

# Connectivity between surface and deep waters determines prokaryotic diversity in the North Atlantic Deep Water

Alexander H. Frank,<sup>1</sup> Juan A. L. Garcia,<sup>1</sup>  
Gerhard J. Herndl<sup>1,2</sup> and Thomas Reinthaler<sup>1\*</sup>

<sup>1</sup>Department of Limnology and Bio-Oceanography,  
University of Vienna, Vienna, Austria.

<sup>2</sup>Department of Biological Oceanography, Royal  
Netherlands Institute for Sea Research, Texel, The  
Netherlands.

## Summary

To decipher the influence of depth stratification and surface provincialism on the dark ocean prokaryotic community composition, we sampled the major deep-water masses in the eastern North Atlantic covering three biogeographic provinces. Their diversity was evaluated using ordination and canonical analysis of 454 pyrotag sequences. Variance partitioning suggested that 16% of the variation in the bacterial community composition was based on depth stratification while 9% of the variation was due to geographic location. General linear mixed effect models showed that the community of the subsurface waters was connected to the dark ocean prokaryotic communities in different biogeographic provinces. Cluster analysis indicated that some prokaryotic taxa are specific to distinct regions in bathypelagic water masses. Taken together, our data suggest that the dark ocean prokaryotic community composition of the eastern North Atlantic is primed by the formation and the horizontal transport of water masses.

## Introduction

The ocean harbors about  $10^{29}$  prokaryotic cells with an enormous phylogenetic diversity (Whitman *et al.*, 1998). In terms of volume, the open ocean below 200 m depth represents the largest biome on earth with Bacteria and

Archaea playing a major role in the organic matter cycling in the meso- and bathypelagic waters (Azam *et al.*, 1983; Reinthaler *et al.*, 2010). However, our knowledge on microbial taxa and their traits in the dark ocean is still limiting our mechanistic understanding of the functioning of dark ocean microbial communities.

For the surface ocean, annually repeatable patterns in bacterial community composition have been found, associated with recurring physico-chemical changes in the environment (Fuhrman *et al.*, 2006; Vergin *et al.*, 2013). The distribution of surface water bacterial communities also suggested a bipolar distribution where the community composition in the Arctic and Antarctic regions was more similar to each other than compared to those at lower latitudes (Ghiglione *et al.*, 2012). Moreover, an extended global epipelagic dataset indicated biogeographic patterns of Bacteria similar to macroscopic organisms (Sul *et al.*, 2013). The reason for the global distribution pattern of open ocean Bacteria is still controversial and might be the result of dispersal limitation or environmental selection (Martiny *et al.*, 2006). A recent simulation using a global surface ocean circulation model supports the notion that neutral evolution in combination with dispersal limitation generates biogeographic patterns in bacterial community composition in surface waters (Hellweger *et al.*, 2014).

In addition to the latitudinal trends, prokaryotic community composition in the ocean is depth-stratified (DeLong *et al.*, 2006; Brown *et al.*, 2009), potentially linked to depth-related changes in the quality and quantity of available energy sources. It is generally assumed that the pycnocline at the periphery of distinct deep-water masses is a major constraint for the unrestricted, vertical distribution of the free-living prokaryotes to the dark ocean. This dispersal limitation may result in an apparent depth stratification of the free-living non-sinking prokaryotic community (Galand *et al.*, 2010; Agogue *et al.*, 2011). In contrast, microbial communities attached to sinking particles are likely able to cross density gradients and thus might significantly contribute to the diversity of deep-water prokaryotic communities (Nagata *et al.*, 2010; Moeseneder *et al.*, 2012), particularly if the sinking particles enter denser deep waters and thereby become neutrally buoyant

Received 13 May, 2015; revised: 17 December, 2015; accepted 18 January, 2016. \*For correspondence. E-mail: thomas.reinthal@univie.ac.at; Tel. +43-1-4277-76432; Fax: +43-1-4277-876401

†Disclosure: The authors declare no conflict of interest.

(Bochdansky *et al.*, 2010). The extent of surface derived prokaryotes shaping deep-water diversity depends on the particle properties and thus indirectly on the surface water primary production (Cram *et al.*, 2015), the strength of water column stratification and the source community composition (Hewson *et al.*, 2006; Herndl and Reinthaler, 2013).

Oceanic regions with similar physical and chemical conditions and dominant plankton species have been assigned as distinct biogeographic provinces (Longhurst *et al.*, 1995; Oliver and Irwin, 2008). The widely adopted scheme by Longhurst (2007) suggests several biogeographic provinces based on wind regimes, water column structure, nutrient concentrations, satellite-derived primary production, and phytoplankton community composition. Although this classification is based on surface ocean properties, individual biogeographic provinces might be connected to a variable extent to the processes in the deeper water column.

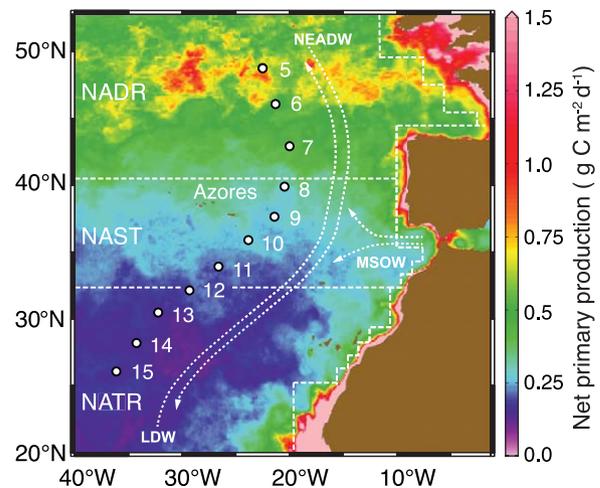
Thus far, only few studies have examined prokaryotic activity and diversity in relation to biogeographic provinces in the ocean. Furthermore, coherent datasets considering biogeographic provinces and the structuring forces of the deep-water masses on microbial communities are essentially missing. Studies conducted in the western North Atlantic suggest that bacterial abundance and prokaryotic heterotrophic production in the meso- and bathypelagic waters of the North Atlantic roughly reflect biogeographic provinces (De Corte *et al.*, 2012).

In the current study, we explicitly addressed the question of the connectivity between the dark ocean prokaryotic community composition and the surface ocean biogeographic provinces. Samples were obtained along a 3000 km long transect in the eastern North Atlantic (Fig. 1). We estimated the prokaryotic community composition in the subsurface waters influenced by atmospheric conditions and in five major deep-water masses over three biogeographic provinces. DNA was sequenced using pyrosequencing technology to include the rare biosphere (Sogin *et al.*, 2006). Our analyses suggest that water mass association is the predominant predictor of prokaryotic community composition in the dark ocean. However, a considerable amount of variation in community changes is attributable to the biogeographic provinces of the North Atlantic surface waters.

## Results

### Sequencing effort and yield

The primer set yielded a total of 168,634 sequences with an average of  $1,971 \pm 665$  reads and an average read length of  $370 \pm 44$  bp per sample. Sequence analysis identified a total of 1,544 bacterial OTUs with an average of  $157 \pm 35$  per sample.



**Fig. 1.** Map of occupied stations in the North Atlantic. The stations are indicated by white dots and numbers. Borders between biogeographic provinces are indicated by white dashed lines (NADR—North Atlantic Drift Province, NAST—North Atlantic Subtropical Province, NATR—North Atlantic Tropical Gyral Province). Net primary production ( $\text{g C m}^{-2} \text{d}^{-1}$ ) on the day of sampling is color-coded. The flow of the main deep water masses is indicated by white lines and arrows. MSOW—Mediterranean Sea Outflow Water, NEADW—North East Atlantic Deep Water, LDW—Lower Deep Water.

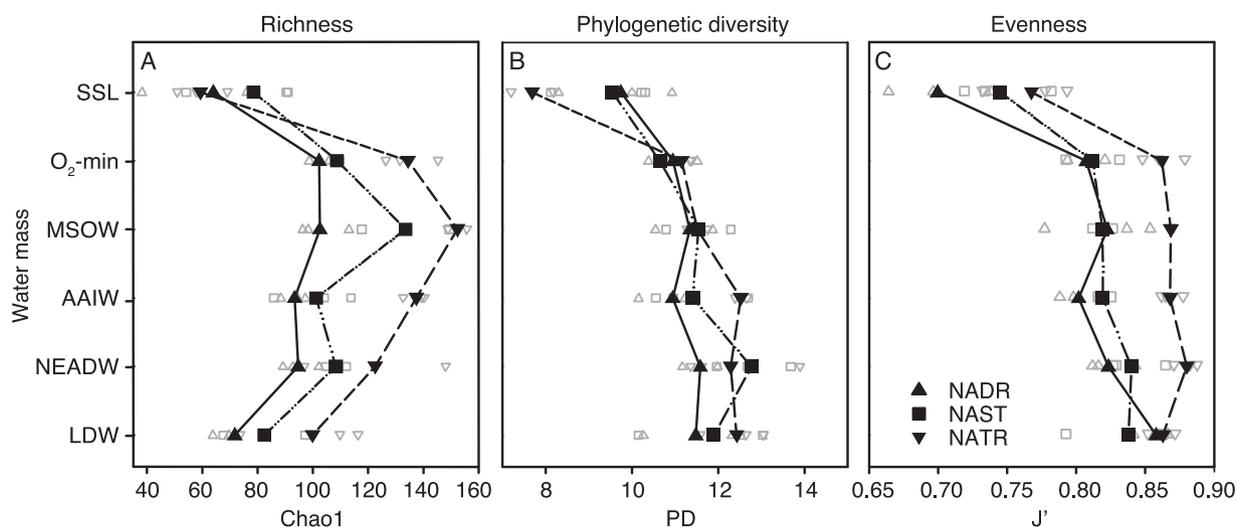
### General depth distribution of OTUs

A general categorization of the OTU depth distribution indicated that  $\sim 12\%$  of all bacterial OTUs were present exclusively in the SSL, 26% were specific to the mesopelagic ( $\text{O}_2$ -min, MSOW, and AAIW) while the bathypelagic water masses (NEADW and LDW) harbored 19% unique OTUs; 23% were found in both, the meso- and bathypelagic layer (Supporting Information Fig. S1).

### Alpha—Diversity

Sample coverage of the bacterioplankton was  $> 80\%$  in all samples, as indicated by rarefaction curves (Supporting Information Fig. S2). The Chao1 index, used as index for OTU-based *alpha*-diversity, changed significantly between the water masses (ANOVA,  $F_{5,44} = 20.33$ ,  $p < 0.001$ ) and biogeographic provinces (ANOVA,  $F_{2,47} = 17.75$ ,  $p < 0.001$ , Fig. 2A). On average, the MSOW, the AAIW and the NEADW exhibited a significantly higher diversity (Chao1 =  $118 \pm 8$ ) as compared to the SSL, the  $\text{O}_2$ -min and the LDW (Chao1 =  $87 \pm 20$ ). Additionally, *alpha*-diversity increased from the North Atlantic Drift Province to the North Atlantic Gyral Province (Tukey HSD,  $p < 0.001$ ; Fig. 2A, Supporting Information Table S1) while the influence of the water masses on diversity was independent from the biogeographic provinces.

The phylogenetic diversity (PD), however, increased from the SSL towards the NEADW (ANOVA,  $F_{5,44} = 13.99$ ,  $p < 0.001$ , Fig. 2B). Using phylogenetic diversity, we found



**Fig. 2.** Rarefied  $\alpha$ -diversity of the water masses and provinces. Filled symbols depict the mean, open symbols indicate the results of individual samples.

(A) Chao1 index, (B) phylogenetic diversity (PD), and (C) Pielou's evenness ( $J'$ ) in the different the water masses: SSL—subsurface layer (100 m depth),  $O_2$ -min—oxygen minimum, MSOW—Mediterranean sea outflow water. AAIW—Antarctic Intermediate Water; NEADW—North East Atlantic Deep Water, LDW—Lower Deep Water. The individual samples are indicated with the respective province: NADR—North Atlantic Drift Province. NAST—North Atlantic Subtropical Province. NATR—North Atlantic Tropical Gyral Province.

no significant influence of the biogeographic provinces on the phylogenetic diversity as determined by a two-way ANOVA (Fig. 2B).

Pielou's evenness ( $J'$ ) was significantly lower in the SSL compared to all the other water masses (ANOVA,  $F_{5,44} = 24.48$ ,  $p < 0.001$ ) and varied over a relatively narrow range from the mesopelagic to the bathypelagic water masses (Fig. 2C). Evenness of the North Atlantic Drift Province (NADR;  $0.800 \pm 0.006$ ) and the North Atlantic Subtropical Province (NAST;  $0.812 \pm 0.007$ ) were significantly lower than in the southern North Atlantic Tropical Gyral Province ( $0.851 \pm 0.006$ ; NATR; ANOVA,  $F_{2,47} = 12.71$ ,  $p < 0.001$ , Fig. 2C).

