology, 12, 1132–1143, 2010.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P.: Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates, Environmental Microbiology, 12, 118–123, 2010.

Legendre, P. and Legendre, L.: Numerical Ecology, Volume 24, Second Edition (Developments in Environmental Modelling), Elsevier Science, 1998.

Longhurst, A.: Ecological geography of the sea, Academic Press, San Diego, CA, 1998.

Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C.: An estimate of global primary production in the ocean from satellite radiometer data, J. Plankton Res., 17, 1245–1271, 1995.

Martín-Cuadrado, A.-B., López-García, P., Alba, J.-C., Moreira, D., Monticelli, L., Strittmatter, A., Gottschalk, G., and Rodríguez-Valera, F.: Metagenomics of the deep Mediterranean, a warm bathypelagic habitat, PLoS One, 2, e914, 2007.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J., COLWELL, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M., Kane, M., Krumins, J., Kuske, C., Morin, P., Naeem, S., Øvreås, L., Reysenbach, A.-L., Smith, V., and Staley, J.: Microbial biogeography: putting microorganisms on the map, Nature Reviews Microbiology, 4, 102–112, 2006a.

Martiny, A. C., Coleman, M. L., and Chisholm, S. W.: Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation, Proc. Natl. Acad. Sci. USA, 103, 12552–12557, 2006b.

Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F., and Chisholm, S. W.: Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*, Environmental Microbiology, 11, 823–832, 2009.

Massana, R., Karniol, B., Pommier, T., Bodaker, I., and Beja, O.: Metagenomic retrieval of a ribosomal DNA repeat array from an uncultured marine alveolate, Environmental Microbiology, 10, 1335–1343, 2008.

Nagata, T., Fukuda, H., Fukuda, R., and Koike, I.: Bacterioplankton distribution and production in deep Pacific waters: Large-scale geographic variations and possible coupling with sinking particle fluxes, Limnol. Oceanogr., 45, 426–435, 2000.

Nekola, J. and White, P.: The distance decay of similarity in biogeography and ecology, J. Biogeography, 26, 867–878, 1999.

Pinhassi, J., Winding, A., Binnerup, S., Zweifel, U., Riemann, B., and Hagstrom, A.: Spatial variability in bacterioplankton community composition at the Skagerrak-Kattegat Front, Marine Ecology-Progress Series, 255, 1–13, 2003.

Pinhassi, J., Sala, M., Havskum, H., Peters, F., Guadayol, O., Malits, A., and Marrase, C.: Changes in bacterioplankton composition under different phytoplankton regimens, Applied And Environmental Microbiology, 70, 6753–6766, 2004.

Pohl, C., Croot, P., Hennings, U., Daberkow, T., Budeus, G., and Loeff, M.: Synoptic transects on the distribution of trace elements (Hg, Pb, Cd, Cu, Ni, Zn, Co, Mn, Fe, and Al) in surface waters of the Northern-and Southern East Atlantic, J. Marine Syst., 84, 28–41, 2011.

Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., Tunlid, A., and Hagstrom, A.: Global patterns of diversity and community structure in marine bacterioplankton, Molecular Ecology, 16, 867–880, 2006.

Quince, C., Curtis, T. P., and Sloan, W. T.: The rational exploration of microbial diversity, The ISME Journal, 2, 997–1006, 2008.

Reeder, J. and Knight, R.: The 'rare biosphere': a reality check, Nature Methods, 6, 636–637, 2009.

Reinthaler, T., van Aken, H. M., and Herndl, G. J.: Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior, Deep-Sea Res. Part II-Topical Studies in Oceanography, 57, 1572–1580, 2010.

Ricklefs, R. E. and Schluter, D. (Eds.): Species Diversity in Ecological Communities, University Of Chicago Press, Chicago, 1 edn., 1994.

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., Daroub, S. H., Camargo, F. A. O., Farmerie, W. G., and Triplett, E. W.: Pyrosequencing enumerates and contrasts soil microbial diversity, The ISME Journal, 1, 283–290, 2007.

Rossi, J.: Statistical tool for soil biology: Autocorrelogram and Mantel test, European Journal of Soil Biology, 32, 195–203, 1996.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y.-H., Falcon, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Nealson, K., Friedman, R., Frazier, M., and Venter, J. C.: The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific, PLoS Biology, 5, 398–431, 2007.

