environmental microbiology

Environmental Microbiology (2015) **17**(10), 4133–4142 doi:10.1111/1462-2920.12984

Seasonality in molecular and cytometric diversity of marine bacterioplankton: the re-shuffling of bacterial taxa by vertical mixing

Francisca C. García,[1]* Laura Alonso-Sáez,[1†] Xosé Anxelu G. Morán[1,2] and Ángel López-Urrutia[1]**

[1]*Instituto Español de Oceanografía, Centro Oceanográfico de Gijón/Xixón, 33212 Gijón/Xixón, Spain.*

[2]*Red Sea Research Center, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, 23955-6500 Thuwal, Saudi Arabia.*

Summary

The 'cytometric diversity' of phytoplankton communities has been studied based on single-cell properties, but the applicability of this method to characterize bacterioplankton has been unexplored. Here, we analysed seasonal changes in cytometric diversity of marine bacterioplankton along a decadal time-series at three coastal stations in the Southern Bay of Biscay. Shannon–Weaver diversity estimates and Bray–Curtis similarities obtained by cytometric and molecular (16S rRNA tag sequencing) methods were significantly correlated in samples from a 3.5 year monthly time-series. Both methods showed a consistent cyclical pattern in the diversity of surface bacterial communities with maximal values in winter. The analysis of the highly resolved flow cytometry time-series across the vertical profile showed that water column mixing was a key factor explaining the seasonal changes in bacterial composition and the winter increase in bacterial diversity in coastal surface waters. Due to its low cost and short processing time as compared with genetic methods, the cytometric diversity approach represents a useful complementary tool in the macroecology of aquatic microbes.

Received 21 October, 2014; revised 10 July, 2015; accepted 10 July, 2015. For correspondence. *E-mail paquigrcgrc@gmail.com; Tel. (+34) 985308672; Fax (+34) 985326277. **E-mail alop@gi.ieo.es; Tel. (+34) 985308672; Fax (+34) 985326277. †Present address: AZTI-Tecnalia, Marine Research Division, Txatxarramendi Irla, s/n. 48395 Sukarrieta, Spain.

Introduction

Marine bacteria are the most diverse component of planktonic communities (Giovannoni and Rappé, 2000; Venter *et al*., 2004) and constitute one of the largest biomass stocks in the ocean (Buitenhuis *et al*., 2012). Identifying microbial taxa present in the environment and understanding the changes in the composition of microbial communities over spatio-temporal gradients has been a challenge in microbial ecology studies (Pielou, 1975; Giovannoni and Stingl, 2005; Green *et al*., 2008; Zinger *et al*., 2011).

During the last two decades, the use of molecular tools has revolutionized the way ecologists study the distribution and diversity of marine bacteria taxa (Zinger *et al*., 2012). Different seasonal, spatial and latitudinal patterns in bacterioplankton phylogenetic composition have been reported using fingerprinting methods such as denaturing gradient gel electrophoresis (DGGE), automated ribosomal intergenic spacer analysis or terminal fragment length polymorphysm (Schauer *et al*., 2000; Morris *et al*., 2005; Fuhrman *et al*., 2008; Chow *et al*., 2013). Although these methods enable a fast comparison of samples, other approaches such as fluorescence in situ hybridization (FISH), have permitted the determination of the abundance of specific target populations or quantitative changes in the phylogenetic composition of bacterial communities with a reasonable accuracy (Cottrell and Kirchman, 2000; Pernthaler *et al*., 2002; Amann and Fuchs, 2008; Schattenhofer *et al*., 2009). Recently, the development of high-throughput sequencing technologies has represented a breakthrough in this field by increasing several orders of magnitude the number of sequences typically obtained per sample (Sogin *et al*., 2006; Zinger *et al*., 2011; Vergin *et al*., 2013; Salter *et al*., 2015).

While all molecular approaches have their own advantages and disadvantages (e.g. Alonso-Sáez *et al*., 2007), a common limitation is that they are time-consuming and/or expensive for large-scale studies requiring the analysis of hundreds to thousands of samples. In early ecological studies, Margalef (1968) proposed other classification criteria, the photosynthetic pigment composition, to study the structure and composition of phytoplankton

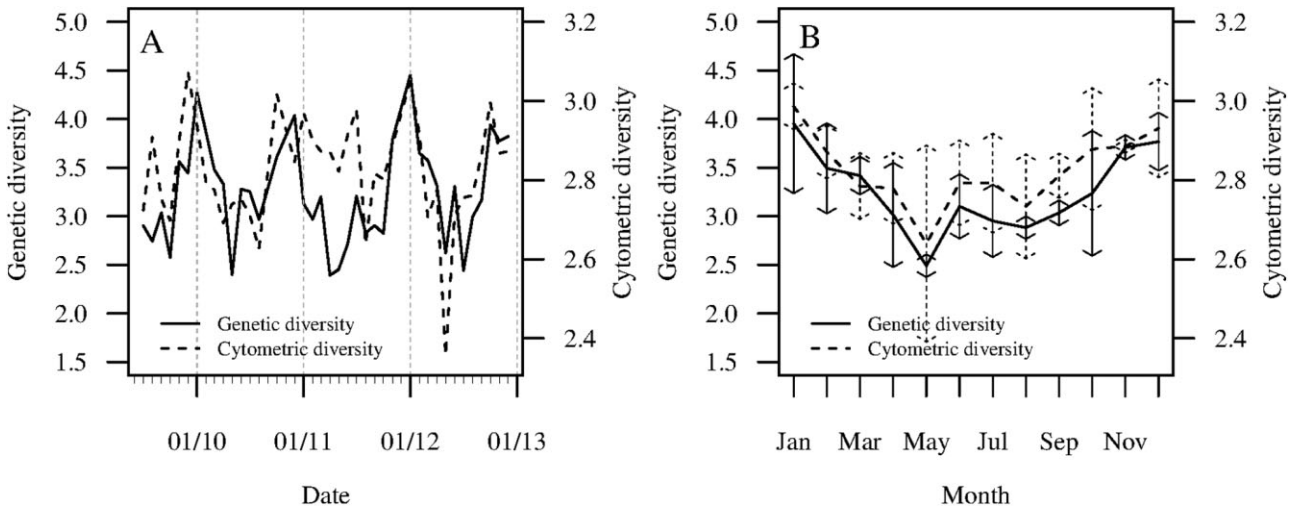

Fig. 1. Trends of genetic (solid line) and cytometric (dashed line) diversity (Shannon–Weaver index) of surface bacterial communities at E2 during a 3.5 year monthly time-series (A) and as a mean annual cycle (B).