#### Beta—Diversity

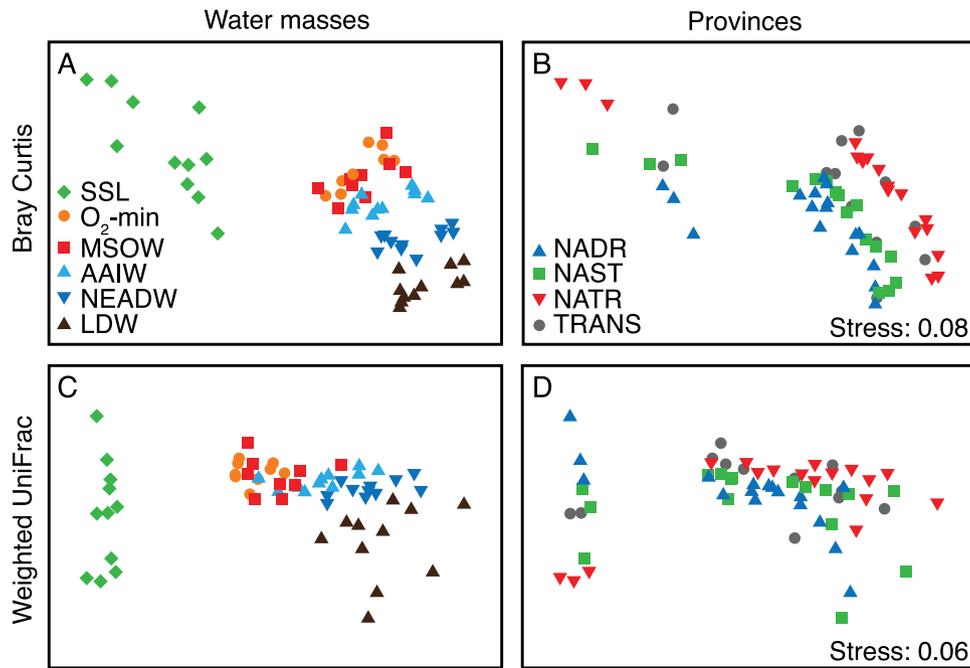
The communities clustered according to water masses as derived from the nonmetric multidimensional scaling (NMDS) plots based on Bray–Curtis similarity of the Hellinger-transformed OTU table (Fig. 3A). This strong effect of water masses was also supported by a two-way ANOSIM (global  $R = 0.85$ ,  $p < 0.001$ ). A similar clustering was obtained in plots based on weighted UniFrac (Fig. 3C).

A clear separation of the community according to biogeographic provinces was obtained between the northern provinces (NADR and NAST) and the most southern province (NATR) based on NMDS (Fig. 3B). However, less variability was explained as compared to the influence of water masses (two-way ANOSIM,  $R = 0.059$ ,  $p < 0.001$ ). Using weighted UniFrac analysis, the clustering according to biogeographic provinces in the NMDS seemed complex

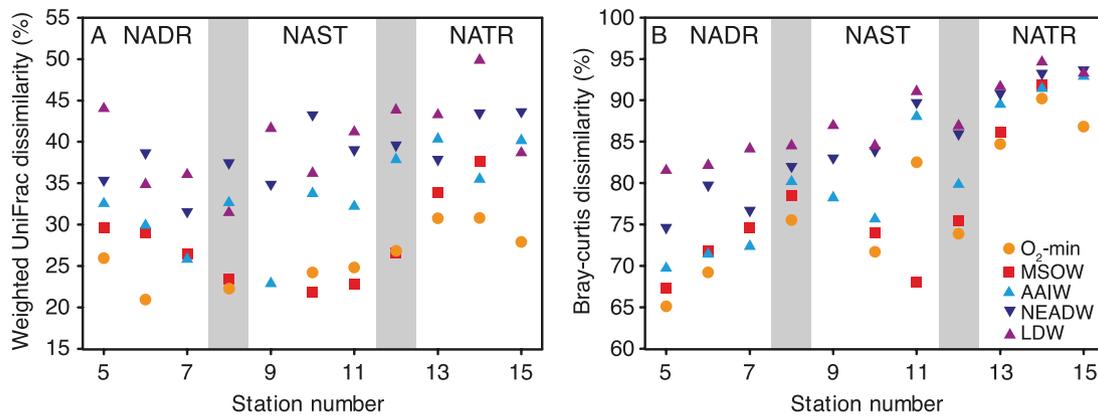
but was statistically significant [ANOSIM (NADR, NATR),  $R = 0.70$ ,  $p < 0.001$ ; Fig. 3D]. This clustering remained significant even after removing the rather strong influence of the SSL [ANOSIM (NADR, NATR),  $R = 0.63$ ,  $p < 0.001$ ].

Additionally, the OTU based analysis suggested that bacterial community composition changed within water masses between biogeographic provinces, as indicated by the uniform distances from the northern provinces (NADR and NAST) to the North Atlantic Tropical Province (NATR) for each water mass (Fig. 3A,B). Moreover, the weighted UniFrac analysis suggested that the latitudinal influence on the bacterial community composition was dependent on the specific water mass. For example, changes between provinces were minor for the  $O_2$ -min layer and more pronounced in the LDW. Similar to the OTU based analysis, analysis of similarity on the weighted UniFrac analysis indicated close resemblance of the community composition of  $O_2$ -min layer and MSOW, but also of the MSOW and the AAIW. In contrast to the OTU-based analysis, however, AAIW showed some similarity with the NEADW (ANOSIM,  $R = 0.38$ ,  $p < 0.05$ ), indicated by the overlapping clusters of both water masses (Fig. 3C).

We performed linear mixed effect models (LME) on the weighted UniFrac distance matrix. The Linear mixed effect models indicated a significant province-dependent change in the link between the bacterial community composition of the SSL and the underlying water masses (Fig. 4). The dissimilarity between the bacterial community composition of the SSL and the deep communities increased by  $\sim 21\%$  from the North Atlantic Drift Province to the North Atlantic Tropical Gyral Province (LME,  $\text{chi-square}_{10} = 27.0$ ,



**Fig. 3.** Nonmetric multidimensional scaling (NMDS) plots of community composition using Bray–Curtis similarity and bacterial OTUs (A, B) and weighted UniFrac dissimilarity (C, D). Samples are color coded according to water mass in (A) and (C) or biogeographic province in (B) and (D). SSL—subsurface layer (100 m depth),  $O_{2min}$ —oxygen minimum, MSOW—Mediterranean Sea Outflow Water. AAIW—Antarctic Intermediate Water; NEADW—North East Atlantic Deep Water, LDW—Lower Deep Water. The individual samples are indicated with the respective province: NADR—North Atlantic Drift Province. NAST—North Atlantic Subtropical Province. NATR—North Atlantic Tropical Gyral Province. TRANS indicates samples lacking distinct affiliation to one province.

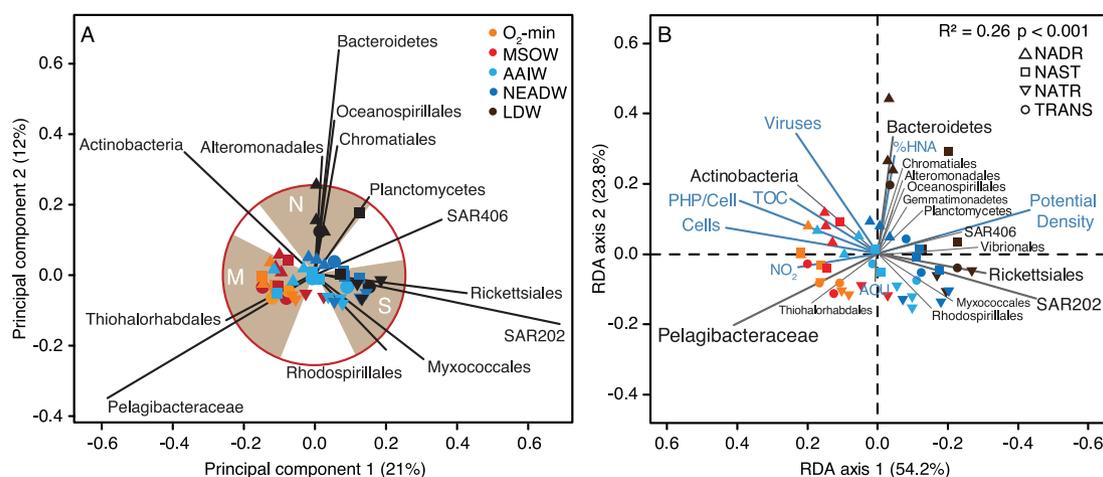


**Fig. 4.** Dissimilarity of water-mass specific communities to the epipelagic community composition given as weighted UniFrac (A) and the Bray–Curtis dissimilarity metric (B) at each station from north (station 5) to south (station 15) crossing three Provinces: NADR—North Atlantic Drift Province, NAST—North Atlantic Subtropical Province, NATR—North Atlantic Tropical Gyral Province. Grey bars indicate transition zones between provinces. Samples are color coded according to water mass.  $O_{2min}$ —oxygen minimum, MSOW—Mediterranean Sea Outflow Water, AAIW—Antarctic Intermediate Water, NEADW—North East Atlantic Deep Water, LDW—Lower Deep Water.

$p < 0.01$ ). This province-dependent change of similarity between the bacterial community composition of the SSL and the deep-water masses was least in the LDW (15%) and greatest in the AAIW (31%).

*Relative abundance of bacterioplankton phylotypes in the water masses*

The mean relative abundances of the dominant phylotypes in the water masses indicated a major contribution of



**Fig. 5.** Principal component analysis (PCA) and redundancy analysis (RDA) of meso- and bathypelagic phylotypes. In the PCA (A) only phylotypes contributing to the spreading more than the average (indicated by the red circle) are indicated. Brown pie slices within the circle indicate northern (N) and southern (S) regions of the bathypelagic, as well as the mesopelagic realm (M). In RDA (B), the significant environmental response-vectors are shown (downscaled by the factor 2). Vectors for species and constrained sites in the RDA are scaled to the square root of their Eigen value. Colors indicate water masses, symbol shapes correspond to provinces. The explained variation in percent is given for the two main axes, which are both highly significant. Abbreviation of the water masses and provinces as follows: SSL—subsurface layer (100 m depth), O<sub>2</sub>-min—oxygen minimum, MSOW—Mediterranean Sea Outflow Water, AAIW—Antarctic Intermediate Water, NEADW—North East Atlantic Deep Water, LDW—Lower Deep Water. The individual samples are indicated with the respective province: NADR—North Atlantic Drift Province, NAST—North Atlantic Subtropical Province, NATR—North Atlantic Tropical Gyral Province.

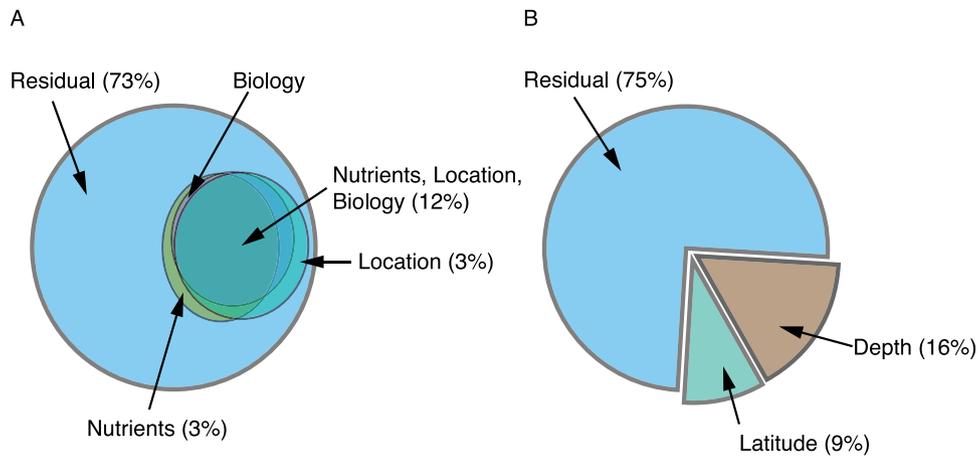
*Pelagibacteraceae* (*Alphaproteobacteria*), decreasing from 61% in the SSL to 39% in the LDW (Supporting Information Fig. S3). An unclassified member of the *Deltaproteobacteria* related to the SAR324 cluster (Sva0853) contributed substantially to the deep-water masses, i.e., in the range of 13–18%. The relative abundance of members of the SAR202 and SAR406 clusters increased with depth, contributing on average 8% to the bacterioplankton community in the NEADW and LDW (Supporting Information Fig. S3). Other groups such as the *Desulfobacterales*, *Thiohalorhabdadales* (*Gammaproteobacteria*) and the *Planctomycetes* were present in all water masses, albeit at low relative abundances (Supporting Information Figs. S3 and S4). In contrast to the NADR, the community changes in the southern province NATR were largely influenced by several groups of low abundance bacterial taxa (SAR406 cluster, *Rickettsiales*, SAR202 and *Pseudoalteromonadales*) as indicated by the RDA triplot depicting the correlation of phylotypes with environmental factors (Fig. 5B).

