Received Date : 27-Jul-2015

Revised Date : 29-Sep-2015

Accepted Date : 08-Oct-2015

Article type : Original Article

Title

Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes

Running title: Particle-related lifestyle in deep-sea prokaryotes

Authors: Guillem Salazar[1], Francisco M. Cornejo-Castillo[1], Encarna Borrull[1], Cristina Díez-Vives[2], Elena Lara[3], Dolors Vaqué[1], Jesús M. Arrieta[4], Carlos M. Duarte[4], Josep M. Gasol[1]* and Silvia G. Acinas[1]*

Affiliations:

1 Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC. Pg Marítim de la Barceloneta 37-49, E08003 Barcelona, Catalunya, Spain

2 School of Biotechnology and Biomolecular Sciences, Centre for Marine Bio-Innovation, The University of New South Wales, Sydney, NSW, Australia.

3 Institute of Marine Sciences (CNR-ISMAR), National Research Council, Castello 2737/F Arsenale-Tesa 104, 30122 Venezia, Italy

4 King Abdullah University of Science and Technology (KAUST), Red Sea Research Center (RSRC), Thuwal, 23955-6900, Saudi Arabia

* To whom correspondence should be addressed: E-mail: pepgasol@icm.csic.es or sacinas@icm.csic.es. Tel (+34) 93 230 8565; Fax (+34) 93 230 95 55.

Keywords: amplicon sequencing / bathypelagic microbial diversity / free-living / lifestyle / particle-attached / prokaryotic community structure /

Subject Category: Microbial population and community ecology.

Abstract

The free-living (FL) and particle-attached (PA) marine microbial communities have repeatedly been proved to differ in their diversity and composition in the photic ocean and also recently in the bathypelagic ocean at a global scale. However, although high taxonomic ranks exhibit preferences for a PA or FL mode of life, it remains poorly understood whether two clear lifestyles do exist and how these are distributed across the prokaryotic phylogeny. We studied the FL (<0.8 μm) and PA (0.8 – 20 μm) prokaryotes at 30 stations distributed worldwide within the bathypelagic oceanic realm (2,150 – 4,000 m depth) using high throughput sequencing of the small subunit ribosomal RNA gene (16S rRNA). A high proportion of the bathypelagic prokaryotes were mostly found either attached to particles or freely in the surrounding water but rarely in both types of environments. In particular, this trait was deeply conserved through their phylogeny suggesting that the deep-ocean particles

and the surrounding water constitute two highly distinct niches and that transitions from one to the other have been rare at an evolutionary time-scale. As a consequence, PA and FL communities had clear alpha- and beta-diversity differences that exceeded the global-scale geographical variation. Our study organizes the bathypelagic prokaryotic diversity into a reasonable number of ecologically coherent taxa regarding their association to particles, a first step for understanding which are the microbes responsible for the processing of the dissolved and particulate pools of organic matter that have a very different biogeochemical role in the deep ocean.

Introduction

The deep ocean contains 70% of the ocean's microbial cells and represents 60% of its heterotrophic activity (Arístegui *et al.* 2009). This activity is supported by a flux of particles produced in the upper ocean, which is considered to be the dominant input of organic carbon to the deep-sea (Ducklow *et al.* 2001; Arístegui *et al.* 2002) as dissolved organic carbon is largely consumed within the mesopelagic layer (Arístegui *et al.* 2002). Hence, the deep-sea pelagic ecosystem is not an homogenous habitat where microbes grow in suspension, but contains a variety of particles that represent important sources of organic matter fueling the dark ocean food web (Herndl & Reinthaler 2013). These include fast-sinking particles, sinking through the deep-ocean over a few weeks, as well as buoyant or slow-sinking organic particles, which remain suspended in the deep ocean over annual times scales (Herndl & Reinthaler 2013). Thus, the identification of the microorganisms inhabiting deep ocean's organic particles and the ones living freely in the water is a crucial first step for deciphering the ecological functioning of the deep ocean.