communities. Later, Li (1997) introduced the concept of 'cytometric diversity' to characterize marine phytoplankton communities using individual cell properties measured by flow cytometry. In natural and stained samples, cytometric properties such as light scatter and fluorescence reflect changes in cell size, complexity, pigment and nucleic acid content, and therefore, they can be used to classify planktonic cells. In an initial application of this method in freshwater bacterial communities, its power in differentiating bacterial groups was limited to a few (i.e. < 10) groups, presumably made up of a variable number of species (Schiaffino *et al*., 2013). However, the applicability of cytometric diversity based on cytometric bins potentially allows the differentiation of a substantially larger number (> 100) of bacterial groups. Thus, this methodology would represent an important advancement in detecting changes in community composition, as flow cytometry is technically simple, requires small volumes and allows a relatively fast analysis of hundreds to thousands of samples.

Here, we analysed the seasonal dynamics of heterotrophic bacterioplankton in a decadal time-series in the Southern Bay of Biscay using the cytometric diversity approach and compared it with genetic diversity estimated through molecular methods. Although seasonal changes in the composition of coastal marine bacterioplankton have been widely reported (Alonso-Sáez *et al*., 2007; 2015; Campbell *et al*., 2009; Gilbert *et al*., 2012; Chow *et al*., 2013), the factors controlling this seasonality are less understood. Using the power of the cytometric method, we performed a large-scale analysis at three continental shelf stations in the Cantabrian Sea (2586 samples in total), at a spatio-temporal resolution not easily achievable with molecular techniques. Our objectives were: (i) to test whether ecological meaningful patterns can be derived from cytometric diversity analysis in marine bacterioplankton and (ii) to understand the factors responsible for seasonal changes in the diversity and composition of coastal bacterial communities.

Results

Comparison of genetic and cytometric diversities

In the molecular dataset from station E2 (obtained by 16S rRNA tag pyrosequencing over 3.5 years), a cyclical pattern of diversity with maximal Shannon–Weaver index values in winter and minima in early summer was found (using an operational taxonomic unit [OTU] genetic distance cut-off of 0.03; Fig. 1). This seasonality is clear both in the monthly time-series (Fig. 1A) and in the average annual cycle (Fig. 1B). As expected, the Shannon–Weaver diversity indices were higher when using a more stringent distance cut-off for OTU construction (0.01), but the same monthly and average annual patterns of diversity were observed (Figs S1 and S2).

Although the number of taxa obtained by the molecular approach is ultimately defined by the genetic distance cut-off, cytometric diversity relies on the number of bins defined (e.g., 1024 categories in this study). Therefore, the absolute values of diversity indices obtained by the molecular and the cytometric approaches cannot be meaningfully compared. Rather, our focus was on comparing spatio-temporal trends. The cytometric approach provided a similar, albeit weaker, seasonal pattern in the diversity of bacterial communities along the 3.5 years (Fig. 1A). The largest differences between genetic and

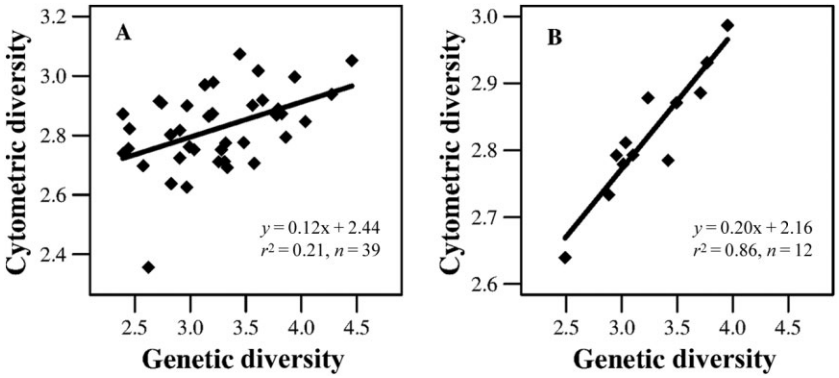

Fig. 2. Correlation analysis between genetic and cytometric diversity (Shannon–Weaver index). The solid-filled diamond symbols represent the diversity estimates performed by the genetic and cytometric methods using the 3.5 year monthly sampled database (A) and the mean annual cycle (B).

cytometric diversities were found during 2011, when relatively high cytometric diversity values were observed as compared with other years, especially from March to June (Fig. 1A).

Despite this discrepancy, the average annual cycles obtained with both methodologies were similar, with highest Shannon–Weaver index values during winter and lowest in late spring (Fig. 1B). Whereas a significant, although weak correlation, was found between individual estimates of Shannon–Weaver diversity obtained by both methods (Fig. 2A, r = 0.46, $P < 0.01$), monthly averages were highly correlated (Fig. 2B, $r = 0.93$, $P < 0.01$). These results indicate that despite some individual discrepancies, cytometric diversity can reliably capture general trends in the bacterioplankton community composition. A Mantel test confirmed a significant relationship between the similarity matrices calculated using the Bray–Curtis metric obtained by both techniques (Fig. S3, Mantel statistic r = 0.59, $P = 0.001$).