#### Distribution and variability of phylotypes

The distribution and relative dominance of phylotypes (as defined in the “Methods” section) across the water masses and biogeographic provinces were analyzed by principal component analysis (PCA). The SSL showed a distinctly different composition in bacterial phylotypes as compared to the deep-water masses and was dominated by *Pelagibacteraceae*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, *Rhodobacterales*, and *Oceanospirillaceae*. For evaluating

the changes in the meso- and bathypelagic bacterial community composition, the data of the SSL were removed from the analysis to increase the analytical resolution of the vertical and horizontal changes in relative phylotype abundance in the meso- and bathypelagic water masses along the transect. Two main axes were retained in the principal component analysis and accounted for 42% of the variation in the dataset (Fig. 5A). A potential third principal component axis was driven by the relative abundance of *Vibrionales* (*Gammaproteobacteria*) that peaked in the LDW south of the Azores (Station 11) at a relative abundance of 29% and sharply decreased to the north and to the south. PCA indicated latitudinal and water mass-related changes in bacterial community composition in the AAIW, NEADW, and LDW in addition to the varying dominance of specific phylotypes (Fig. 5A). Similar to weighted UniFrac, the latitudinal changes in the MSOW and O<sub>2</sub>-min layer were less pronounced compared to the deeper water masses.

Overall, PCA indicated a decreasing relative abundance of *Pelagibacteraceae* with depth, representing a major contribution to the changing bacterial community composition from the meso- to the bathypelagic water masses (Fig. 5A). Similarly, the *Thiohalorhabdadales* significantly decreased with depth, while the SAR406 cluster significantly increased with depth. In the NEADW and LDW of the North Atlantic Drift Province, mainly subgroups of *Bacteroidetes* (*Flavobacteria*), *Planctomycetes* (*Phycisphaera*), and *Gammaproteobacteria* (*Alteromonadales*, *Oceanospirillales*, *Chromatiales*) structured the bacterioplankton.



**Fig. 6.** Venn diagram and pie chart of the variance partitioning analysis. A model of bacterial phylotypes including the complete set of environmental parameters (A) and a reduced model using depth and latitude (B) is shown. The areas correspond to the amount of variance explained by each factor. Overlapping areas indicate shared variation of the parameter effect on community composition in (A).

Despite exhibiting a lower relative abundance than the former, some *Alphaproteobacteria* (i.e., *Rickettsiales* excluding SAR11) together with members of the SAR202 cluster were responsible for the variability in the bathypelagic community composition in the southern provinces NAST and NATR. *Actinobacteria* (*Acidimicrobia*) were relatively more important in the northern part of the transect while *Rhodospirillales*, and *Deltaproteobacteria* (*Myxococcales*) showed a relatively higher abundance toward the south. These groups substantially influenced the latitudinal changes of the bacterial community composition in the mesopelagic water masses (Fig. 5A).

#### *Environmental parameters influencing bacterial community composition*

Changes in the bacterial community composition in relation to the environment were explored by redundancy analysis (RDA). Bacterial phylotypes in the deep-water masses were positively correlated with several environmental and biotic parameters (Fig. 5B). RDA indicated a correlation of total organic carbon (TOC) with latitude despite its rather low variability along the transect (Supporting Information Table S2). Additionally, the increasing TOC concentrations towards the high latitudes correlated with the increasing abundance of *Actinobacteria* and *Thiohalorhabdales* within the mesopelagic waters (Fig. 5B). Particularly in the North Atlantic Drift Province, the percentage of high nucleic acid cells (HNA) increased with depth and correlated with *Alteromonadales*, *Bacteroidetes*, *Oceanospirillales*, and *Chromatiales* (compare Fig. 5A and B). *Actinobacteria*, dominating in the North Atlantic Drift Province, correlated with viral abundance that, in turn, generally decreased towards the NAST and NATR in the meso- and bathypelagic water masses. RDA also suggested a link between

increasing cell-specific heterotrophic production in the O<sub>2</sub>-min and the MSOW and the relative abundance of *Thiohalorhabdales*.

Variance partitioning of the RDA showed that 27% of the total variability in the community composition below the SSL were explained by the measured environmental variables (Fig. 6A). A rather small fraction of the variability (3% each) was exclusively associated with location (geographic position and depth) or chemical parameters (salinity, O<sub>2</sub>, AOU, TOC, PO<sub>4</sub>, SiO<sub>4</sub>, and NO<sub>3</sub>). Adding biological parameters (viral and prokaryotic abundance, percentage of HNA, ectoenzyme activity and cell-specific leucine incorporation) did not increase the explained overall variance. A more reduced model using depth and latitude had a similar explanatory power (25%) as compared to the more complex model including all measured variables (Fig. 6B).

## Discussion

### *Prokaryotic diversity in the deep North Atlantic*

In the subsurface layer, an increase in OTU-based bacterial *alpha*-diversity was detected from the North Atlantic Drift Province to the North Atlantic Tropical Province (Fig. 2A). This was also reported in a global study on bacterial richness, where an increasing diversity towards lower latitudes has been found (Fuhrman *et al.*, 2008). Also, a meta-analysis showed significant latitudinal gradients in bacterial diversity, both in the southern and northern hemisphere (Sul *et al.*, 2013).

The richness-based *alpha*-diversity indices have been questioned, however, because changes in the community composition cannot be derived from OTU abundances alone (Hanson *et al.*, 2012). Thus, in an attempt to circumvent some of the ambiguities concerning current OTU-based *alpha*-diversity measures, we additionally analyzed

the data using phylogenetic diversity (Faith, 1992). In contrast to OTU richness, this approach showed that diversity increased with depth and decreased with temperature (Fig. 2B) coinciding with a decreasing productivity over the broad scale of environmental parameters obtained from the lower epipelagic towards the LDW. It has been shown that evenness supports functional stability of ecosystems (Wittebolle *et al.*, 2009). In our study evenness indicated a similar pattern in all three provinces with relatively low values in the SSL and higher evenness throughout much of the deep-water column. A higher taxonomic evenness in the bathypelagic, with an overall higher variability was also reported in earlier studies (Hewson *et al.*, 2006; Agogu e *et al.*, 2011).

Beta-diversity is characterized by partitioning diversity among communities or along environmental gradients using the number of shared species between two communities (Lozupone and Knight, 2008). Mixed effect models on the dissimilarity between the epipelagic and the respective deep-water communities suggested that the changes in the prokaryotic community composition of the deep-water masses are connected to the latitudinal changes in the epipelagic community of the different biogeographic provinces. Furthermore, the relation between the epipelagic bacterial communities and the respective deep-water mass associated community was dependent on the province, suggesting a variable degree of connectivity of surface water and dark ocean community composition (Figs. 3A,B and 4B). Prokaryotic communities might evolve with ageing of a particular water mass and the associated changes in the physicochemical characteristics. However, no such effect was found in this study. In contrast, a low but significant similarity was obtained between communities in the same water mass at different locations along the transect (Fig. 3). Thus, our data suggest that dispersal limitation is decoupling the communities of adjacent water masses, while the productivity of the surface community is modulating the intensity of the epipelagic influence on the deep-water masses.

In contrast to methods that are purely based on species counts, weighted Unifrac seems robust to variations in the region sequenced (Liu *et al.*, 2007) and also chimeras have been shown to not affect network clustering of Unifrac distances (Ley *et al.*, 2008). The major pattern of an increasing north to south dissimilarity between the SSL communities and the bathypelagic water masses was also revealed analyzing the weighted Unifrac distance plots (Fig. 3C and D). We suggest that differences between weighted UniFrac and Bray–Curtis distances are due to changes in phylogenetic affiliation not resolved in the analysis using OTUs. It has been proposed that phylogenetic similarity and functional traits are connected to a certain degree (Philippot *et al.*, 2010; Ortmann and Ortell, 2014) although the extent of the link between phylogeny and metabolic function is still under debate (Prosser *et al.*, 2007). In our study, the

weighted Unifrac plots indicated a high similarity of the AAIW and the MSOW/O<sub>2</sub>-min layer in the North Atlantic Drift Province, while in the oligotrophic gyre (NATR) the AAIW was more similar to the NEADW. Similarly, the linear mixed effect model also suggested a higher connectivity of the AAIW to the SSL in the NADR as compared to the NATR. This might indicate relatively less dispersal of the surface bacterial community into the mesopelagic waters in the subtropical gyre than at higher latitudes where export flux of particles is higher and water column stratification less intense (Buesseler and Boyd, 2009). A recent study suggests that the particle-association lifestyle is conserved within the phylogeny of the bathypelagic prokaryotes (Salazar *et al.*, 2015). Therefore, the apparent change in connectivity between epipelagic and mesopelagic waters in the different provinces might be due to environmental filtering or directly related to province-specific particle flux and its attached microbial community.

#### *Biogeography of abundance categories*

Typically, natural prokaryotic communities are characterized by a large number of low abundance OTUs coined the 'rare biosphere' (*sensu* Sogin *et al.*, 2006). This rare biosphere might constitute a seed bank and become abundant once environmental conditions change (Pedros-Alio, 2006). Additionally, some members of the rare biosphere may influence biogeochemical cycles disproportionately to their (low) abundance (Fuhrman, 2009). To test the impact of the rare bacterial biosphere on the changes in diversity, we assigned the OTUs to three different groups according to their relative abundance. Only ten OTUs exceeded at least once a relative abundance of 10% of the per-sample total OTU counts and thus were categorized as abundant. In total, 122 OTUs constituted between 1 and 10% of the bacterial community, categorized as common. Based on the sequencing effort (~1,500 sequences per sample), we considered a relative abundance of <1% as rare OTUs. This low abundance fraction was by far the most diverse category comprising ~1,410 bacterial OTUs (data not shown). The general resemblance pattern between deep-water communities was similar for the three abundance categories (Mantel test based on weighted Spearman:  $R > 0.8$ ,  $p < 0.001$  for all three abundance categories combined; Supporting Information Table S3). This finding is in agreement with a report on Arctic bacterioplankton where the rare Bacteria exhibited a comparable biogeographic distribution pattern as the abundant taxa (Galand *et al.*, 2009). However, low abundance taxa are apparently more important for the overall changes in community composition in deep-water masses than the abundant taxa (Supporting Information Table S3 and Supporting Information Fig. S5), a finding also reached by Agogu e *et al.* (2011).

*Distribution patterns of bacterial phylotypes*

Similar to other reports from the region (Schattenhofer *et al.*, 2009; Agogue *et al.*, 2011), the ubiquitous SAR11 clade (*Pelagibacteraceae*) dominated the relative abundance of the prokaryotic community in the sub-surface layer and decreased to the LDW (Supporting Information Fig. S3). The relative abundance of *Bacteroidetes* increased in the LDW flowing from the south to the north and correlated with the percentage of high nucleic acid (HNA) containing cells (Fig. 5B). *Bacteroidetes* are known to be associated with phytoplankton blooms and marine snow particles (DeLong *et al.*, 1993; Rath *et al.*, 1998; Fandino *et al.*, 2001). Thus, it seems likely that in the productive northern North Atlantic *Bacteroidetes* is transported into the deep waters with senescent and sedimenting phytoplankton blooms and marine snow. Members of the *Oceanospirillales* were mainly encountered in the LDW of the North Atlantic Drift Province (Fig. 5 and Supporting Information Fig. S4). Recently piezophilic *Oceanospirillales* have been isolated, thriving at depth between 2000 and 7000 m (Cao *et al.*, 2014), and some members of this group also contain RuBisCO and sulfur-oxidation genes that may enable them to switch between an auto- and mixotrophic lifestyle (Swan *et al.*, 2011). Thus, fast sinking particles may provide microscale sub- and anoxic habitats supporting an autotrophic lifestyle based on sulfur-oxidation (Karl *et al.*, 1984; Shanks and Reeder, 1993). Together with *Alteromonadales* that are abundant in the North Atlantic Drift Province, these two groups seem to prefer high productivity regions with relatively fresh input from the epipelagic layer or alternatively, relatively fresh material from the formation site of the water masses. The uncultured deltaproteobacterial group Sva0853 (Supporting Information Fig. S4), belonging to the SAR324 cluster, was abundant in the bathypelagic realm as it contributed up to 20% to the communities in the AAIW and NEADW; it was also found abundantly at the Pacific Hawaii Ocean Time-Series (HOT) station in a similar depth range (Shi *et al.*, 2011). Members of the *Thiohalorhabdales*, a hitherto unknown group for marine pelagic waters, were relatively abundant in all deep-water masses ( $\geq 2\%$ ) (Supporting Information Figs. S3 and S4). *Thiohalorhabdales* were only recently isolated from a hypersaline lake and are characterized as facultatively anaerobic and obligate sulfur-oxidizing bacteria (Sorokin *et al.*, 2008). Interestingly, we found a correlation of the relative contribution of *Thiohalorhabdales* to the total bacterial community with nitrite concentration (Fig. 5B). This supports the notion, that some of their members are denitrifiers with a possible function in the sulfur cycling of the mesopelagic layer.