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., and Pernthaler, J.: Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean, Environmental Microbiology, 11, 2078–2093, 2009.

Schlitzer, R.: Ocean Data View, http://odv/awi.de, 2011.

Schloss, P. D.: A high-throughput DNA sequence aligner for microbial ecology studies., PloS One, 4, e8230–e8230, 2009.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., van Horn, D. J., and Weber, C. F.: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Applied and Environmental Microbiology, 75, 7537–7541, 2009.

Schols, P., D'hondt, C., Geuten, K., Merckx, V., Janssens, S., and Smets, E.: MorphoCode: coding quantitative data for phylogenetic analysis, Phyloinformatics, 4, 1–4, 2004.

Smouse, P., Long, J., and Sokal, R.: Multiple-regression and correlation extensions of the mantel test of matrix correspondence, Systematic Zoology, 35, 627–632, 1986.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J.: Microbial diversity in the deep sea and the underexplored "rare biosphere",

Proc. Natl. Acad. Sci. USA, 103, 12115–12120, 2006.

Sokal, R. R. and Sneath, P. H. A.: Principles of Numerical Taxonomy (A Series of Books in Biology), W. H. Freeman and Company, 1st edn., 1963.

Sokal, R. R. and Sneath, P. H. A.: Numerical Taxonomy: The Principles and Practice of Numerical Classification, W. H. Freeman and Company, 1973.

Stott, J.: Jcoord, version 1,0, http://www.jstott.me.uk/jcoord, 2011.

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., Piera, J., and Bracher, A.: Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance, Biogeosciences, 8, 3609–3629, doi:10.5194/bg-8-3609-2011, 2011.

Thiele, K.: The holy grail of the perfect character: the cladistic treatment of morphometric data, Cladistics, 9, 275–304, 1993.

Treusch, A. H., Vergin, K. L., Finlay, L. A., Donatz, M. G., Burton, R. M., Carlson, C. A., and Giovannoni, S. J.: Seasonality and vertical structure of microbial communities in an ocean gyre, The ISME Journal, 3, 1148–1163, 2009.

Van Hannen, E., Zwart, G., Van Agterveld, M., Gons, H., Ebert, J., and Laanbroek, H.: Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses, Applied and Environmental Microbiology, 65, 795–801, 1999.

Varela, M. M., van Aken, H. M., Sintes, E., and Herndl, G. J.: Latitudinal trends of Crenarchaeota and Bacteria in the meso- and bathypelagic water masses of the Eastern North Atlantic, Environmental Microbiology, 10, 110–124, 2007.

Venter, J., Remington, K., Heidelberg, J., Halpern, A., Rusch, D., Eisen, J., Wu, D., Paulsen, I., Nelson, K., Nelson, W., Fouts, D., Levy, S., Knap, A., Lomas, M., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y., and Smith, H.: Environmental genome shotgun sequencing of the Sargasso Sea, Science, 304, 66–74, 2004.

VLIZ: Longhurst Biogeographical Provinces, http://www.vliz.be/vmdcdata/vlimar/downloads.php, 2009.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R.: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Applied and Environmental Microbiology, 73, 5261–5267, 2007.

West, N. J., Obernosterer, I., Zemb, O., and Lebaron, P.: Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean, Environmental Microbiology, 10, 738–756, 2008.

Wietz, M., Gram, L., Jorgensen, B., and Schramm, A.: Latitudinal patterns in the abundance of major marine bacterioplankton groups, Aquatic Microbial Ecology, 61, 179–189, 2010.

Wright, T. D., Vergin, K. L., Boyd, P. W., and Giovannoni, S. J.: A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer, Applied and Environmental Microbiology, 63, 1441–1448, 1997.

Xie, Z., Koch, B. P., Möller, A., Sturm, R., and Ebinghaus, R.: Transport and fate of hexachlorocyclohexanes in the oceanic air and surface seawater, Biogeosciences, 8, 2621–2633, doi:10.5194/bg-8-2621-2011, 2011.

Yokokawa, T., De Corte, D., Sintes, E., and Herndl, G. J.: Spatial patterns of bacterial abundance, activity and community composition in relation to water masses in the eastern Mediterranean Sea, Aquatic Microbial Ecology, 59, 185–195, 2010.

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., Martiny, J. B. H., Sogin, M. L., Boetius, A., and Ramette, A.: Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems, PloS One, 6, e24570, 2011.