Cytometric diversity patterns in the Cantabrian Sea

Overall, the decadal seasonal dynamics for the three stations analyzed (Fig. 3) were similar to the pattern found for station E2 during the 3.5 year period (Fig. 1), with higher cytometric diversity in early winter and late autumn than in summer (Fig. 4). In order to understand the factors leading to temporal changes in surface bacterial cytometric diversity, we performed correlation analyses with different ancillary environmental variables (chlorophyll, temperature, salinity, inorganic nutrients, day length and mixed-layer depth [MLD]) using the 10 year time-series. Three variables showed significant correlations of similar strength with the cytometric diversity of surface bacterial communities: day length, inorganic nutrients, (particularly NO_2–NO_3) and MLD. The three variables showed similar correlations at the three stations studied (day length: $r = -0.69$, -0.61 and -0.54; NO_2–NO_3: $r = 0.66$, 0.62 and 0.23; MLD: $r = 0.61$, 0.64 and 0.44, for E1, E2 and E3 respectively), and as they were correlated, it is hard to determine the specific influence of each variable on surface cytometric diversity patterns. Therefore, we explored the influence of these factors on the bacterioplankton composition across another spatio-temporal dimension, the vertical gradient. In particular, for each month and station during the 10 year time-series, we compared the depth profiles of nutrient concentration and potential density with the similarity index calculated

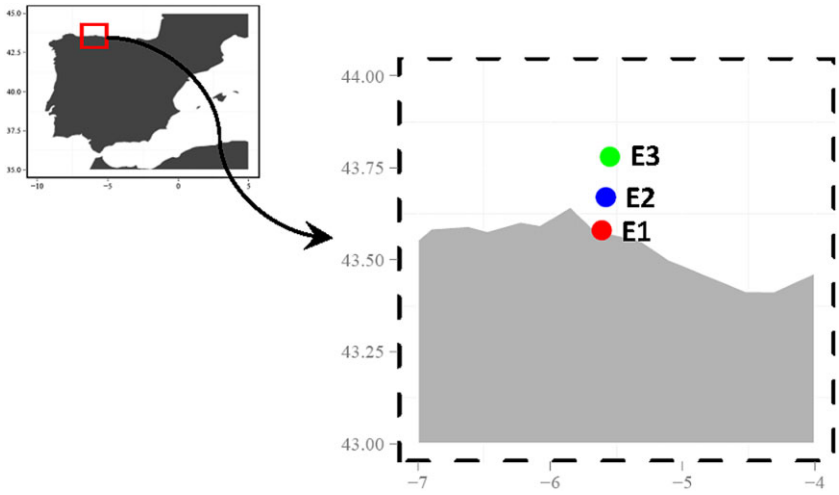

Fig. 3. Sampling stations along a coastal gradient at the Cantabrian Sea. Samples from stations E1, E2 and E3 were used for cytometric diversity estimates. Genetic diversity analyses were performed on surface samples from E2.