Differences between free-living (FL) and particle-attached (PA) microbial communities have been observed in relation to their local abundance and biomass, substrates incorporation or adaptation to different ecological features such as the degradation of organic matter compounds (Caron *et al.* 1982; Pedrós-Alió & Brock 1983; Fernández-Gómez *et al.* 2013). However, comparative analyses of FL and PA microbial communities have been restricted mostly to the photic ocean, where particles remain in suspension less than one month (Lande & Wood 1987) and/or to specific locations in the aphotic ocean and have used a diversity of approaches that made robust comparisons difficult among studies (Acinas *et al.* 1999; Ghiglione *et al.* 2007; Eloe *et al.* 2011; Smith *et al.* 2013; Crespo *et al.* 2013). Hence, there is a need for geographically extensive and coherent sampling efforts to examine the consistency of the differences between PA and FL communities across distant locations in the deep ocean.

The diversity and biogeography of bathypelagic prokaryotic communities has only been recently described at a global scale showing that PA and FL communities differ greatly in composition and appear to be structured by different ecological drivers (Salazar *et al.* 2015). Additionally, high taxonomic ranks (such as Orders or Phyla) have been shown to exhibit contrasting abundance patterns between FL and PA communities (Eloe *et al.* 2011; Smith *et al.* 2013; Crespo *et al.* 2013) suggesting that some degree of ecological coherence exists for high taxonomic ranks in relation to the degree of particle-association. Despite that, only the recent emergence of high-throughput sequencing techniques allows the exploration of the diversity of the FL and PA communities from a phylogenetic point of view and thus understanding whether and how these two lifestyles are linked to the prokaryotic evolutionary history.

Here we examine the phylogenetic patterns of free-living and particle-associated prokaryote communities in the global deep ocean. We do so on the basis of a global survey of samples collected between 2,000 - 4,000 m depth along the track of the Malaspina 2010 Circumnavigation Expedition (Duarte 2015). Samples were divided into two size fractions: 0.2-0.8 (FL) and 0.8-20 µm (PA) in order to operationally separate free-living cells from those attached to small particles, and analyzed by high-throughput sequencing of the16S rRNA prokaryotic genes. We combined ecological and phylogenetic analyses in order to: i) test whether two clear lifestyles (FL and PA) exist in the bathypelagic ocean and whether these are consistent at a global scale, ii) quantify the proportion of prokaryotic taxa associated with each lifestyle, iii) test the phylogenetic conservation of these two lifestyles and iv) identify the abundant and cosmopolitan members of the FL and PA communities in the dark ocean.

Material and Methods

Sample collection

A total of 60 samples of deep, bathypelagic water were obtained during the Malaspina 2010 expedition corresponding to 30 different sampling stations globally distributed across the world's oceans (Fig. S1, Table S1). We focused our efforts in sampling at a depth of 4,000 m, although a few samples were taken at shallower depths within the bathypelagic zone where orographic constraints prevented deeper sampling. Two different size fractions were analyzed in each sample to characterize each of the free-living (FL, 0.2-0.8 µm) and particle-attached (PA, 0.8-20 µm) prokaryotes. For each sample 120 L of seawater were sequentially filtered through a 200 µm and a 20 µm mesh to remove large plankton. Further filtering was done by filtering water serially through 142 mm polycarbonate membrane filters of 0.8 µm

(Merk Millipore, Isopore polycarbonate) and 0.2 µm (Merk Millipore, Express Plus) pore size with a peristaltic pump (Masterflex, EW-77410-10) obtaining a final set of 60 samples. The filters were then flash-frozen in liquid N_2 and stored at -80°C until DNA extraction. For that purpose, the filters were cut in small pieces with sterile razor blades and half of each filter was used for DNA extractions, which were performed using the standard phenol-chloroform protocol with slight modifications (Logares *et al.* 2013). Details regarding the methodological approach have been presented before (Salazar *et al.* 2015).

Sample sequencing and processing

Prokaryotic diversity was assessed using amplicon sequencing of the V4 region of the 16S rRNA gene with the Illumina MiSeq platform using paired-end reads (2 X 250 bp) and primers targeting prokaryotes (i.e. both Bacteria and Archaea). All library construction and sequencing was carried out at the JGI (www.jgi.doe.gov) following a standard protocol (Caporaso *et al.* 2011). Briefly, the variable region V4 of the 16S rRNA gene was amplified using primers F515/R806. Primer 806R has been recently shown to underestimate the abundance of SAR11 and Thaumarchaeota (Apprill *et al.* 2015; Parada *et al.* 2015). This dataset, however, was shown to be in good agreement with data derived from metagenomes, and thus not dependent on primers. Although SAR11 abundances were underestimated, Thaumarchaeota abundances derived from 16S rRNA sequencing and metagenomes were highly consistent (Salazar *et al.* 2015). Sequence processing included the removal of contaminants, disrupted pair-end reads and PhiX spike-in shotgun library reads included as internal standards, trimming and assembling of remaining pair-end reads, removal of primer sequences, quality control using sliding window and clustering at 97% identity for the construction of Operational Taxonomic Units (OTUs). Singletons (i.e. OTUs occurring once