**Conclusion**

Our data suggest that water masses are strong predictors of the community composition in the deep North Atlantic, however, Longhurst provinces have an impact on the diver-

sity of the bacterial community in different deep-water masses of the North Atlantic. Probably not unexpected, the *Pelagibacteraceae* were the most abundant group in all water masses and all provinces. We found that members of the *Bacteroidetes*, *Chromatiales*, *Alteromonadales*, *Oceanospirillales* and *Gemmatimonadetes* mainly occur in the northern more productive NADR. However, toward the southern, less productive North Atlantic Tropical Gyral province the *Rickettsiales*, the SAR202-cluster and the *Myxococcales* as well as the *Rhodospirillales* increased in relative abundance. Furthermore, our analyses indicate that the connectivity of the bacterial communities between the surface and the dark ocean is more pronounced for the mesopelagic than the bathypelagic water masses. The strength of this connectivity likely depends on the interaction between horizontal dispersal by advection, the magnitude of particle flux and the strength of water column stratification. The biogeographic distribution of the rare taxa is comparable to the more abundant groups. However, the low-abundant taxa are more important for the overall changes in the bacterial community composition in the deep-water masses than the more abundant taxa. Thus taken together, the dark ocean microbial biogeography is far more complex than hitherto assumed.

**Experimental procedures***Sampling sites*

During the MEDEA-1 research cruise (October 2011), a 2,840 km long transect following the eastern branch of the North Atlantic Deep Water (NEADW) was performed from 48.75°N, 22.51°W to 26.06°N, 36.12°W crossing three biogeographic provinces as defined in Longhurst (2007): the North Atlantic Drift Province (NADR), the North Atlantic Subtropical Province (NAST) and the North Atlantic Tropical Gyral Province (NATR, Fig. 1). Satellite derived net primary production (NPP), averaged over the time of the cruise, was used to set the flexible borders of the provinces to best fit their described characteristics in Longhurst (2007). NPP rates for each sampling day were retrieved from the ocean productivity website (<http://www.science.oregonstate.edu/ocean.productivity>) using the Vertically Generalized Production Model (VGPM) of Behrenfeld and Falkowski (1997).

To determine the prokaryotic community composition and environmental parameters, samples at 11 stations and 6 depths ranging from 100 to 5,000 m were collected using 19 × 25 liter Niskin bottles (from Ocean Test Equipment) mounted in a rosette frame equipped with calibrated conductivity-temperature-depth sensors (CTD; Sea-Bird electronic SBE911plus). The factory calibration of the conductivity sensor was checked for drift with on board measurements of salinity samples and standards using a Guildline 8400B salinometer. The deep-water masses were identified by their salinity-temperature characteristics and samples were collected from the epipelagic 100 m depth layer, subsequently coined subsurface layer (SSL); from the mesopelagic, the oxygen minimum layer (O<sub>2</sub>-min), the Mediterranean Sea Outflow

Water (MSOW) and the Antarctic Intermediate Water (AAIW) were sampled. From the bathypelagic zone, water from the North East Atlantic Deep Water (NEADW) and the Lower Deep Water (LDW) was collected. A detailed description of the physico-chemical properties of the individual water masses encountered in our study region is given in Supporting Information Table S2.

### Environmental parameters

Oxygen concentrations were measured with an oxygen sensor (SBE43, Seabird Electronics) mounted on the CTD and calibrated daily with discrete oxygen measurements using a spectrophotometric Winkler approach (Reinthal *et al.*, 2006). Apparent oxygen utilization (AOU) was calculated as the difference between the oxygen concentration at saturation level and the observed oxygen concentration. AOU was calculated with the software package Ocean Data View (Schlitzer, 2006) using the output of the calibrated CTD and the O<sub>2</sub> sensor.

Determination of inorganic nutrient concentrations followed Joint Global Ocean Fluxes Study recommendations (Gordon *et al.*, 1993). On board, the concentrations of the dissolved inorganic nutrients SiO<sub>4</sub>, NO<sub>3</sub>, NO<sub>2</sub>, PO<sub>4</sub><sup>3-</sup> were determined using an autoanalyzer (TRAACS, Technicon) immediately after collecting and filtering the samples through pre-rinsed 0.2 µm filters (Acrodisc, Pall Corporation).

For total organic carbon (TOC) measurements, aliquots of 10 mL seawater were transferred into combusted glass ampoules. Subsequently, the pH was adjusted to <2 with concentrated phosphoric acid. Thereafter, the ampoules were sealed and stored frozen at -20°C until analysis in the lab. TOC concentrations were determined by a high temperature combustion method using a Shimadzu TOC 5000A analyzer (Benner and Strom, 1993). The TOC concentrations were calculated from quadruplicate sample injections compared to a three-point standard curve prepared with potassium hydrogen phthalate. The instrument's performance and the validity of the calibration were determined using reference material of the Hansell consensus reference materials program (44–46 µmol L<sup>-1</sup> for the reference samples; *n* = 3 and 1–2 µmol L<sup>-1</sup> for low carbon water; *n* = 3). The average analytical precision of the instrument was <3%.

### Prokaryotic abundance and heterotrophic production

Prokaryotic abundance was determined as previously described (Marie *et al.*, 1997). Briefly, duplicate seawater samples were fixed with electron microscopy grade glutaraldehyde (0.5% final concentration), flash-frozen and stored at -80°C until analysis. In the home laboratory, the samples were thawed, stained with SYBR Green I (Life Technologies) in the dark for 10 min and counted on a FACS Aria II flow cytometer (BD Biosciences). Prokaryotes were analyzed on cytograms of side scatter *versus* green fluorescence and high nucleic acid containing cells (HNA) were distinguished from low nucleic acid containing cells (LNA) by gating (Gasol and Del Giorgio, 2000).

Leucine incorporation was measured as described in Reinthal *et al.* (2010). Briefly, <sup>3</sup>H-leucine was added to the seawater samples to a final concentration of 5–10 nM. Follow-

ing incubation at *in situ* temperature in the dark for 4–24 h, samples were filtered on to 0.2-µm polycarbonate filters (Nuclepore, Whatman). Filters were extracted twice with 5% ice-cold trichloroacetic acid and subsequently dried in 20 mL scintillation vials. Eight mL scintillation cocktail (FilterCount, Canberra-Packard) was added to the samples and after 18 h, filters were counted in a liquid scintillation counter (Tricarb 3100TR, Perkin Elmer). Prokaryotic heterotrophic production was calculated from the blank-corrected leucine incorporation rates assuming a theoretical conversion factor of 1.55 kg mol<sup>-1</sup> leucine (Kirchman, 1993). A detailed description of the biological parameters of the individual water masses encountered in our study region is given in Supporting Information Table S4.

### DNA extraction and sequencing

Samples for DNA analysis were collected from six depth layers at each station. Ten liter of seawater of each depth were filtered through 0.2-µm polycarbonate filters (47 mm diameter, Millipore). Filters were flash-frozen in liquid nitrogen for 10 min and then stored at -80°C until further processing. Back in the lab, the filters were cut into small pieces to increase the subsequent lysis efficiency using sterile scissors. DNA from filter sections was extracted using the UltraClean Soil DNA Isolation Kit according to recommendations (MoBio). The commercial kit includes lysis, bead-based homogenization by vortexing, and DNA cleanup with silica membranes in microfuge tubes. Tag sequencing and prior PCR-amplification were performed at the Research and Testing Laboratory (Lubbock, TX) on a Roche 454 FLX platform using Titanium series reagents (Dowd *et al.*, 2008). Primers 28F (GAGTTTGATCNTGGCTCAG) and 519R (GTNTTACNGCGGCKGCTG) were used for amplification of the variable regions V1-3 for Bacteria (Fan *et al.*, 2012). Per sample ~1,500 sequences with an average read length of ~370 bp were obtained. After quality filtering and denoising following Reeder and Knight (2010), the sequences were analyzed with QIIME 1.6.0 (Caporaso *et al.*, 2010) using the *de novo* OTU picking workflow described in detail on the QIIME website. Adopting a 97% sequence similarity threshold, operational taxonomic units (OTUs) were assigned with the Greengenes 12.0 database using the RDP classifier implemented in QIIME. Relative OTU abundances were calculated from the obtained OTUs normalized to the total OTU counts.

Pyrotag sequences have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under bioproject number PRJNA262973.

### Alpha- and beta-diversity assessment

To constrain the diversity assessment, we applied two approaches to estimate *alpha*- and *beta*-diversity by using the occurrence and relative abundance of OTUs as well as measures of sequence similarity serving as proxies for phylogenetic distinctness between communities.

The Chao1 index was used to assess OTU richness of samples. Samples were rarefied using the web-based program iNEXT to account for varying sampling depth (Hsieh, 2013). iNEXT uses the occurrence of singletons and doubletons to estimate completeness of all the samples, identifying the sample with the lowest species coverage. Thereafter, the

samples are compared at equal species completeness, which is a more sensitive approach than rarefaction based on sample size to uncover relative differences in richness between the communities (Chao and Jost, 2012).

Divergence-based *alpha*-diversity was determined by calculating phylogenetic diversity (PD) in QIIME (Faith, 1992). A tree was constructed using FastTree (Price *et al.*, 2009) on Lane-mask filtered sequences (Lane, 1991). The resulting sequences were aligned with the PyNAST algorithm (Caporaso *et al.*, 2010). PD considers the branch lengths between different species in each sample. The PD values were compared at similar sample size using rarefaction curves.

OTU based *beta*-diversity was analyzed in PRIMER6 (Clarke, 1993). Relative OTU abundances were square-root transformed resulting in linear species rank-abundance curves. Nonmetric multidimensional scaling plots (NMDS) and analyses of similarity (ANOSIM) with at least 999 permutations were performed on Bray–Curtis similarity matrices.

For divergence-based *beta*-diversity, weighted Unifrac analysis was used. In brief, weighted Unifrac analysis results in a similarity matrix by using comparisons of branch length differences between communities (samples) in a phylogenetic tree (Lozupone and Knight, 2005). Stronger weights are given to branches with higher relative abundances. Thus, phylogenetically more similar communities are clustered more closely together than phylogenetically more distant communities.

Additionally, linear mixed effect models (LME) were used to investigate the influence of biogeographic provinces on the dissimilarity of the deep-water communities to its overlying SSL (for details see Supporting Information).

#### *Analysis of environmental influence on community composition*

For analysis of environmental parameters influencing the community composition, the OTUs were phylogenetically classified. The original OTUs were clustered into groups of different taxonomic ranks including order, class and family. The result was subsequently compared to weighted Unifrac similarity using Mantel tests (Mantel, 1967). Clustering with the highest similarity to weighted Unifrac ( $\rho = 0.87$ ) resulted in 129 bins, which are referred to as phylotypes. To reveal the major environmental factors and their contribution to the encountered community composition variance partitioning was performed on the Hellinger-transformed dataset using the varpart function of vegan (Oksanen *et al.*, 2013).

#### *Additional statistical analysis*

After validation of homoscedasticity, analysis of variance (ANOVA) followed by Tukey's HSD tests was used to examine significant differences between factors. If homoscedasticity was not achieved, Mann–Whitney *U* tests with Bonferroni–Holm correction were performed to determine statistically homogenous groups. Community composition and its environmental forcing was analyzed by principal component analyses (PCA) and redundancy analyses (RDA) using Hellinger-transformed data, essentially following the recommendations of Legendre and Legendre (1998). The significance of the principal component axes was checked with the broken stick

model (Frontier, 1976). Redundancy analysis can be understood as a direct extension of multiple regression to modeling of multivariate response data and as a canonical extension of principal component analysis (Legendre and Legendre, 1998). It has been argued that in ecological studies, redundancy analysis is more robust than canonical correlation analysis (CCA), especially when the appearance or disappearance of a certain taxa is of no exceptional importance for the data set (Legendre and Gallagher, 2001). One major problem in the statistical analysis with many oceanographic datasets is co-correlation of environmental factors with depth and latitude. However, through stepwise reduction of the constraining variables, we obtained a reasonable amount of unbiased factors describing the dataset without redundancy. LME, PCA, RDA, variance partitioning, and multiple single factor analyses were carried out using R 2.15.0 (R Core Team, 2013). For details on the used packages see the supplementary information.

#### **Acknowledgements**

We thank the captain and the crew of the R/V Pelagia for their help during work at sea. The technical and analytical support of Santiago Gonzalez and Jan van Ooijen from the Netherlands Institute for Sea Research (NIOZ) is highly appreciated. We are indebted to Christian Baranyi for his help on board and technical advice in the lab. Many thanks go to Franziska Eibenberger for her analytical help and enthusiasm and to Stefan Dullinger for his advices on the linear mixed effect models. Part of the shiptime and lab work was funded by the Austrian Science Fund (FWF) project I486-B09 and by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement No. 268595 (MEDEA project) both to G.J.H. and by the Austrian Science Fund (FWF) through the PADOM project P23221-B11 to T.R. This work is in partial fulfillment of the requirements for a Ph.D. degree from the University of Vienna to A.H.F.