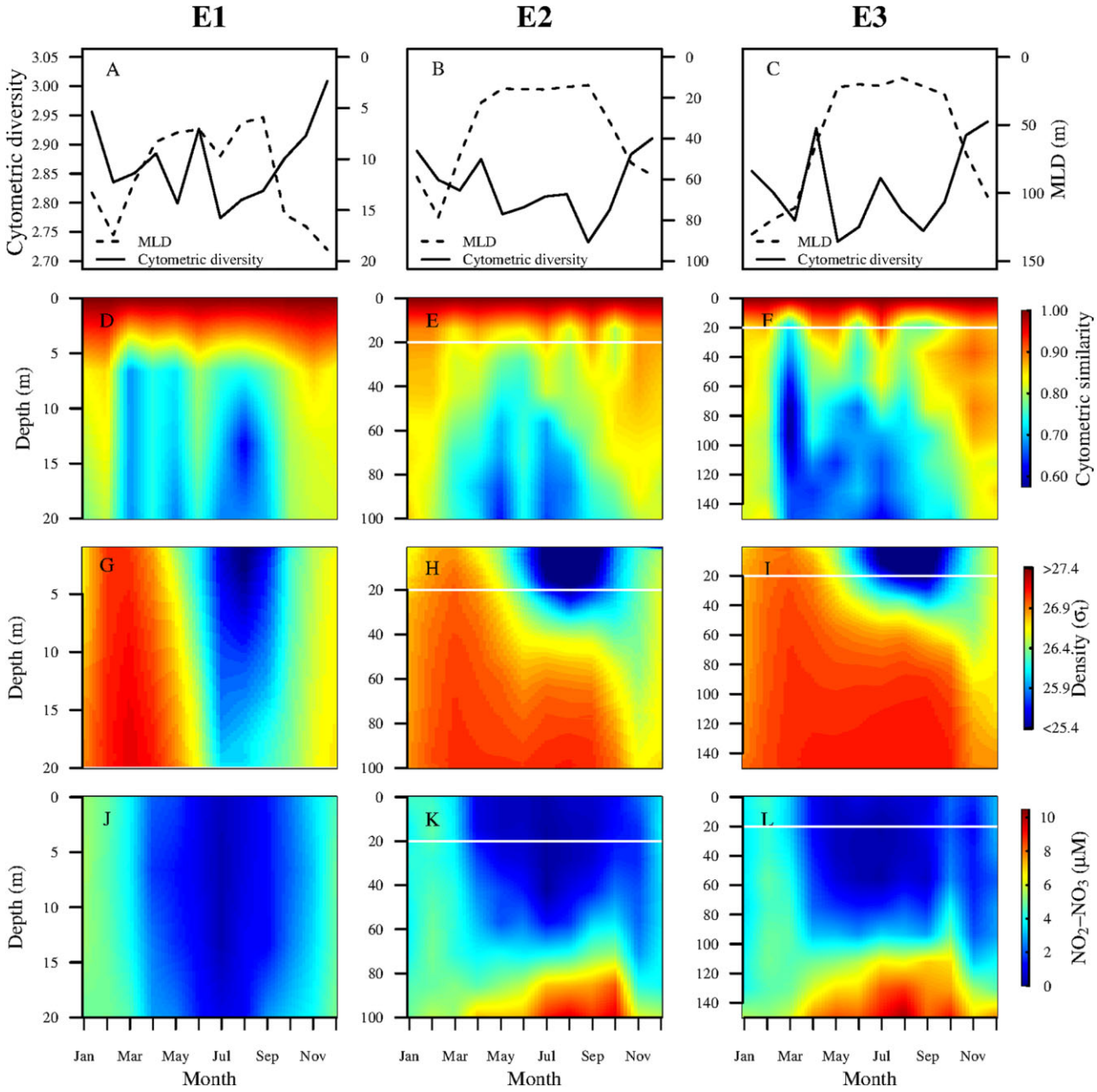

Fig. 4. Seasonal dynamics along the spatial and vertical gradient of the cytometric diversity and MLD (A–C), community similarity (D–F), potential density (G–I) and nutrient concentration (J–L) data estimated from a 10 year monthly time-series database collected in the Cantabrian Sea. Community similarity was estimated as the Bray–Curtis similarity between surface data and the different depths. The white line (E, F, H, I, K, L) represents the 20 m depth. Warm colours represent high similarity with the surface sample, and cool colours represent low similarity with the surface sample. Surface similarity value is 1 because it represents similarity with itself.

between the cytometric diversity of the surface sample and each of the remaining sample depths. We found that the similarity in cytometric diversity always decreased with depth (Fig. 4D–F).

However, during the spring–summer months this decrease in similarity was more pronounced than during late autumn–early winter. This seasonal pattern was found at the three stations but, interestingly, the decay in similarity with depth was sharper at E1 than at E2 or E3 (Fig. 4D–F). During the summer months the similarity between surface and 20 meters at E1 was around 0.6 while at E2 and E3, the similarity was higher, ca. 0.8 (see Fig. 4D and white line in Fig. 4E and F). Similarly, during summer, at E1 there was a sharp change in the cytometric similarity at around 5 m, whereas this drastic change occurred at around 20 m at E2 and E3.

Vertical nutrient profiles varied between stations and seasonally, and the shallowing of the nitracline was coincident in time with the decrease in cytometric similarity between depth and surface samples. Yet, the nitracline was deeper than the depth where the change in bacterial cytometric similarity occurred. The nutrient profiles (NO_2–NO_3) revealed a sharp gradient at *c.* 80 and 120 m at stations E2 and E3 respectively (Fig. 4J–L), whereas the gradient in density was much shallower (Fig. 4G–I). From April to September, no vertical gradient in nutrients was found at E1.

In contrast to nutrients, the MLD was coincident both in time and vertically with changes in bacterial cytometric composition, suggesting that this variable was the best to explain the cytometric similarity patterns observed at the stations (Fig. 4A–C). During late fall and winter, MLD frequently reached the bottom, whereas during spring–summer conditions, it shoaled at all stations resembling the pattern in cytometric similarity. At E1, the MLD during the summer was at around 5–10 m, whereas at E2 and E3, it shoaled up to only 20 m, similar to the patterns found for the decay in cytometric diversity at the same stations (Fig. 4D–F). The agreement between the dynamics of the MLD and cytometric diversity was also supported by the potential density profiles (Fig. 4G–I).

We used the Bray–Curtis similarity metric for all the analyses explained above. Thus, we compared similarity in terms of bacterial community composition based on cytometric categories and their relative abundance (i.e. evenness). In order to clarify whether the spatio-temporal variation found was mainly caused by changes in the cytometric categories present or by their evenness, we re-estimated the similarity matrix by using a metric based only on presence/absence data (i.e., the Jaccard's index). Hence, we replaced the abundance value in each cytometric category by 1 (when there were cells present) or 0 (when there were no cells in that category). In general, the pattern found using Jaccard's index was similar to that found by using the Bray–Curtis index (Fig. S4), suggesting that the trend was due to a change in the presence/absence of particular cytometric categories rather than in their relative abundance. In summary, although we found a high correlation between surface bacterial community composition and day length, nutrients and MLD, the analysis of vertical patterns indicated that the dynamics of the MLD is the factor (of those tested) mostly affecting the composition of bacterioplankton communities.

Discussion

Ribosomal phylogenetic markers are the *de facto* standard to analyse compositional and diversity changes in bacterial assemblages (Giovannoni and Rappé, 2000). However, high-throughput molecular approaches are still expensive and time-consuming when applied on a large number of samples, posing limitations in our knowledge of spatial and temporal patterns in oceanic microbial communities. A fast and easy processing of many microbial samples can be achieved by flow cytometry, a method that records bio-optical information of hundreds to thousands of cells per minute. While the single-cell properties obtained by flow cytometry have been used to study the diversity of phytoplankton communities (Li, 1997; 2002), its potential for the smallest-sized heterotrophic microorganisms (bacteria and archaea) remains unexplored and, in general, the power of this method to detect ecological patterns in microbial diversity is unclear. Schiaffino and colleagues (2013) estimated cytometric diversity of freshwater picoplankton by analysing distinct cytometric populations (clusters) that shared similar properties, rather than the properties of each individual cell, as proposed by Li (1997). In this study, we have adapted the methodology proposed by Li (1997) to study bacterioplankton using three cytometric properties to separate the cells falling into each cytometric category instead of the two originally used by Li (1997) for phytoplankton communities. The advantage of incorporating the third parameter is that we have another dimension to separate the cells falling in each cytometric category, thus increasing the resolution of the methodology (see Experimental procedures).