in just one sample) and chimerical OTUs were removed. The remaining OTUs were taxonomically annotated using both the online RDP Naïve Bayesan Classifier (Wang *et al.* 2007) and the BLAST-based classifier within the QIIME pipeline (Caporaso *et al.* 2011) using the SILVA database (release 115) as reference. The best match to SILVA (minimum similarity of 70%) was used to annotate each OTU by this approach. A minimum confidence value of 90 was used as the criterion for the RDP-based annotation. An OTU abundance table was constructed containing the number of reads belonging to every OTU in each sample. Details on the sequence processing have also been described before (Salazar *et al.* 2015).

Phylogenetic reconstruction

Short sequences, such as the reads obtained through Illumina sequencing, may be problematic for phylogenetic reconstruction, especially for the satisfactorily resolving the evolutionary relations between broad taxonomic groups (Moret *et al.* 2002). However, new tools have been developed for this purpose that use reference phylogenies, usually constructed with longer sequences, and add the short reads through the use of new algorithms developed for this purpose (Matsen *et al.* 2010; Berger *et al.* 2011). This approach, which has been applied before for microbial eukaryotes (Dunthorn *et al.* 2014; Monier *et al.* 2014), bacteria (Brazelton *et al.* 2013; Larsson *et al.* 2014) and viruses (Mengual-Chuliá *et al.* 2012) was used for the present study: A phylogeny was inferred for all the representative OTU sequences with an average of 250 bp (from 219 to 278 bp) through its phylogenetic placement into a previously constructed phylogenetic tree with full-length 16S rRNA sequences. The closest sequence to each OTU in SILVA v.115 database was found and collected using BLAST (Altschul *et al.* 1990) and used for the construction of an initial phylogeny (the alignment provided by SILVA v.115 release was used). The phylogeny was

constructed using maximum likelihood inference with RAxML v. 8.0.19 (Stamatakis 2014) and the GTR evolutionary model with optimization of substitution rates and of site-specific evolutionary rates (GTRCATI). The best tree was selected from a total of 100 trees constructed for the topology and 100 extra trees were generated for computing the bootstrap values. This initial phylogeny was used for the insertion of the representative OTU sequences using the evolutionary placement algorithm (Berger *et al.* 2011) as implemented in RAxML v. 8.0.19 (Stamatakis 2014). For that purpose we used the previosuly constructed tree and an alignment containing both set of sequences (the representative OTU sequences and the sequences used for the first tree). This alignment was constructed with MOTHUR (Schloss *et al.* 2009) by aligning the first set of sequences while using the second as a reference. The alignment was trimmed to the common 16S rRNA gene fragment covered by both sets. The same evolutionary model (GTRCATI) was used for the inclusion of the representative OTU sequences within the initial phylogeny. The final phylogeny was visually inspected and 8 OTUs were removed because they corresponded to very large branches and were closely related to mithocondria sequences (confirmed using BLAST against NCBI). An additional reduced phylogeny was constructed containing only the OTUs with more than 10 reads (see motivation in *Results*).

Statistical analyses

All data treatment and statistical analyses were conducted with the R Statistical Software (R Core Team 2015) using version 3.1.0 and *vegan* (Oksanen *et al.* 2015), *ape* (Paradis *et al.* 2004), *picante* (Kembel *et al.* 2010), *geiger* (Harmon *et al.* 2008), *MASS* (Venables & Ripley 2002) and *indicspecies* (De Cáceres & Legendre 2009) packages. All the analyses were performed using an OTU abundance table that was previously sampled down

to the minimum number of reads (10,617 reads/sample) in order to avoid artifacts due to an uneven sequencing effort among samples.