#### **References**

- Agogue, H., Lamy, D., Neal, P.R., Sogin, M.L., and Herndl, G.J. (2011) Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol Ecol* **20**: 258–274.
- Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., and Thingstad, F. (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10**: 257–263.
- Behrenfeld, M.J., and Falkowski, P.G. (1997) Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol Oceanogr* **42**: 1–20.
- Benner, R., and Strom, M. (1993) A critical-evaluation of the analytical blank associated with doc measurements by high-temperature catalytic-oxidation. *Mar Chem* **41**: 153–160.
- Bochdansky, A.B., van Aken, H.M., and Herndl, G.J. (2010) Role of macroscopic particles in deep-sea oxygen consumption. *Proc Natl Acad Sci USA* **107**: 8287–8291.
- Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bissett, A., Lauro, F.M. *et al.* (2009) Microbial community structure in the North Pacific Ocean. *ISME J* **3**: 1374–1386.
- Buesseler, K.O., and Boyd, P.W. (2009) Shedding light on processes that control particle export and flux attenuation in

- the twilight zone of the open ocean. *Limnol Oceanogr* **54**: 1210–1232.
- Cao, Y., Chastain, R.A., Eloë, E.A., Nogi, Y., Kato, C., and Bartlett, D.H. (2014) Novel psychropiezophilic Oceanospirillales species *Profundimonas piezophila* gen. nov., sp. nov., isolated from the deep-sea environment of the Puerto Rico Trench. *Appl Environ Microbiol* **80**: 54–60.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Chao, A., and Jost, L. (2012) Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology* **93**: 2533–2547.
- Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117–143.
- Cram, J.A., Xia, L.C., Needham, D.M., Sachdeva, R., Sun, F., and Fuhrman, J.A. (2015) Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *ISME J*. 10.1038/ismej.2015.76.
- De Corte, D., Sintes, E., Yokokawa, T., Reinthaler, T., and Herndl, G.J. (2012) Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. *ISME J* **6**: 1566–1577.
- DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924–934.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.
- Fan, L., McElroy, K., and Thomas, T. (2012) Reconstruction of ribosomal RNA genes from metagenomic data. *PLoS One*. doi:10.1371/journal.pone.0039948.
- Fandino, L.B., Riemann, L., Steward, G.F., Long, R.A., and Azam, F. (2001) Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. *Aquat Microb Ecol* **23**: 119–130.
- Frontier, S. (1976) Decrease of Eigenvalues in principal component analysis—Comparison with broken stick model. *J Exp Mar Biol Ecol* **25**: 67–75.
- Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.
- Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., et al. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774–7778.
- Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature* **459**: 193–199.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106**: 22427–22432.
- Galand, P.E., Potvin, M., Casamayor, E.O., and Lovejoy, C. (2010) Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J* **4**: 564–576.
- Ghigliione, J.F., Galand, P.E., Pommier, T., Pedros-Alio, C., Maas, E.W., Bakker, K., et al. (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc Natl Acad Sci USA* **109**: 17633–17638.
- Gordon, L.I., Jennings, J.C.J., Ross, A.A., and Krest, J.M. (1993) A suggested protocol for continuous flow automated analysis of seawater nutrients (phosphate, nitrate, nitrite and silicic acid) in the WOCE hydrographic program and the Joint Global Ocean Fluxes Study. Corvallis, Oregon, USA: College of Oceanic and Atmospheric Sciences, Oregon State University, p 55.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Hellweger, F.L., van Sebille, E., and Fredrick, N.D. (2014) Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* **345**: 1346–1349.
- Herndl, G.J., and Reinthaler, T. (2013) Microbial control of the dark end of the biological pump. *Nat Geosci* **6**: 718–724.
- Hewson, I., Steele, J.A., Capone, D.G., and Fuhrman, J.A. (2006) Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol Oceanogr* **51**: 1274–1283.
- Hsieh, T.C., Ma, K.H., and Chao, A. (2013) iNEXT online: Interpolation and extrapolation (Version 1.3.0) <http://chao.stat.nthu.edu.tw/next/>. Accessed on May 8, 2013.
- Karl, D.M., Knauer, G.A., Martin, J.H., and Ward, B.B. (1984) Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* **309**: 54–56.
- Kirchman, D.L. (1993) Leucine incorporation as a measure of biomass production by heterotrophic bacteria. In *Handbook of Methods in Aquatic Microbial Ecology*. Kemp, P.F., Sherr, B.F., Sherr, E.B., Cole, J.J. (eds) Boca Raton, USA: Lewis Publishers, pp. 509–512.
- Lane, D.J. (1991) 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*. Stackebrandt, E., Goodfellow, M. (eds) New York: Wiley, pp. 115–175.
- Legendre, P., and Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.
- Legendre, P., and Legendre, L. (1998) *Numerical Ecology*, 2nd english edn, Vol. 20. Amsterdam: The Netherlands.
- Ley, R.E., Hamady, M., Lozupone, E.C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., et al. (2008) Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.
- Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., Knight, R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res*. doi: 10.1093/nar/gkm541.
- Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C. (1995) An estimate of global primary production in the ocean from satellite radiometer data. *J Plankton Res* **17**: 1245–1271.
- Longhurst, A. (2007) *Ecological Geography of the Sea*, 2nd edn. Amsterdam, The Netherlands: Academic Press.
- Lozupone, C., and Knight, R. (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.
- Lozupone, C.A., and Knight, R. (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* **32**: 557–576.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

- Marie, D., Partensky, F., Jacquet, S., and Vaultot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microbiol* **63**: 186–193.
- Martiny, J.B., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: Putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.
- Moeseneder, M.M., Smith, K.L., Ruhl, H.A., Jones, D.O.B., Witte, U., and Prosser, J.I. (2012) Temporal and depth-related differences in prokaryotic communities in abyssal sediments associated with particulate organic carbon flux. *Deep Sea Res Part I* **70**: 26–35.
- Nagata, T., Tamburini, C., Aristegui, J., Baltar, F., Bochdansky, A.B., Fonda-Umani, S., *et al.* (2010) Emerging concepts on microbial processes in the bathypelagic ocean—Ecology, biogeochemistry, and genomics. *Deep Sea Res Part II* **57**: 1519–1536.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al.* (2013) vegan: Community Ecology Package. <http://CRAN.R-project.org/package=vegan>.
- Oliver, M.J., and Irwin, A.J. (2008) Objective global ocean biogeographic provinces. *Geophys Res Lett.* doi:10.1029/2008GL034238.
- Ortmann, A.C., and Ortell, N. (2014) Changes in free-living bacterial community diversity reflect the magnitude of environmental variability. *FEMS Microbiol Ecol* **87**: 291–301.
- Pedros-Alio, C. (2006) Marine microbial diversity: Can it be determined? *Trends Microbiol* **14**: 257–263.
- Philippot, L., Andersson, S.G., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., and Hallin, S. (2010) The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* **8**: 523–529.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.
- Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al.* (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**: 384–392.
- R Core Team (2013) R: A language and environment for statistical computing. <http://www.R-project.org/>.
- Rath, J., Wu, K.Y., Herndl, G.J., and De Long, E.F. (1998) High phylogenetic diversity in a marine-snow-associated bacterial assemblage. *Aquat Microb Ecol* **14**: 261–269.
- Reeder, J., and Knight, R. (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7**: 668–669.
- Reinthal, T., Bakker, K., Manuels, R., Van Ooijen, J., and Herndl, G.J. (2006) Fully automated spectrophotometric approach to determine oxygen concentrations in seawater via continuous-flow analysis. *Limnol Oceanogr Meth* **4**, 358–366.
- Reinthal, T., van Aken, H.M., and Herndl, G.J. (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Res Part II* **57**: 1572–1580.
- Salazar, G., Cornejo-Castillo, F.M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D. *et al.* (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Mol Ecol* **24**: 5692–5706
- Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078–2093.
- Schlitzer, R (2011) Ocean Data View 4, <http://odv.awi.de>.
- Shanks, A.L., and Reeder, M.L. (1993) Reducing microzones and sulfide production in marine snow. *Mar Ecol Prog Ser* **96**: 43–47.
- Shi, Y., Tyson, G.W., Eppley, J.M., and DeLong, E.F. (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* **5**: 999–1013.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Sorokin, D.Y., Tourova, T.P., Galinski, E.A., Muyzer, G., and Kuenen, J.G. (2008) *Thiohalorhabdus denitrificans* gen. nov., sp nov., an extremely halophilic, sulfur-oxidizing, deep-lineage gammaproteobacterium from hypersaline habitats. *Int J Syst Evol Microbiol* **58**: 2890–2897.
- Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, L.A., and Sogin, M.L. (2013) Marine bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci USA* **110**: 2342–2347.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D. *et al.* (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.
- Vergin, K.L., Done, B., Carlson, C.A., and Giovannoni, S.J. (2013) Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *Aquat Microb Ecol* **71**: 1–13.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., *et al.* (2009) Initial community evenness favours functionality under selective stress. *Nature* **458**: 623–626.

## Supporting information

Additional Supporting Information may be found in the online version of this article.

**Table S1.** Diversity indices from different water mass and biogeographic province

**Table S2.** Mean physico-chemical parameters of the water masses and biogeographical provinces.

**Table S3.** Statistics of rare, abundant and common bacteria

**Table S4.** Mean biological parameters of the water masses and biogeographical provinces.

**Table S5.** Probe sequences and formamide concentrations used for CARD-FISH

**Fig. S1.** Shared and unique OTUs in the different pelagic realms.

**Fig. S2.** Rarefaction and coverage plots the richness estimator Chao1

**Fig. S3.** Water mass-specific bacterial phylotypes.

**Fig. S4.** Heat-plot of all bacterial groups in the different water masses

**Fig. S5.** NMDS of the different bacterial abundance classes.

## **Supplementary Information**

Connectivity between surface and deep waters determines prokaryotic diversity in the North Atlantic Deep Water

Alexander H. Frank, Juan A. Garcia, Gerhard J. Herndl, Thomas Reinthaler

## **Supplementary Methods**

**Linear mixed effect models**

**CARD-FISH procedure**

**R packages used**

**Evaluation of the sequencing approach**

## **Supplementary Tables**

Table S1 **Diversity indices from different water mass and biogeographic province**

Table S2 **Mean physico-chemical parameters of the water masses and biogeographical provinces.**

Table S3 **Statistics of rare, abundant and common bacteria**

Table S4 **Mean biological parameters of the water masses and biogeographical provinces.**

Table S5 **Probe sequences and formamide concentrations used for CARD-FISH**

## **Supplementary Figures**

Figure S1 **Shared and unique OTUs in the different pelagic realms.**

Figure S2 **Rarefaction and coverage plots the richness estimator Chao1**

Figure S3 **Water mass-specific bacterial phylotypes.**

Figure S4 **Heat-plot of all bacterial groups in the different water masses**

Figure S5 **NMDS of the different bacterial abundance classes.**

## **Supplementary References**

## Methods

### Linear Mixed Effect Model

Many common statistical models can be expressed as linear models that incorporate both fixed effects (i.e., parameters associated with certain repeatable levels of experimental factors), and random effects (i.e., parameters associated with individual experimental units drawn randomly from a population). A model with both, fixed and random effects is called a mixed-effects model (Pinheiro, J.C. and Bates, D.M., 2002). Mixed-effects models can describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors (e.g. longitudinal data, repeated measures, multilevel and block designs). By associating common random effects to observations sharing the same level of a classification factor, mixed-effects models flexibly represent the covariance structure induced by the grouping of the data (Pinheiro, J.C. and Bates, D.M., 2002). In our case, we hypothesized, that the effect of the surface community composition on the deep-water masses was dependent on the water mass (W) and the province (P), while each station (S) was treated as introducing random effects; denoted as 1|S in the function lmer of the R-package lme4 (Bates, D. *et al.*, 2013). Specifically, we extracted the weighted UniFrac dissimilarity ( $D_{wu}$ ) between the subsurface layer (SSL at ~100 m depth) and each sample below at the same station ( $n = 5$ ).

Firstly, we built a linear mixed effect model (lm) where province (P) was the fixed and the station (S) was the random effect on  $D_{wu}$ , obtaining our first hypothetical model  $A_H$ :

$$A_H = \text{lm}(D_{wu} \sim P + (1|S)) \quad (1)$$

A simple null model ( $A_0$ ) was introduced that omits the effect of the provinces:

$$A_0 = \text{lm}(D_{wu} \sim (1|S)) \quad (2)$$

Subsequently the full model and the null model were tested against each other. The test was significant ( $\text{Chi}^2_2 = 8.02, p < 0.05$ ) for our data set.