Our results show a significant agreement in the temporal pattern of bacterial community diversity using their cytometric properties and a molecular standard approach, opening the door to a promising application of the routine flow cytometry method in aquatic systems. At the compositional level, important changes in the cytometric properties of bacterial communities were also found seasonally and vertically, as commonly found by molecular analyses (Fuhrman *et al.*, 2006; Gilbert *et al.*, 2012; Giovannoni and Vergin, 2012). The significant correlation in bacterial community composition based on cytometric and genetic criteria (Fig. S3) confirms that to a certain extent, changes in the cytometric properties of cells in a community reflect changes in microbial community composition. In agreement with this idea, several studies have found that different cytometric groups have a distinct phylogenetic composition. In particular, the two main bacterial groups detected in environmental samples based on their nucleic acid content (high nucleic acid content (HNA) and low nucleic acid content (LNA)), appear to be taxonomically different (Zubkov *et al.*, 2001; Mary *et al.*, 2006; Vila-Costa *et al.*, 2012) with little overlap between them (Schattenhofer *et al.*, 2011). In particular, the widespread SAR11 clade has been associated with LNA cells in different studies (Mary *et al.*, 2006; Schattenhofer *et al.*, 2011; Morán *et al.*, 2015). Here, we found a drastic decrease in cytometric diversity in a sample collected in

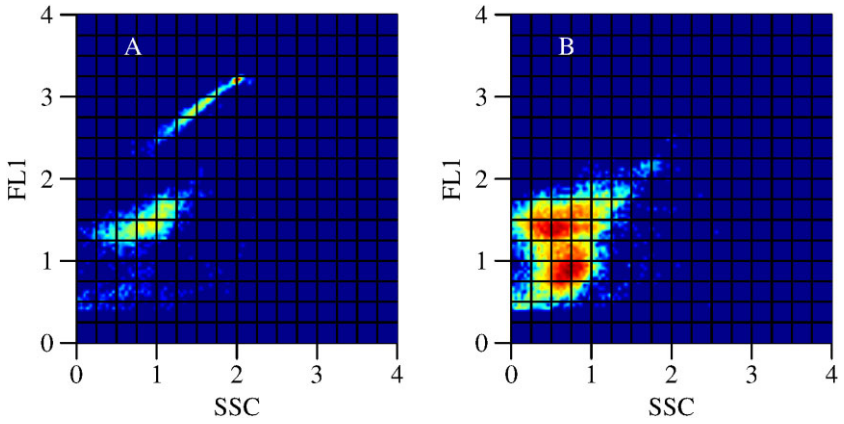

Fig. 5. Comparison between two flow cytometry standard (FCS) files corresponding to May 2012, when a low cytometric diversity was found (A) and January 2012, when a high cytometric diversity was found (B). The vertical axis is green fluorescence (FL1) and the horizontal axis is side scatter (SSC).

May 2012, when almost no LNA cells were recorded in the samples (Fig. 5A) and SAR11 populations were strongly depleted (Alonso-Sáez *et al*., 2015; Morán *et al*., 2015). This data point with the lowest cytometric diversity increased the correlation found with the Shannon–Weaver index between both methods (Fig. 2) and also in the Mantel test results (Fig. S3), although the relationship between both approaches remained significant after removing this value (Mantel test, $r = 0.44$, $P = 0.001$).

Despite the significant correlation found between both approaches, it should be acknowledged that their level of resolution is clearly different. The absolute value of the cytometric diversity index is dependent on the number of cytometric bins, which is arbitrarily selected and the degree of spread of cytometric signals. Different taxa can fall in the same cytometric category, and the same taxa may occupy different categories due to variations in morphology and cell size over their growth cycle. The use of regularly spaced divisions provided by the flow cytometer detection channels is also likely to split cells belonging to the same taxa into multiple bins and/or to create multitaxa bins. Yet, the evidence presented here is that the cytometric method can pick out large signals in community variability at coarse scales (both over time and across depths; Fig. 4). Thus, the utility of the proposed method should be enhanced at detecting large signals over spatial and/or temporal scales. This apparent limitation would be offset by its potential to provide a large coverage of data at a low cost and effort, unfeasible by molecular techniques.

In agreement with previous multiyear time-series of marine bacterioplankton based on molecular data (Fuhrman *et al*., 2006; Caporaso *et al*., 2012; Gilbert *et al*., 2012; Smyth *et al*., 2014), we found a pattern of increasing diversity during winter. Day length and, more recently, net heat flux have been shown to be the factors best correlated with patterns of bacterial diversity in the English Channel L4 station, but the mechanistic reason for this correlation remains unclear (Gilbert *et al*., 2010; Smyth *et al*., 2014).

Of the variables similarly correlated with cytometric diversity, we found that MLD was the one most closely resembling changes in cytometric diversity along the vertical profile. Moreover, the changes in bacterial cytometric composition occurring in surface waters when the stratification breaks were coincident with increases in diversity at all three stations (Fig. 4). Therefore, we hypothesize that the higher values of bacterial diversity in early winter were due to a vertical re-shuffling of bacterial taxa during the mixing period. The effects of mixing on the patterns of bacterial diversity have been already highlighted in recent studies (Vergin *et al*., 2013; Fuhrman *et al*., 2015). During stratification, taxa that adapted to thrive in oligotrophic, warmer waters outcompete other taxa, resulting in an overall decrease in diversity. During the winter months, the MLD is deeper and, therefore, the whole water column is well mixed, providing similar environmental conditions at any depth and indirectly causing an increase in the number of bacterial taxa found at the surface able to grow at the surface. The finding of rare taxa typically abundant in the deep ocean such as SAR324 and SAR406, in surface samples from station E2 during winter (Alonso-Sáez *et al*., 2015), further supports our hypothesis. However, based on our data, we cannot rule out the possibility that the increase in diversity was due to taxa that were originally in surface waters but below the detection limits, which grew favoured by other drivers associated with water column mixing not measured in this study (e.g. micronutrients or vitamins).

Changes in community composition following mixing by overturning have been described in the Bermuda Atlantic Time-Series Study (Morris *et al*., 2005). Salter and colleagues (2015) have recently reported seasonal increases in the diversity of SAR11 in the NW Mediterranean strongly associated with mixing using a 7 year time-series. However, a mechanistic explanation for the increase of bacterial diversity has been hindered by the lack of high vertical resolution in bacterial composition. The use of flow cytometry to complement molecular tech-

niques has allowed us to increase the spatio-temporal resolution of the time-series analysed and to propose the species reshuffling by vertical mixing hypothesis to explain the late autumn–early winter maxima in bacterial diversity in coastal systems.