Alpha and beta-diversity

We calculated prokaryotic richness/diversity metrics using two approaches: an OTU-based approach (i.e. considering the OTUs as unrelated biological entities) and a phylogenetic approach (i.e. considering the evolutionary relationships among OTUs using the complete computed phylogeny). The number of OTUs, the Chao extrapolative richness estimator (Colwell & Coddington 1994) and the Shannon entropy index (Shannon 1948) were computed as OTU-based metrics and the Faith's phylogenetic diversity (PD) (Faith 1992), the PD divided by the number of OTUs (PD/OTUs, hereafter) and the mean nearest taxon distance (MNTD) (Webb *et al.* 2002) were used as phylogenetic measures of diversity. Differences between FL and PA for richness/diversity measures were tested using Mann-Whitney test, as data normality was not assured.

The study of how prokaryotic assemblages vary along sites (i.e. beta-diversity), was also approached by using both OTU-based and phylogenetic beta-diversity distances. The Bray-Curtis dissimilarity index of community composition was used as the OTU-based beta-diversity distance, and betaMNTD (Webb *et al.* 2008) as the phylogenetic beta-diversity distance. Non-metric multidimensional scaling (NMDS) (Minchin 1987) analysis using random starts was used for visualization of beta-diversity and Permutational MANOVA (McArdle & Anderson 2001; Anderson 2001) using 1000 permutations was used to test for significant differences and to partition the beta-diversity matrix variance between FL and PA group of samples.

Particle-association niche index analyses

To numerically characterize each OTU in relation to its occurrence and relative abundance in the PA and FL sets of samples we defined a "particle-association niche index" (PAN index) for each OTU as a measure of the position of an OTU in a continuous niche space ranging from a completely free-living to a completely particle-attached lifestyle. We computed the PAN index by using an abundance-weighted mean: for a given OTU we recorded its abundance in every sample and recorded the size-fraction that every sample belonged to. FL samples were given a value of 0 and PA a value of 1. We then found the abundance-weighted mean of these values. Thus, an OTU occurring only in PA samples would have a PAN-index value of 1 and an OTU strictly occurring in FL samples would have a value of 0. An OTU equally distributed across FL and PA samples would have a PAN-index value of 0.5. This index allows positioning every OTU in a continuum describing its lifestyle preference. This approach has been previously used to define microbial niches in soil microbial ecology in relation to variables such as subsurface depth or soil mud (Stegen *et al.* 2012, 2013; Wang *et al.* 2013).

PAN-index values were compared to null model communities constructed with randomization methods in order to test whether bathypelagic prokaryotes exhibit an association to particles different from what is expected if populations had unlimited dispersal across samples (and thus, across PA and FL fractions) and were free of selection pressures. We constructed null communities by randomly permuting the counts across our abundance OTU table, maintaining row and column sums, (i.e., the total number of counts per sample and the global absolute abundance of each OTU), and thus controlling for sampling design and global taxon abundance. We permuted the matrix 1,000 times and computed the PAN index for each OTU for these null matrices. These 1,000 null PAN-index values were

compared to the real PAN-index values for each OTU. Randomizations were performed using the *quasiswap* algorithm (Miklós & Podani 2004) in *permatswap* function within the *vegan* R package.

PAN-index values were also used to explore the phylogenetic signal of the particle-association lifestyle (i.e. to ask whether closely related OTUs have similar associations to PA or FL habitats). To summarize major trends in this relationship, the between-OTU niche differences (the difference for each pair of OTUs in its PAN index) were placed in phylogenetic distance bins and the mean niche difference was computed for each bin (Stegen *et al.* 2012). A bin interval of 0.01 units was used (the maximum phylogenetic distance was 3.74 arbitrary units). This allowed the identification of the phylogenetic distance threshold beyond which niche differences no longer increased with phylogenetic distance. The phylogenetic signal was also statistically tested using Pagel's λ index (Pagel 1999), which yields values close to 0 when there is phylogenetic independence of a trait and close to 1 when species' traits are distributed as expected under a Brownian model of trait evolution (Münkemüller *et al.* 2012). Pagel's λ was estimated by maximum likelihood using the *fitContinuous* function within the *geiger* R package. To test whether the estimate was significantly different from 0 (i.e. whether there was a significant phylogenetic signal) it was compared to a model assuming a λ equal to 0 (i.e. no phylogenetic signal) using a likelihood ratio test (LRT).