Secondly, a model ( $B_H$ ) was introduced, which assumed an interaction between water mass (W) and province (P) influencing the effect of the surface community on the deep-water community:

$$B_H = \text{lm}(D_{wu} \sim W * P + (1|S)) \quad (3)$$

$B_H$  was tested against  $A_H$ , and as expected, a highly significant effect of the water mass was found ( $\text{Chi}^2_{12} = 65.9, p < 0.001$ ).

To examine the effect of province on the dissimilarity between the deep waters and the SSL we formulated:

$$C_H = \text{lm}(D_{wu} \sim W + (1|S)) \quad (4)$$

Finally we tested  $C_H$  against  $B_H$ . The result was highly significant ( $\text{Chi}^2_{10} = 27.0, p < 0.01$ ) indicating a significant effect of province to explain the dissimilarity of the water mass to the SSL.

### **Fixation, preparation, and analysis of CARD-FISH samples**

CARD-FISH was used to determine the abundance of major groups of prokaryotes at nine stations (Stns 5, 6, 7, 9, 10, 11, 13, 14, and 15) along the transect, essentially following the method described by Pernthaler, A. *et al.* (2002). Immediately after collecting the samples from the Niskin bottles, 20–80 mL of water were fixed by adding formaldehyde (37%; 2% final concentration) and stored at 4°C in the dark. After 18 h, the samples were filtered through 0.2- $\mu\text{m}$  polycarbonate filters (Millipore GTTP, 25 mm filter diameter) supported by nitrate cellulose filters (Millipore, HAWP, 0.45  $\mu\text{m}$ ), washed twice with 10 mL Milli-Q water, dried and stored in a microfuge vial at -20°C until further processing in the laboratory. For hybridization we used horseradish peroxidase-labeled oligonucleotide probes to determine the fraction of 4',6-diamidino-2-phenylindole (DAPI)-stained cells. The hybridization conditions and probes targeting Bacteria, SAR11 (Alphaproteobacteria) and SAR202 (Chloroflexi), are given in Supplementary Table S4. Cells were counter-stained with a DAPI-mix [5.5 parts of Citifluor (Citifluor), 1 part of Vectashield (Vector Laboratories) and 0.5 parts of phosphate-buffered saline (PBS) with DAPI (final concentration 1  $\mu\text{g mL}^{-1}$ )]. Enumeration of DAPI-stained cells and cells stained with the specific probes was performed with a Zeiss Axioplan 2 epifluorescence microscope equipped with a 100-W Hg lamp and appropriate filter sets for DAPI, and Alexa448. A minimum of 600 DAPI-stained cells was counted per sample. An AxioCam MRm black and white camera (Carl Zeiss, Germany) was used to capture pictures using a plan apochromatic 100x/1.46 oil objective. To obtain relative abundances comparable to the 454-data, the number of SAR11 and SAR202 probe-positive cells was divided by the EUB-positive cells of the same sample.

### **R packages used for the analysis of datasets**

The statistical tests used to analyze the data were performed in R with the help of the following freely available packages: Ade4 (Dray and Dufour, 2007), Agricolae (de Mendiburu, 2013), FactoMineR (Husson *et al.*, 2013), MASS (Venables and Ripley, 2002), sna (Butts, 2013), Vegan (Oksanen *et al.*, 2013) and Venneuler (Wilkinson, 2011).

### **Evaluation of the 454 sequencing approach**

Commonly, three-percent sequence dissimilarity in the 16S rRNA gene is used as a threshold to delineate operational taxonomic units (OTUs) (Martin, 2002) although this might not be appropriate for all prokaryotes and likely masks a considerable amount of micro-diversity (Garcia-Martinez and Rodriguez-Valera, 2000; Acinas *et al.*, 2004). Sequencing the V1–V3 region has been shown to yield similar phylogenetic affiliations compared to Sanger sequencing (Kumar *et al.*, 2011) and the region was found better for phylogenetic community characterization when compared to short sequences of the hypervariable V6 region used in the past (Liu *et al.*, 2007; Jeraldo *et al.*, 2011). However, to check the consistency of 454 pyrosequencing-derived relative OTU abundances, we performed a comparison with quantitative CARD-FISH analysis using a set of SAR11 and SAR202 probes (see Supplementary Table S5 and Supplementary Figure S6). This comparison yielded a good agreement between both methods similarly to that recently reported by Logares *et al.* (2013). Based on these methodological issues, however, the explanatory power of commonly used diversity indices remains uncertain (Tuomisto, 2010).

## Tables

**Table S1** Diversity indices from different water masses and biogeographical provinces

WM	Province	Bac Seq <sup>a</sup>	Bac OTU <sup>b</sup>	Chao1 <sup>c</sup>	PD <sup>d</sup>
<b>SSL</b>		<b>814 ± 284</b>	<b>130 ± 36</b>	<b>63 ± 17</b>	<b>8.99 ± 1.36</b>
	NADR	665 ± 475	116 ± 55	60 ± 21	9.74 ± 1.33
	NAST	928 ± 152	154 ± 27	71 ± 22	9.55 ± 1.24
	NATR	849 ± 135	120 ± 12	57 ± 9	7.69 ± 0.49
<b>O<sub>2</sub>-min</b>		<b>804 ± 229</b>	<b>180 ± 23</b>	<b>115 ± 19</b>	<b>10.95 ± 0.45</b>
	NADR	696 ± 175	157 ± 17	98 ± 9	10.95 ± 0.81
	NAST	1050 ± 296	198 ± 28	103 ± 10	10.64 ± 0.02
	NATR	712 ± 114	184 ± 14	133 ± 10	11.15 ± 0.36
<b>MSOW</b>		<b>821 ± 260</b>	<b>189 ± 42</b>	<b>122 ± 23</b>	<b>11.45 ± 0.62</b>
	NADR	713 ± 320	155 ± 29	101 ± 10	11.34 ± 0.7
	NAST	1014 ± 295	221 ± 52	127 ± 15	11.54 ± 1.06
	NATR	792 ± 43	207 ± 1	149 ± 3	11.52 ± 0.34
<b>AAIW</b>		<b>741 ± 112</b>	<b>169 ± 20</b>	<b>106 ± 21</b>	<b>11.62 ± 0.97</b>
	NADR	801 ± 182	162 ± 24	91 ± 4	10.94 ± 0.69
	NAST	707 ± 81	159 ± 18	95 ± 13	11.4 ± 1.14
	NATR	713 ± 51	187 ± 9	133 ± 1	12.52 ± 0.14
<b>NEADW</b>		<b>664 ± 136</b>	<b>159 ± 18</b>	<b>106 ± 17</b>	<b>12.22 ± 0.99</b>
	NADR	694 ± 103	153 ± 8	91 ± 4	11.59 ± 0.41
	NAST	756 ± 111	169 ± 11	108 ± 3	12.77 ± 0.86
	NATR	541 ± 117	154 ± 30	118 ± 25	12.29 ± 1.39
<b>LDW</b>		<b>551 ± 107</b>	<b>127 ± 24</b>	<b>82 ± 18</b>	<b>11.94 ± 1.09</b>
	NADR	526 ± 74	115 ± 5	71 ± 8	11.48 ± 1.08
	NAST	526 ± 154	121 ± 27	78 ± 14	11.9 ± 1.53
	NATR	599 ± 106	145 ± 29	97 ± 23	12.43 ± 0.76

For abbreviation of water masses and biographical provinces see Supplementary Table S1.

<sup>a</sup>Bac Seq: bacterial sequences after trimming and quality filtering

<sup>b</sup>Bac OTU: Operational taxonomic units at 97% similarity thresholds

<sup>c</sup>Chao 1: Richness estimator Chao 1 rarefied to base coverage (lowest covered sample = 80%)

<sup>d</sup>PD: Phylogenetic diversity (Faith, 1992) at lowest sample size (rarefied)

Numbers indicate averages ± standard deviations (n = 3 per province)

**Table S2** Physico-chemical parameters in the water masses (WM) and biogeographical provinces of the transect in the North East Atlantic

WM <sup>a</sup>	Province <sup>b</sup>	Depth (m)	Theta (°C)	Salinity	Si (μmol/kg)	O <sub>2</sub> (μmol/kg)	PO <sub>4</sub> (μmol/kg)	NO <sub>3</sub> (μmol/kg)	NO <sub>2</sub> (μmol/kg)	TOC (μmol/kg)
SSL		<b>98 ± 5</b>	<b>16.86 ± 3.37</b>	<b>36.34 ± 0.53</b>	<b>1.4 ± 0.9</b>	<b>226 ± 5</b>	<b>0.21 ± 0.23</b>	<b>3.4 ± 3.9</b>	<b>0.048±0.045</b>	<b>64 ± 10</b>
	NADR	100 ± 1	13.12 ± 0.79	35.76 ± 0.14	2.5 ± 0.5	231 ± 3	0.49 ± 0.12	8.1 ± 1.8	0.045±0.009	67 ± 3
	NAST	96 ± 8	17.10 ± 1.95	36.36 ± 0.26	1.1 ± 0.4	222 ± 3	0.12 ± 0.12	2.0 ± 2.0	0.086±0.064	64 ± 18
O <sub>2</sub> -min	NATR	100 ± 1	20.36 ± 1.25	36.90 ± 0.27	0.6 ± 0.1	225 ± 5	0.01 ± 0.00	0.0 ± 0.0	0.014±0.005	62 ± 4
		<b>783 ± 138</b>	<b>9.67 ± 0.83</b>	<b>35.47 ± 0.21</b>	<b>10.2 ± 1.9</b>	<b>175 ± 9</b>	<b>1.20 ± 0.15</b>	<b>19.4 ± 2.2</b>	<b>0.022±0.004</b>	<b>53 ± 7</b>
	NADR	792 ± 256	9.46 ± 0.98	35.49 ± 0.35	9.6 ± 0.6	183 ± 4	1.14 ± 0.10	18.3 ± 1.2	0.022±0.005	55 ± 2
MSOW	NAST	732 ± 50	10.38 ± 0.15	35.54 ± 0.11	8.9 ± 0.2	177 ± 3	1.11 ± 0.03	18.3 ± 0.6	0.023±0.003	56 ± 6
	NATR	826 ± 24	9.18 ± 0.78	35.37 ± 0.12	12.1 ± 2.4	166 ± 9	1.34 ± 0.17	21.7 ± 2.5	0.019±0.005	48 ± 10
		<b>1024 ± 111</b>	<b>8.32 ± 1.42</b>	<b>35.53 ± 0.23</b>	<b>11.3 ± 1.6</b>	<b>194 ± 16</b>	<b>1.17 ± 0.10</b>	<b>18.8 ± 1.4</b>	<b>0.022±0.004</b>	<b>56 ± 10</b>
AAIW	NADR	972 ± 124	8.39 ± 2.49	35.53 ± 0.37	9.9 ± 0.9	202 ± 29	1.11 ± 0.08	17.8 ± 0.7	0.023±0.004	61 ± 11
	NAST	1066 ± 59	8.60 ± 0.62	35.64 ± 0.16	11.4 ± 0.5	192 ± 3	1.14 ± 0.07	18.5 ± 1.0	0.019±0.001	53 ± 8
	NATR	1033 ± 153	7.98 ± 1.09	35.42 ± 0.10	12.5 ± 1.9	189 ± 6	1.26 ± 0.09	20.2 ± 1.2	0.023±0.006	55 ± 14
NEADW		<b>1530 ± 156</b>	<b>4.98 ± 0.71</b>	<b>35.14 ± 0.11</b>	<b>13.1 ± 2.7</b>	<b>239 ± 12</b>	<b>1.19 ± 0.07</b>	<b>18.6 ± 1.0</b>	<b>0.019±0.003</b>	<b>51 ± 11</b>
	NADR	1364 ± 115	5.08 ± 1.08	35.10 ± 0.16	10.8 ± 0.1	245 ± 18	1.15 ± 0.01	17.9 ± 0.2	0.021±0.003	58 ± 9
	NAST	1533 ± 58	5.16 ± 0.79	35.19 ± 0.12	12 ± 0.4	240 ± 11	1.15 ± 0.01	18.2 ± 0.1	0.019±0.002	52 ± 7
LDW	NATR	1693 ± 12	4.69 ± 0.17	35.14 ± 0.03	16.4 ± 1.6	232 ± 7	1.27 ± 0.07	19.7 ± 0.9	0.016±0.003	41 ± 10
		<b>2749 ± 2</b>	<b>2.75 ± 0.10</b>	<b>34.95 ± 0.01</b>	<b>25.5 ± 5.4</b>	<b>254 ± 8</b>	<b>1.29 ± 0.09</b>	<b>19.5 ± 1.2</b>	<b>0.021±0.003</b>	<b>49 ± 10</b>
	NADR	2749 ± 3	2.82 ± 0.10	34.94 ± 0.01	19.5 ± 2.7	263 ± 4	1.20 ± 0.03	18.1 ± 0.5	0.022±0.002	56 ± 5
LDW	NAST	2750 ± 1	2.71 ± 0.12	34.95 ± 0.01	26.4 ± 3.6	253 ± 3	1.29 ± 0.05	19.6 ± 0.6	0.021±0.003	51 ± 10
	NATR	2747 ± 1	2.72 ± 0.03	34.96 ± 0.00	30.5 ± 2	245 ± 4	1.38 ± 0.05	20.7 ± 0.6	0.018±0.003	41 ± 9
		<b>4304 ± 334</b>	<b>2.13 ± 0.11</b>	<b>34.91 ± 0.01</b>	<b>43.3 ± 2.0</b>	<b>244 ± 2</b>	<b>1.49 ± 0.01</b>	<b>22.3 ± 0.3</b>	<b>0.019±0.004</b>	<b>47 ± 6</b>
LDW	NADR	4039 ± 58	2.23 ± 0.02	34.92 ± 0.00	41.7 ± 0.8	241 ± 1	1.48 ± 0.01	22.1 ± 0.2	0.022±0.003	52 ± 3
	NAST	4170 ± 60	2.15 ± 0.03	34.91 ± 0.00	42.8 ± 0.7	244 ± 1	1.48 ± 0.01	22.3 ± 0.1	0.018±0.003	49 ± 5
	NATR	4701 ± 265	2.00 ± 0.08	34.90 ± 0.01	45.5 ± 1.6	247 ± 1	1.50 ± 0.01	22.6 ± 0.2	0.017±0.004	43 ± 7

<sup>a</sup>Water masses: SSL – Subsurface Layer (100m), O<sub>2</sub>-min – O<sub>2</sub> Minimum, MSOW – Mediterranean Sea Outflow Water, AAIW – Antarctic Intermediate Water, NEADW – North East Atlantic Deep Water, LDW – Lower Deep Water.