In summary, as in the seminal application of Li (1997) to small phytoplankton, our study reveals the utility and the potential of the cytometric diversity to assess heterotrophic bacterioplankton community structure and dynamics in aquatic systems, especially for large-scale studies where the application of molecular techniques is more limited. We took advantage of the high throughput offered by flow cytometry to analyse changes in diversity over spatial, vertical and temporal scales and found that the re-shuffling of species by vertical mixing may be an important factor for the recurrent early winter maxima in bacterial diversity in temperate ecosystems of the oceans. The use of flow cytometry for studying the diversity of different components of microbial communities can be a useful complementary tool to molecular approaches to provide new insights in the understanding of global microbial community spatio-temporal distribution.

Experimental procedures

Sampling sites

Samples were collected monthly off Gijón/Xixón, from April 2002 to December 2012, as part of the programme RADIALES of the Spanish Institute of Oceanography. This 10 year time-series study includes three stations sampled along a transect perpendicular to the coast located in the continental shelf at 43.58°N, 5.61°W (Station 1, E1), 43.67°N, 5.58°W (Station 2, E2) and 43.78°N, 5.55°W (Station 3, E3) (Fig. 3). Samples for flow cytometry analysis were collected over the 10 year period at 4 depths at E1 (from the surface to 20 m depth), 8 depths at E2 (from the surface to 100 m depth) and 9 depths at E3 (from the surface to 150 m depth). DNA samples for 454 pyrosequencing were obtained over 3.5 years (July 2009 to December 2012) at the surface (0 m) at E2.

A SeaBird 25 CTD was used to collect temperature and salinity data. The density (σt) gradient was used to estimate the MLD. The MLD was defined as the depth where a change in density of 0.05 kg m^{-3} for a 5 m interval was first found. Samples for nutrients analyses (NO_2, NO_3 and PO_4) were collected concomitantly with flow cytometry samples and were frozen until analysis with a Technicon autoanalyzer following the methodology described by Grasshoff (1976).

Flow cytometry analysis and cytometric diversity estimates

Samples for flow cytometry were preserved with 1% paraformaldehyde + 0.05 % glutaraldehyde and stored at −80°C until processing, usually within 1 month of collection. For flow cytometry analysis, 0.4 ml of sample was stained with 2.5 μmol l^{-1} dimethylsulphoxide (DMSO)-diluted SYTO-13 nucleic acid fluorochrome (Molecular Probes) and analysed with a FACSCalibur flow cytometer (BD/Becton, Dickinson and Company) with a laser emitting at 488 nm. 1.0 μ dm fluorescent latex beads (Molecular Probes) were added to each sample as an internal standard. Flow cytometry standard (FCS) files were gated automatically following the methodology described in García and colleagues (2014). This methodology uses a model-based clustering for automatically detecting the different groups of cells that appears in a sample and allowed us to separate heterotrophic cells from noise or phototrophic cells. Side scatter (SSC), green fluorescence (FL1) and red fluorescence (FL3) are common flow cytometric variables used to identify bacterioplankton groups (Gasol and del Giorgio, 2000). The settings used for detecting these cytometric parameters for bacterioplankton cells were 400 nm for SSC, 511 nm for FL1 and 590 nm for FL3. These variables were used in the flowClust function for the automatic clustering of the heterotrophic bacteria populations.

The methodology used by Li (1997) to estimate phytoplankton diversity was followed and adapted to the analysis of bacterioplankton. For each bacterioplankton FCS file, we selected the clusters corresponding to heterotrophic prokaryotes.

As we specifically targeted heterotrophic bacteria, we analysed cytograms obtained from samples stained with SYTO 13. In these samples, the quantification of autotrophs is less reliable than in live-cell samples (Calvo-Díaz and Morán, 2006), and thus, we decided to exclude them from the analysis. In any case, the abundance of cyanobacteria was generally low in the samples, contributing to $3.2 \pm 4.6\%$ of total cells. The flow cytometer detection channels allow differentiating the cells according to their optical properties. SSC and FL1 signals are more important for differentiation of heterotrophic bacterioplankton communities than FL3, which results from fluorescence spillover rather than a true signal. Hence, in our analysis we gave more weight to SSC and FL1 by clustering the detection channels into 16 bins each, whereas FL3 channels were grouped into only 4 bins. The rational of using FL3 was to obtain a three-dimensional view of the community to separate heterotrophic bacteria from noise or cyanobacteria groups more efficiently (Gasol and del Giorgio, 2000). We then estimated the diversity of each sample using the number of cells falling into each of the 1024 ($16 \times 16 \times 4$) resulting categories. Figure 5 shows an example with two FCS files where SSC and FL1 parameters were divided into 16×16 bins obtained from a sample with relatively low and high cytometric diversity (Fig. 5A and B respectively).

Genetic diversity estimates by 454 pyrosequencing

Samples (0.5–2 l) for pyrosequencing were collected monthly from surface waters at E2 (Fig. 3) using Niskin bottles. Samples were pre-filtered through GF/A filters (Whatman) and filtered onto 0.2 μm diameter polycarbonate filters (Millipore). DNA extraction was performed using Power Water DNA Isolation kit (Mobio). DNA was quantified using a Nanodrop (Thermo) and bacterial sequences were amplified by the use of domain-specific bacterial primers (341F and 805R; Herlemann *et al.*, 2011) complemented with a 10 bp sample-specific bar-code sequence on the reverse primer.