Analysis of specific lineages analyses in PA and FL microbial communities

The detection of specific lineages with a PA or FL lifestyle was performed using two different strategies: one at broad taxonomic levels (Phylum and finer levels) and a second approach at the OTU level. For the first strategy, OTUs belonging to the same Phylum were grouped using SILVA-derived taxonomy. The group PAN-index values were tested for

significance using the null hypothesis that the mean PAN-index of all the OTUs belonging to a same Phylum equaled 0.5 (i.e. the expected PAN index for an OTU that is equally distributed across PA and FL samples). The significance was tested with one-sample Wilcoxon signed rank tests, as data normality was not assured. In order to assure a minimum sample size, only Phyla containing more than 40 OTUs were tested. P-values were adjusted for multiple comparisons using False Discovery Rate correction (Benjamini & Hochberg 1995). The same was done for the main lineages clearly annotated within each Phylum, which were selected according to the SILVA-based taxonomical annotation of OTUs. A second strategy consisted on detecting "indicator OTUs" for each of the two lifestyles, i.e. cosmopolitan OTUs (widely distributed OTUs across sampling stations) but restricted to the PA or FL group of samples. This was addressed using the "indicator species" approach (Dufrene & Legendre 1997; De Cáceres & Legendre 2009). Indicator OTUs for the two size-fractions were identified using the IndVal index, which is a combined measure of "specificity" (A, the proportion of the total reads of an OTU that appear in a given size-fraction) and "fidelity" (B, the proportion of samples of a given size-fraction where an OTU occurs) (De Cáceres & Legendre 2009). The significance of the association was tested using permutation tests. Those indicator OTUs with a p-value<0.05 and both, a fidelity and specificity value ≥0.8, were considered valid. This assures that potential indicator OTUs are both widely distributed among stations and restricted to any of the two size-fractions.

Results

The samples of bathypelagic prokaryotes formed two non-overlapping clusters that exactly corresponded to the PA and FL samples in an NMDS ordination space built from OTU-based Bray-Curtis distances (Fig. 1a), accounting for a third of the variance in

community composition across samples (Permutational MANOVA: F=26.295, R^2= 0.312, P-value=0.001). The mean Bray-Curtis distance between samples belonging to the same size-fraction was in both cases (i.e. in FL and PA samples) lower than the mean distance between the two size-fractions of the same station (Fig. S2; Mann-Whitney test: FL: U=11612, P-value<0.0001; PA: U=20252, P-value<0.0001).

Both, phylogenetic beta-diversity (i.e. betaMNTD) and OTU-based beta-diversity (i.e. Bray-Curtis distance) were highly correlated (Mantel test: r=0.77, P-value<0.001) across samples although, irrespectively of the absolute Bray-Curtis value, betaMNTD values within the FL group of samples tended to be lower than within PA (Fig. S3). All three metrics of OTU-based alpha-diversity used here differed significantly between FL and PA samples (Mann-Whitney test: nº of OTUs: U=745.5, P-value<0.0001; Chao richness estimator: U=714, P-value<0.0001; Shannon diversity index: U=749, P-value<0.0001) with the FL set of samples being richer and more diverse in OTUs than their PA counterpart (Fig. 1b). The three measures of phylogenetic diversity were also significantly different (Mann-Whitney test: PD: U=630, P-value=0.007; PD/ OTUs: U=44, P-value<0.0001; MNTD: U=37, P-value<0.0001) with a higher phylogenetic diversity in the FL group of samples while the phylogenetic diversity per OTU and MNTD were higher in the PA fraction (Fig. 1b).

The placement of each OTU into a continuous niche space described by the PAN-index, ranging from a completely free-living to a completely particle-attached lifestyle, showed a bimodal distribution (Fig. 2), with most OTUs accumulating in the extreme values (close to 0 or 1) in contrast to the unimodal aggregation around 0.5, with two secondary peaks at the extreme values, expected for null, randomization-based communities (Fig. 2a).