<sup>b</sup>Provinces: NADR – North Atlantic Drift Province, NAST – North Atlantic Subtropical Province, NATR – North Atlantic Tropical Gyral Province  
Numbers indicate averages ± standard deviations (n = 3 per province)

**Table S3** ANOSIM (global R) and Mantel (weighted Spearman rho) statistics (all  $p < 0.001$ ) to test the influence of the abundance fractions on the differences in community composition between provinces and water masses in the meso- and bathypelagic waters.

ANOSIM	Rare	Common	Abundant	All
Province <sup>a</sup>	0.512	0.518	0.442	0.558
Water mass <sup>b</sup>	0.647	0.763	0.722	0.803
Ratio (a/b)	0.79	0.69	0.61	0.69

Mantel	Rare	Common	Abundant	All
Rare	1	0.832	0.7	0.902
Common		1	0.787	0.926
Abundant			1	0.825
All				1

<sup>a</sup>Fractions rare (< 1% of the community), abundant (1–10 % of the community) and common (> 10% of the community) as defined in the main text.

**Table S4** Biological parameters according to water mass (WM) and biogeographical provinces

WM	Province	AOU ( $\mu\text{mol L}^{-1}$ )	FLC (ng Chla $\text{L}^{-1}$ )	VA ( $10^6 \text{ ml}^{-1}$ )	PA ( $10^5 \text{ ml}^{-1}$ )	Leu inc ( $\text{pmol L}^{-1} \text{ h}^{-1}$ )	PHP ( $\mu\text{mol C m}^{-3} \text{ d}^{-1}$ )
SSL		<b>17 ± 13</b>	<b>99 ± 48</b>	<b>2.69 ± 0.73</b>	<b>3.34 ± 0.63</b>	<b>3.2654 ± 3.3593</b>	<b>10.1229 ± 10.4138</b>
	NADR	28.0 ± 2.6	28.4 ± 6.7	2.297 ± 0.899	3.89 ± 0.91	1.1283 ± 0.2439	3.4977 ± 0.7562
	NAST	19.3 ± 10.3	127.6 ± 18.5	3.004 ± 0.689	2.94 ± 0.44	1.8632 ± 1.3155	5.7759 ± 4.0779
<b>O<sub>2</sub>-min</b>	NATR	-0.2 ± 5.6	113 ± 33.2	2.434 ± 0.815	3.59 ± 0.20	8.2071 ± 2.8113	25.442 ± 8.7149
		<b>104 ± 9</b>	<b>15 ± 2</b>	<b>0.31 ± 0.17</b>	<b>0.80 ± 1.12</b>	<b>0.0486 ± 0.0349</b>	<b>0.1505 ± 0.1083</b>
	NADR	96.9 ± n.a.	17.2 ± n.a.	0.560 ± n.a.	2.76 ± n.a.	0.1078 ± n.a.	0.3341 ± n.a.
MSOW	NAST	99 ± 4.1	15.6 ± 0.9	0.285 ± 0.174	0.48 ± 0.26	0.0433 ± 0.0048	0.1342 ± 0.0148
	NATR	112.6 ± 9.6	12.2 ± 1.3	0.209 ± 0.036	0.13 ± 0.07	0.0242 ± 0.0107	0.075 ± 0.0330
		<b>96 ± 11</b>	<b>14 ± 2</b>	<b>0.33 ± 0.28</b>	<b>0.64 ± 0.56</b>	<b>0.0390 ± 0.0288</b>	<b>0.1210 ± 0.0892</b>
AAIW	NADR	88.0 ± 12.1	14.1 ± 2.8	0.618 ± 0.132	1.04 ± 0.70	0.0657 ± 0.0214	0.2035 ± 0.0663
	NAST	96.6 ± 5.5	13.2 ± 1.3	0.127 ± 0.026	0.32 ± 0.02	0.0242 ± 0.0116	0.0749 ± 0.0360
	NATR	106.4 ± 0	13.1 ± 0.1	0.097 ± 0.004	0.35 ± 0.09	0.014 ± 0.0098	0.0434 ± 0.0305
NEADW		<b>77 ± 10</b>	<b>12 ± 2</b>	<b>0.39 ± 0.39</b>	<b>0.43 ± 0.38</b>	<b>0.0171 ± 0.0135</b>	<b>0.0531 ± 0.0419</b>
	NADR	69.0 ± 10.2	13.1 ± 3.1	0.872 ± 0.228	0.86 ± 0.39	0.0341 ± 0.0061	0.1058 ± 0.0189
	NAST	74.4 ± 5.1	11.9 ± 1.3	0.215 ± 0.096	0.22 ± 0.07	0.0117 ± 0.0021	0.0363 ± 0.0064
LDW	NATR	86.5 ± 7	9.8 ± 1.0	0.092 ± 0.058	0.22 ± 0.06	0.0055 ± 0.0033	0.0171 ± 0.0102
		<b>83 ± 9</b>	<b>3 ± 2</b>	<b>0.33 ± 0.26</b>	<b>0.26 ± 0.25</b>	<b>0.0062 ± 0.0037</b>	<b>0.0193 ± 0.0116</b>
	NADR	73.0 ± 4.3	2.6 ± 1.5	0.641 ± 0.129	0.51 ± 0.31	0.0104 ± 0.0026	0.0322 ± 0.0081
LDW	NAST	84.4 ± 3.9	3.6 ± 3.0	0.255 ± 0.021	0.15 ± 0.02	0.0058 ± 0.0010	0.0181 ± 0.0030
	NATR	92.8 ± 4.6	4.1 ± 2.5	0.088 ± 0.035	0.11 ± 0.02	0.0024 ± 0.0004	0.0075 ± 0.0012
		<b>103 ± 1</b>	<b>4 ± 2</b>	<b>0.29 ± 0.26</b>	<b>0.16 ± 0.12</b>	<b>0.0066 ± 0.0075</b>	<b>0.0206 ± 0.0234</b>
LDW	NADR	102.9 ± 0.9	3.1 ± 2.5	0.596 ± 0.203	0.28 ± 0.16	0.0148 ± 0.0086	0.0457 ± 0.0267
	NAST	102.0 ± 0.8	3.2 ± 0.8	0.191 ± 0.082	0.11 ± 0.02	0.0037 ± 0.0012	0.0115 ± 0.0037
	NATR	102.9 ± 0.5	5.0 ± 3.7	0.093 ± 0.023	0.09 ± 0.05	0.0015 ± 0.0001	0.0045 ± 0.0004

For abbreviation of water masses and biographical provinces see Supplementary Table S1.  
 Numbers indicate mean ± standard deviations (n > 5 per province)



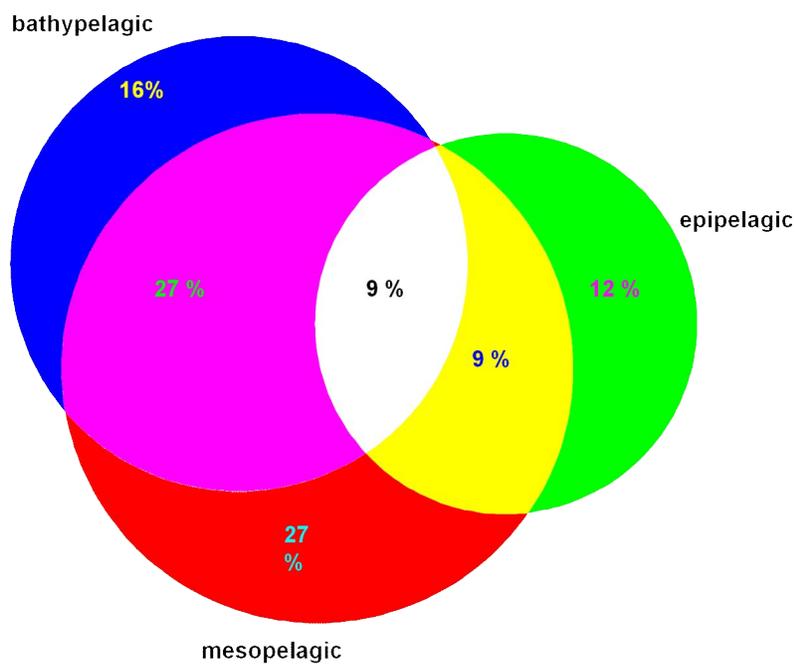
**Table S5** Probe sequences and formamide concentrations used for CARD-FISH

Probe	Sequence (5'→3')	FA <sup>a</sup> (%)	Reference
EUB338 I	GCTGCCTCCCGTAGGAGT	55	(Amann <i>et al.</i> , 1990)
EUB338 II	GCAGCCACCCGTAGGTGT	55	(Daims <i>et al.</i> , 1999)
EUB338 III	GCTGCCACCCGTAGGTGT	55	(Daims <i>et al.</i> , 1999)
SAR11-152R	ATTAGCACAAGTTTCCYCGTGT	45	(Morris <i>et al.</i> , 2002)
SAR11-441R	TACAGTCATTTTCTTCCCCGAC	45	(Morris <i>et al.</i> , 2002)
SAR11-542R	TCCGAACTACGCTAGGTC	45	(Morris <i>et al.</i> , 2002)
SAR11-732R	GTCAGTAATGATCCAGAAAGYTG	45	(Morris <i>et al.</i> , 2002)
SAR202-104R	GTTACTCAGCCGTCTGCC	35	(Morris <i>et al.</i> , 2004)
SAR202-312R	TGTCTCAGTCCCCCTCTG	40*	(Morris <i>et al.</i> , 2004)

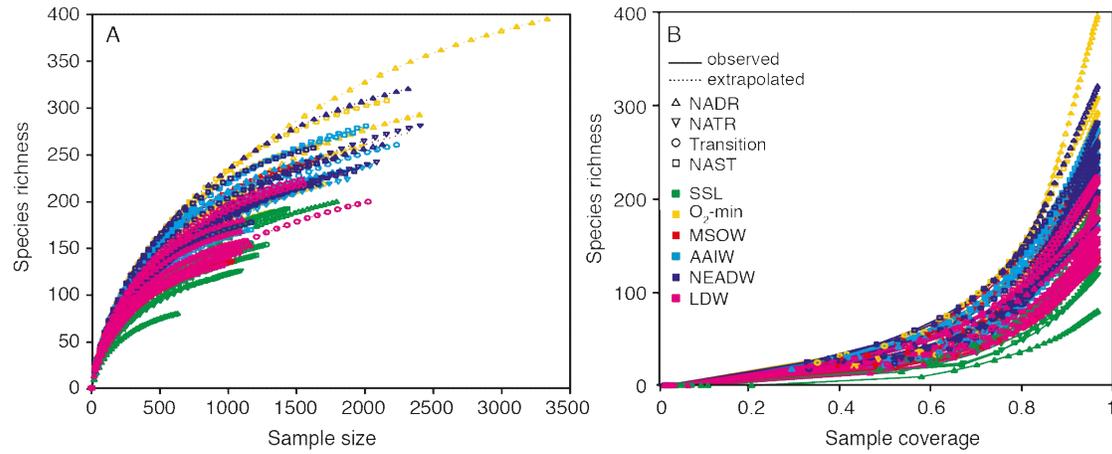
<sup>a</sup>Formamide (FA) concentration in percent of hybridization buffer

\*Applied formamide concentration according to Schattenhofer *et al.* (2009)