Polymerase chain reaction (PCR) (20 μl) contained a final concentration of 1 μmol l^{-1} of each primer, 0.2 mmol l^{-1} of each dNTP and 0.02 nmol l^{-1} of polymerase Phusion High Fidelity Taq Polymerase. The template DNA concentration varied between 1 and 10 ng per reaction. The PCR cycles started with a 5 min initial denaturation at 98°C, followed by 25 cycles of 98°C for 40 s, 53°C for 40 s and 72°C for 60 s and a final 7 min elongation at 72°C. Triplicate PCR reactions were pooled for each sample. The purified amplicons were subject to pyrosequencing from the reverse primer using a 454 FLX + system. The raw reads were quality trimmed (with a minimum Phred average quality score of 35 over a 50 bp window) and aligned against a reference SILVA alignment, keeping only those positions that start and stop in the same alignment space. From 488 593 reads initially screened, 248 474 reads were considered of good quality and with a minimum alignment length of 140 bp for further analyses. The resulting sequences were denoised by a preclustering method allowing one mismatch (Huse *et al*., 2010) as recommended (Schloss *et al*., 2009). Chimeras were removed using UCHIME. For OTU construction, reads were clustered at genetic distance cut-offs 0.01 and 0.03 substitution per nucleotide, using the average linkage method. Taxonomic assignment of the OTUs was obtained by classification with SILVA taxonomy using the Wang approach (Wang *et al*., 2007) as implemented in MOTHUR. The average confidence in the taxonomic assignment was 99.6 ± 3.3% and 98.9 ± 5.6% for OTUs at the 0.01 and 0.03 cut-off levels. Chloroplasts and cyanobacteria identified by this method were removed from the data set, and samples were downsized to 4174 reads by random resampling to equal the depth of sequencing of all samples before calculating Shannon–Weaver diversity indices. Sequences of this study have been deposited in the European Nucleotide Archive (accession number PRJEB6399).

It should be acknowledged that due to mismatches of the reverse primer with some members of the SAR11 clade, the recovery of the total diversity of this group may have been affected. However, we consider that in our data set, the SAR11 group was still fairly well represented as it was the dominant group in our samples (contributing up to 54% of reads), and indeed, a positive correlation in the abundance of SAR11 by tag-sequencing and FISH using the SAR11-441R probe (Morris *et al*., 2002) was found (Fig. S5, Spearman Rho = 0.73, $P < 0.01$, $n = 12$).

Determination of diversity indices

We calculated Shannon–Weaver indices for the molecular and cytometric data set to compare the diversity patterns of bacterioplankton communities. For cytometric diversity, this calculation was based on the total number of cytometric categories detected and the relative abundance of cells in each category. In the molecular data set, the Shannon–Weaver diversity index was calculated based on the total number of OTUs detected (at 0.03 OTU genetic distance cut-off) and their evenness in the samples.

The similarity between samples by each method over a 3.5 year period was calculated using Bray–Curtis similarity index. A Mantel test was then performed to compare the two similarity matrices. Bacterioplankton flow cytometry data and molecular data were standardized before similarity analysis. We also calculated pairwise Bray–Curtis similarity matrices between flow cytometry samples across the depth profile for the three sampling stations. For that purpose, we assessed the vertical variability in species bacterial community composition by calculating, for each station and date, the similarity index between the cytometric composition at the surface and each of the other sampled depths. This vertical similarity analysis was applied to every monthly profile along the seasonal cycle during the 10 year time-series.

All the analyses were performed with R (Team, 2013) and statistical tests were performed using the R package 'vegan' (Oksanen *et al*., 2012).

Acknowledgements

This work was supported by Metabolic Ocean Analysis (METOCA) project funded by Spanish National Investigation + Development + Innovation (I + D + I) Plan (METOCA, CTM2009-13882-MAR), by the time-series project RADIALES from the Spanish Institute of Oceanography (IEO), the Marie Curie European Reintegration Grant 'FUNDIVERSITY' (FP7, Grant Agreement 268331), by the research grant Coastal Ocean Microbial communities and Temperature (COMITE, CTM2010-15840) and by Modelado de las reglas de ensamblado y estabilidad de los ecosistemas de comunidades planctónicas en el oceáno global (MARES – CGL2013-41256-P). Financial support was also provided by the Principado de Asturias FEDER(GRUPIN14-144). F.C.G. was funded by a Formación de Personal Investigador (FPI) grant program and L.A-S. was funded by a 'Juan de la Cierva' Fellowship from the Spanish Ministry of Economy and Competitiveness (MINECO).

References

Alonso-Sáez, L., Balague, V., Sa, E.L., Sánchez, O., González, J.M., Pinhassi, J., *et al*. (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* **60:** 98–112.

Alonso-Sáez, L., Díaz-Pérez, L., and Morán, X.A.G. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol* **17:** 3765–3779.

Amann, R., and Fuchs, B.M. (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6:** 339–348.

Buitenhuis, E.T., Li, W.K.W., Lomas, M.W., Karl, D.M., Landry, M.R., and Jacquet, S. (2012) Picoheterotroph (Bacteria and Archaea) biomass distribution in the global ocean. *Earth Syst Sci Data* **4:** 101–106.

Calvo-Díaz, A., and Morán, X.A.G. (2006) Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay. *Aquat Microb Ecol* **42:** 159–174.

Campbell, B.J., Yu, L., Straza, T.R.A., and Kirchman, D.L. (2009) Temporal changes in bacterial rRNA and rRNA genies in Delaware (USA) coastal waters. *Aquat Microb Ecol* **57:** 123–135.

Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R., and Gilbert, J.A. (2012) The Western English Channel contains a persistent microbial seed bank. *Isme J* **6:** 1089–1093.

Chow, C.-E.T., Sachdeva, R., Cram, J.A., Steele, J.A., Needham, D.M., Patel, A., *et al*. (2013) Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *Isme J* **7:** 2259–2273.

Cottrell, M.T., and Kirchman, D.L. (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl Environ Microbiol* **66:** 5116–5122.

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., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105:** 7774–7778.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13:** 133–146.

García, F.C., López-Urrutia, A., and Morán, X.A. (2014) Automated clustering of heterotrophic bacterioplankton in flow-cytometry data. *Aquat Microb Ecol* **72:** 175–185.