One third of the OTUs (1,163 out of 3,534) had significantly more extreme PAN index than their null expectation (i.e. the observed values were beyond the 2.5% percentile of the simulated ones to either side of the PAN range). These 1,163 OTUs represented 85.1% of the total reads. The removal of low-abundance OTUs (those having 10 or fewer reads) resulted in the disappearance of these peaks at extreme values in the null expectation but not in the real PAN-index values (Fig. 2b). For this reduced dataset, 60% of the OTUs (1,018 out of 1,712), representing an 82.1% of reads, had more extreme values than their null expectation.

Pairwise differences in PAN index between OTUs were closely and positively correlated with between-OTU phylogenetic distances across short phylogenetic distances (up to 0.6-0.8 arbitrary units, representing a 16-20% of the maximum phylogenetic distance across the entire tree), but there was no systematic relationship for OTU pairs at greater phylogenetic distances (Fig. 3a). Most of the OTUs belonging to the same Class or even Phylum (and lower taxonomic ranks) exhibited pairwise differences in the PAN index equal or lower than this value (Fig. 3b). The analysis of the phylogenetic consistency of the PAN index showed that most OTUs exhibited extreme PAN-index values (i.e. close to 0 or 1) and that these values were consistently conserved for relatively broad clades across the phylogeny (Fig. 4). The phylogenetic signal test for the whole phylogeny resulted in a Pagel's λ estimation of 0.887, significantly different from the value of 0 corresponding to the null-hypothesis (LRT: likelihood ratio=1340.132, P-value≤0.0001).

Seven out of the 15 Phyla and 16 out of the 31 lower-level lineages deviated significantly from the mean PAN index of 0.5 expected in the absence of preference toward the FL or PA life styles (Fig. 5, Table S2 and S3). These differences were also evident by the

positioning of the OTUs belonging to each of the Phyla or finer lineages within the NMDS space (Fig. S4 and Fig. S5). Deferribacteres, Choloroflexi, Euryarchaeota and Thaumarchaeota and all the lower-level lineages tested within them (e.g. SAR406, SAR202, Marine Group I and II) had mean PAN indexes significantly < 0.5, indicating a clear preference for the free-living lifestyle. Other lineages (Arctic 97B-4, SAR324, SAR86, AEGEAN-169, DB1-14 and BD2-11) also had a significantly lower mean PAN index compared to the null hypothesis but the Phyla they belong to did not (Verrucomicrobia, Deltaproteobacteria, Gammaproteobacteria, Alphaproteobacteria, and Gemmatimonadetes). Planctomycetes, Bacteroidetes and Firmicutes had mean PAN index significantly greater than 0.5 supporting a preference for the particle-attached lifestyle. All the lineages tested within Planctomycetes (Planctomycetaceae, Phycisphaeraceae and OM90) had also significantly high mean PAN index, together with Desulfuromonadales and OM27 (Deltaproteobacteria) and the SHA-109 lineage (Cyanobacteria). Despite Bacteroidetes, as a Phylum, had a mean PAN index significantly greater than 0.5, none of the lineages tested within it yielded significant departures from 0.5 after FDR correction (although they had significant or almost significant P-values before correction). As OTUs within Firmicutes did not group into finer lineages that could be clearly annotated by SILVA, none were included in the analysis.

We detected 100 OTUs with value as indicator OTUs for the FL and 35 for the PA lifestyles (phylogenetically placed in Fig. 4; analysis results in Table S4 and S5). SAR324, SAR406, SAR202, Marine Group I and II dominated the indicator OTUs for the FL lifestyle. These same lineages comprised the 30 most abundant OTUs across the whole dataset (Fig. S6). PA indicator OTUs were mainly composed of several representatives of Planctomycetes, Alphaproteobacteria, Deltaproteobacteria and Gammaproteobacteria.

Discussion

Comparison of the free-living and particle attached microbial communities in the bathypelagic ocean