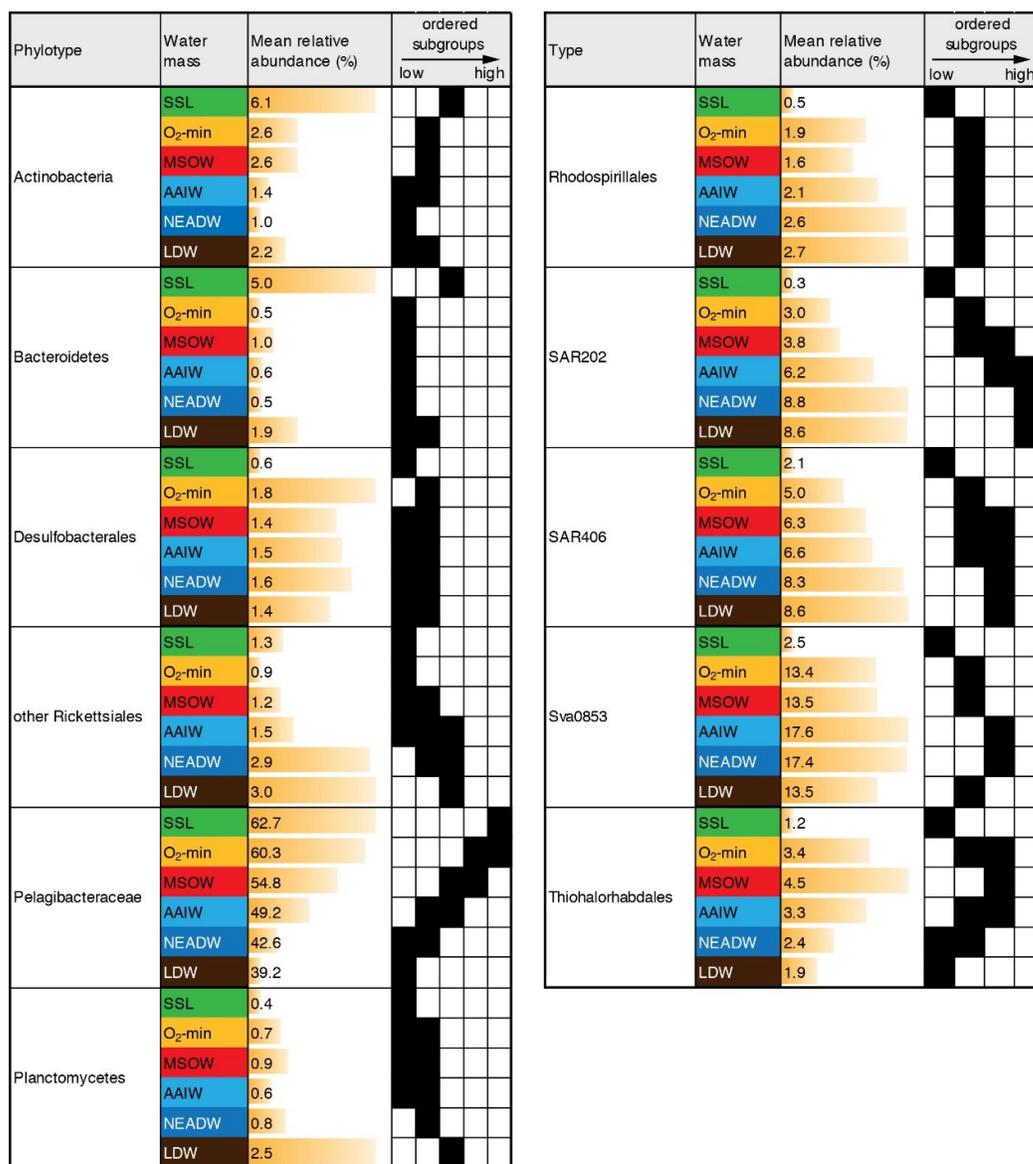
## Figures



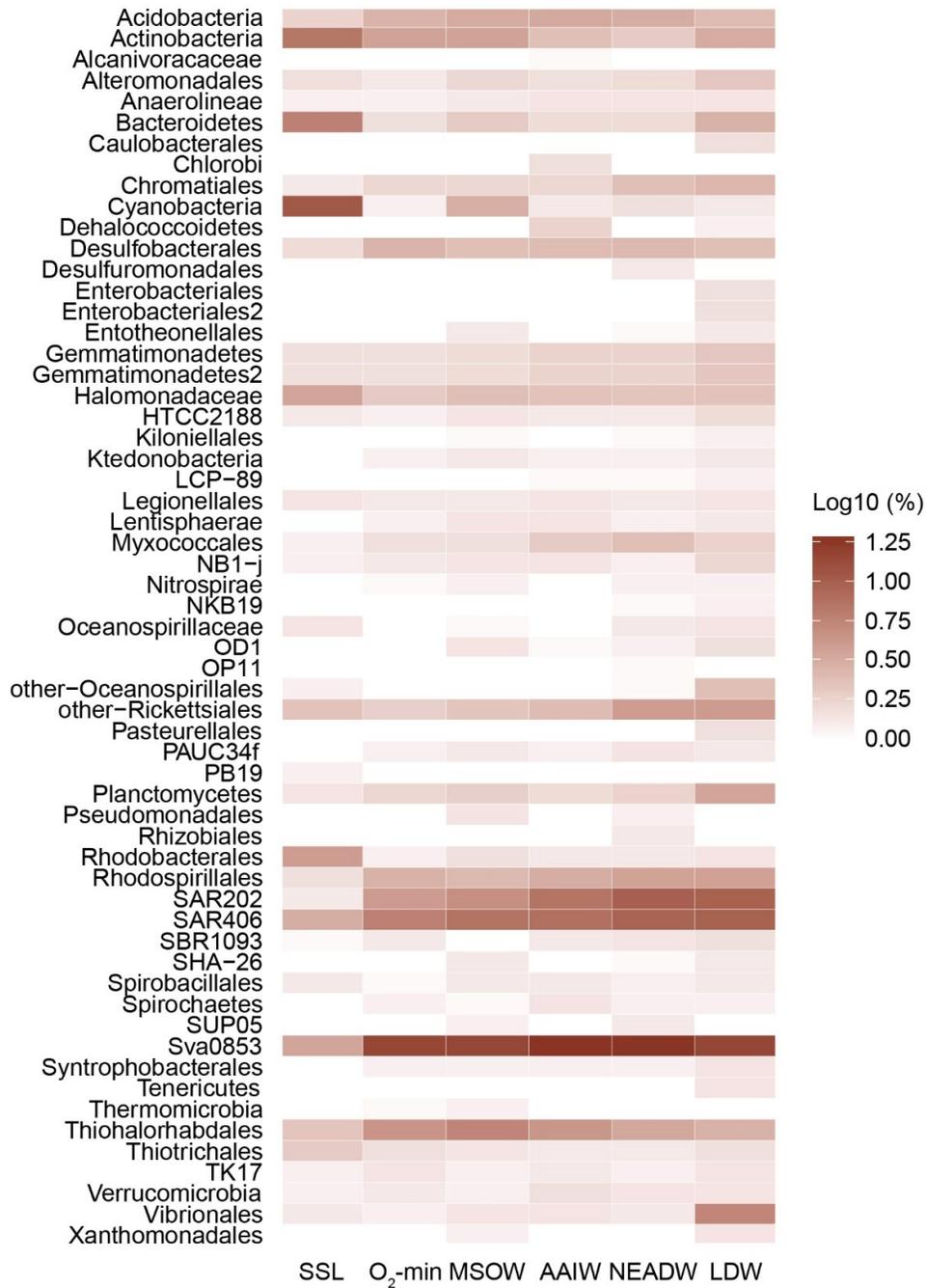
**Figure S1** Percentages of shared and unique OTUs between different pelagic realms represented as Venn diagram. The size of each area is corresponding to the percentage. Epipelagic (green) = SSL; mesopelagic (red) = O<sub>2</sub>-min, MSOW, and AAIW; bathypelagic (blue) = NEADW and LDW.



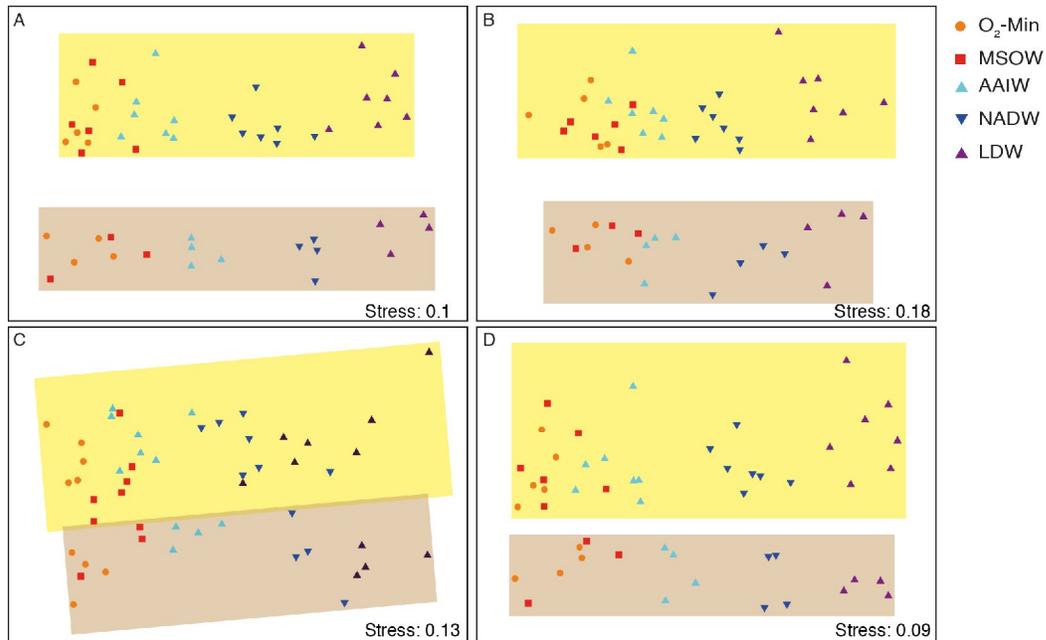
**Figure S2** Rarefaction and extrapolation curves of the Chao1 richness estimator versus sample size (A) and sample coverage (B). The sample coverage was derived from the slope of the species accumulation curves as suggested in Chao *et al.* (2012) using the program iNEXT with 200 bootstraps. The calculated minimal observed coverage of 0.8 was used for downstream statistical analysis and comparison of the Chao1 richness estimators between samples, provinces and water masses.



**Figure S3** Water mass-dependent distribution of the dominant bacterial phylotypes in the NADR, NAST and NATR. Black squares of non-overlapping ordered subgroups represent statistically significant differences at  $p < 0.05$ . The squares also indicate an increase or decrease in relative abundance between the water masses.



**Figure S4** Heatmap of all OTUs including those that were not significantly different between water masses. Original relative abundances in percent are rescaled to log10 for better visibility.



**Figure S5** Non-metric multidimensional scaling (NMDS) plots of the meso- and bathypelagic bacterial community using different abundance categories calculated as Bray-Curtis similarity on square-root transformed relative abundances of OTUs. Full range of OTUs (A), rare OTUs with abundances <1% (B), abundant OTUs occurring >10% (C), and common OTUs in the range of 1 – 10% (D). Yellow background indicates the North Atlantic Drift Province (NADR) and North Atlantic Subtropical Province (NAST), brown background color indicates the North Atlantic Tropical Gyral Province (NAST).

## References

- Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl, DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919-1925.
- Bates D, Maechler M, Bolker, B. (2013) lme4: Linear mixed-effects models using Eigen and Eigenfunctions. R package version 0.999999-2.
- Butts CT (2013) sna: Tools for social network analysis. R package version 2.3-1.
- Daims H, Brühl A, Amann R, Schleifer K-H, Wagner, M. (1999) The Domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434-444.
- de Mendiburu F. (2013) agricolae: Statistical Procedures for Agricultural Research. R package version 1.1-4.
- Dray S, Dufour AB. (2007) The ade4 package: implementing the duality diagram for ecologists. **22**: 1-20.
- Husson F, Josse J, Le S, Mazet J. (2013) FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.25.
- Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. (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.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2013) vegan: Community Ecology Package. <http://CRAN.R-project.org/package=vegan>
- Pernthaler A, Pernthaler J, Amann R. (2002) Fluorescence In Situ Hybridization and Catalyzed Reporter Deposition for the Identification of Marine Bacteria. *Appl Environ Microbiol* **68**: 3094-3101.
- Pinheiro JC, Bates DM. (2002) *Mixed-effects models in S and S-PLUS*, corr. 3. print. (edn.) New York, NY: Springer.
- Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078-2093.
- Tuomisto H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* **33**: 2-22.
- Venables WN, Ripley BD. (2002) *Modern Applied Statistics with S*, Fourth Edition (edn.) New York: Springer.
- Wilkinson L. (2011) venneuler: Venn and Euler Diagrams. R package version 1.1-0.