Gasol, J.M., and del Giorgio, P.A. (2000) Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Sci Mar* **64:** 197–224.

Gilbert, J., Somerfield, P., Temperton, B., Huse, S., and Joint, I. 2010. *Day-length is central to maintaining consistent seasonal diversity in marine bacterioplankton* [WWW document]. Available from Nature Precedings. URL http://hdl.handle.net/10101/npre.2010.4406.1.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrueck, L., Reeder, J., Temperton, B., *et al*. (2012) Defining seasonal marine microbial community dynamics. *Isme J* **6:** 298–308.

Giovannoni, S.J., and Rappé, M.S. (2000) Evolution, diversity and molecular ecology of marine prokaryotes. In *Microbial Ecology of the Oceans*. Kirchman, D.L. (ed.). New York, NY, USA: Wiley-Liss, Inc., pp. 47–84.

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

Giovannoni, S.J., and Vergin, K.L. (2012) Seasonality in ocean microbial communities. *Science* **335:** 671–676.

Grasshoff, K. (1976) *Methods of Sea-water Analysis*. Verlag Chemie, Weinheim and New York.

Green, J.L., Bohannan, B.J.M., and Whitaker, R.J. (2008) Microbial biogeography: from taxonomy to traits. *Science* **320:** 1039–1043.

Herlemann, D.P.R., Labrenz, M., Juergens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *Isme J* **5:** 1571–1579.

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

Li, W.K.W. (1997) Cytometric diversity in marine ultraphytoplankton. *Limnol Oceanogr* **42:** 874–880.

Li, W.K.W. (2002) Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature* **419:** 154–157.

Margalef, R. (1968) *Perspectives in Ecological Theory*. Chicago, IL, USA: University of Chicago Press.

Mary, I., Heywood, J.L., Fuchs, B.M., Amann, R., Tarran, G.A., Burkill, P.H., and Zubkov, M.V. (2006) SAR11 dominance among metabolically active low nucleic acid bacterioplankton in surface waters along an Atlantic meridional transect. *Aquat Microb Ecol* **45:** 107–113.

Moran, X.A.G., Alonso-Saez, L., Nogueira, E., Ducklow, H.W., Gonzalez, N., Lopez-Urrutia, A., *et al*. (2015) More, smaller bacteria in response to ocean's warming? *Proceed Biol Sci/R Soc* **282:** 20150371. http://dx.doi.org/10.1098/rspb.2015.0371

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Morris, R.M., Vergin, K.L., Cho, J.C., Rappe, M.S., Carlson, C.A., and Giovannoni, S.J. (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-Series Study site. *Limnol Oceanogr* **50:** 1687–1696.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*. (2012) *Vegan: Community Ecology Package. R package version 2.0–3*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document]. URL http://CRAN.R-project.org/package=vegan.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68:** 3094–3101.

Pielou, E. (1975) *Ecological Diversity*. New York: Wiley-Interscience.

Salter, I., Galand, P.E., Fagervold, S.K., Lebaron, P., Obernosterer, I., Oliver, M.J., *et al*. (2015) Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. *ISME J* **9:** 347–360.

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.

Schattenhofer, M., Wulf, J., Kostadinov, I., Gloeckner, F.O., Zubkov, M.V., and Fuchs, B.M. (2011) Phylogenetic characterisation of picoplanktonic populations with high and low nucleic acid content in the North Atlantic Ocean. *Syst Appl Microbiol* **34:** 470–475.

Schauer, M., Massana, R., and Pedros-Alio, C. (2000) Spatial differences in bacterioplankton composition along the Catalan Coast (NW Mediterranean) assessed by molecular fingerprinting. *FEMS Microbiol Ecol* **33:** 51–59.

Schiaffino, R.M., Gasol, J.M., Izaguirre, I., and Unrein, F. (2013) Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient. *Aquat Microb Ecol* **68:** 231–250.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Smyth, T.J., Allen, I., Atkinson, A., Bruun, J.T., Harmer, R.A., Pingree, R.D., *et al*. (2014) Ocean net heat flux influences

seasonal to interannual patterns of plankton abundance. *PLoS ONE* **9:** e98709.

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.

Team, R.C. (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document]. URL http://CRAN.R-project.org/.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al*. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

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–U129.

Vila-Costa, M., Gasol, J.M., Sharma, S., and Moran, M.A. (2012) Community analysis of high- and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA pyrosequencing. *Environ Microbiol* **14:** 1390–1402.

Wang, K., Li, M., and Bucan, M. (2007) Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* **81:** 1278–1283.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., *et al*. (2011) Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS ONE* **6:** e24570.

Zinger, L., Gobet, A., and Pommier, T. (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Mol Ecol* **21:** 1878–1896.

Zubkov, M.V., Fuchs, B.M., Burkill, P.H., and Amann, R. (2001) Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea. *Appl Environ Microbiol* **67:** 5210–5218.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Trends of genetic diversity (Shannon–Weaver index) of surface bacterial communities at E2 during a 3.5 year monthly time series (A) and as a mean annual cycle (B) when two different OTU similarity cut-off are used: 0.01 (solid line) and 0.03 (dashed line).

Fig. S2. Correlation analysis between genetic (OTU genetic distance cut-off 0.01) and cytometric diversity (Shannon–Weaver index) using the 3.5 year monthly sampled database and mean annual cycle.

Fig. S3. Bacterioplankton similarity matrix estimated using Bray–Curtis algorithm with 3.5 year monthly surface sample.

Fig. S4. Community similarity data estimated from a 10-year monthly time series database collected in the Cantabrian Sea. Community similarity was estimated as the Jaccard similarity between surface data and the different depths across the water column. Warm colours represent high similarity with the surface sample and cool colours represent low similarity with the surface sample. Surface similarity value is 1 because represent similarity with itself.

Fig. S5. Percentage of SAR11 detected by 454 pyrosequencing method (y-axis) and CARD-FISH (x-axis). Black line is the model fitting and red line represents the line (0,1).