The results presented here demonstrate niche partitioning in deep-sea prokaryotes as reflected in clear differences in the composition of free-living and particle-associated bathypelagic bacteria. This confirms previous evidence from surface waters in various marine sites (Delong *et al.* 1993; Acinas *et al.* 1999; Hollibaugh & Wong 2000; Kellogg & Deming 2009) as well as indications for the limited set of deep-ocean communities examined in the past (Ghiglione *et al.* 2007; Eloe *et al.* 2011; Smith *et al.* 2013; Crespo *et al.* 2013). Free-living and particle-attached fractions had consistent beta-diversity differences at the scale of the global bathypelagic ocean examined here (Fig. 1a), with these differences accounting for a considerable proportion (~31%) of the variance in community composition (Salazar *et al.* 2015). Moreover the communities within any of the two size-fractions from any location worldwide were, on average, more similar than the FL and PA communities within the same location (Fig. S2). That these differences were consistent both with an OTU-based and a phylogenetic-based approach to characterize beta-diversity (Fig. S3) confirms sufficiently strong niche partitioning between the two size fractions (i.e. affecting a sufficiently large number of prokaryotic species) as to be detected at the community level. Indeed, community differences between these two lifestyles were stronger than geographical variation within each lifestyle at the global scale (Salazar *et al.* 2015).

The free-living and particle-associated bathypelagic communities exhibited also significant differences in alpha-diversity, with the FL communities being, on average, richer and more diverse at an OTU-based (Fig. 1b) and phylogenetic-based level than their PA

counterparts. Phylogenetic Diversity was defined as the sum of the lengths of all those branches in the phylogeny that are members of the target sample and, thus, depends on the numbers of OTUs, which is different for both FL and PA fractions. Indeed, a higher phylogenetic diversity, standardized to the number of OTUs, was observed within PA communities compared to FL. Most studies in surface waters (Acinas *et al.* 1999; Hollibaugh & Wong 2000; Ghiglione *et al.* 2007; Kellogg & Deming 2009) also reported higher gross OTU richness for free-living bacteria. Our result contrasts with a recent study in the Northwestern Mediterranean Sea based also on 16S rRNA sequencing in which PA assemblages were found to be richer in OTUs than the FL fraction, both in photic and aphotic samples (Crespo *et al.* 2013) , yet the deep samples in the Mediterranean Sea maintain a relatively high temperature year-round (12 ºC) and deep mixing to the bottom of the bathypelagic is a frequent phenomenon in the area (MEDOC Group 1970). The samples we analyzed had average potential temperatures of 1.4 ºC and never mixed with surface waters. In addition, the FL communities in the present study were less diverse communities when taking into account the phylogenetic relatedness of the OTUs, indicating that, on average, the FL communities are composed of more closely related taxa than their PA counterparts, a pattern that, to our knowledge, had not been described in the past.

The distinction between PA and FL microbes has traditionally been made by size fractionation using a variety of filter's pore sizes ranging from 0.5 µm to 5 µm and no consensus exists on the optimal filter pore size. And it is not clear whether an optimal size can be applicable to different ecosystems. In this work, the PA prokaryotes were defined as those retained in a 0.8 µm filter, this cut-off has been repeatedly used in many other studies (Crump *et al.* 1999; Ghiglione *et al.* 2009; Allen *et al.* 2012). The existence of clear alpha- and beta-diversity differences between PA and FL communities indicates that the 0.8 µm

delineation was effective. The PA fraction includes prokaryotes attached to particles and may also include endosymbiotic or parasitic prokaryotes within small protists (<20 μm) , as well as some elongated or aggregated microbes if they are present. Although some authors have observed large cells, individual cells larger than 0.8 μm do not seem to be abundant in the bathypelagic ocean. Our measured biovolumes for bacteria from these same samples (details not shown) corresponded to a mean diameter of 0.495 μm assuming spherical shape (min = 0.417 μm; max = 0.631 μm), thus lower than the 0.8 μm cutoff. However, the existence of some organisms presenting very elongated cells can not be discarded, and thus could be partially retained in the PA fraction.

Attachment to particles or living freely is a phylogenetically conserved trait

Our next goal was to define an index for the strength of the association of an OTU to particles, providing statistical evidence of the significance of the association of a given OTU to the attached or free-living lifestyle. For that purpose we defined the "particle-association niche index" (PAN index) (see *Material and Methods*) by applying an approach used in the past to delineate "microbial niches" from co-occurrence with specific sets of environmental variables (Stegen *et al.* 2012, 2013). Consistent with the observed differences in alpha and beta-diversity for the two lifestyles, the distribution of the particle-association niche index (PAN index) departed from the randomization-based null model expectation (Fig. 2a). This observation implies that the empirical PAN-index values are differently distributed from the distribution expected for communities assembled under unlimited dispersal between size fractions and under the absence of selective differences between the attached and free-living lifestyles. The PAN-index showed a preference for associations of OTUs to either the free-living or particle-attached fractions higher than expected by chance for a third of the OTUs,

which represented 85.1% of the total reads. However, the PAN index is unreliable for OTUs represented by a few reads, for which the chance of finding them only in one size fraction, even if they had a random distribution across samples, is not negligible: being *i* the number of reads of an OTU across the whole dataset, the probability (P) of finding the i reads in samples belonging only to one size fraction (that is, half of the samples) under a random distribution is $P=0.5^i$; for i=5 reads, P=3.13% and for i=10 reads, P=0.09%. Thus, statistically-robust tests of association with either lifestyle are not possible based on the PAN index for rare OTUs . As a result, the OTUs with 10 or less reads across the whole dataset were excluded from this analysis (Fig. 2b). The majority of the deep-ocean prokaryotes (60%), thus, showed a preference to either be attached to particles or free-living with only about 40% of OTUs being randomly distributed among fractions. These results provide strong evidence for the existence of a dichotomous lifestyle for most bathypelagic prokaryotes regarding the association to particles in the deep-ocean. Indeed, that is the case for the majority of the 30 most abundant OTUs of the dataset, that, irrespectively of their distribution along stations, are abundant only in one of the two size-fractions (Fig. S6A-D). However, members of abundant OTUs such as the genera *Alteromonas*, *Alcanivorax* or *Pseudoalteromonas* were found to be evenly distributed between size-fractions (Fig. S6A-D).

Previous indications suggested that high bacterial taxonomic ranks have consistent lifestyles regarding their association to particles (Eloe *et al.* 2011; Smith *et al.* 2013; Crespo *et al.* 2013). Despite this evidence, the phylogenetic coherence of the particle-association lifestyle, i.e. the hypothesis that closely related prokaryotes have similar association to particles, has never been formally tested. Our results confirmed that closely related prokaryotes exhibit a similar lifestyle in relation to particle attachment, whereas the coherence between their lifestyle decreased with increasing phylogenetic distance between

OTUs. This positive linear relation between phylogenetic proximity and lifestyle held up to a 0.6-0.8 phylogenetic distance units (corresponding to a 16-20% of the maximum distance), a threshold that corresponds to the distance separating most of the OTUs belonging to the same Class or even Phylum (Fig. 3b). A secondary increase in the relationship between pairwise OTU-distance and PAN index differences was observed at phylogenetic distances >2 units (Fig. 3a), corresponding to the comparison of pairs of OTUs belonging to different domains (i.e. Bacteria and Archaea). This evidence of coherence in lifestyles across phylogenetic levels represent a pioneer effort at testing the phylogenetic signal of a specific prokaryotic niche, consistent with the scattered observations of a significant phylogenetic conservation of other niche descriptors, such as abundance profiles through time-series for the Baltic Sea bacterioplankton (Andersson *et al.* 2010). Our result supports the hypothesis that high prokaryotic taxonomic ranks could be ecologically coherent (Philippot *et al.* 2010; Koeppel & Wu 2012), at least for some niche axes, and identifies the free-living/particle-attached axis as one of those showing phylogenetic coherence for bathypelagic prokaryotes at the global scale.

The fact that FL and PA prokaryotic communities are consistently composed of distinct members across a worldwide survey, together with the observation that these differences in composition are phylogenetically conserved at a Class/Phylum level suggests that the deep-ocean's particles and the water surrounding them are two highly distinct environments that impose a trade-off for a majority of the bathypelagic prokaryotes, which seem to rarely be able to adapt to both environments. From an evolutionary point of view, transitions from one lifestyle to the other, thus, seem to have been rare in the extended evolutionary history of deep-sea prokaryotes, as has also occurred for other ecological barriers, such as the marine-freshwater transitions (Logares *et al.* 2009). Our results are also

consistent with the fact that complex traits are more deeply conserved in the phylogeny than traits that depend on a few genes (Martiny *et al.* 2013), as a complex set of functions seems to be responsible for the general trophic strategy of marine bacteria (Lauro *et al.* 2009) and specifically for attachment to particles (González *et al.* 2008; Ivars-Martinez *et al.* 2008; Fernández-Gómez *et al.* 2013). This metabolic complexity associated to the PA lifestyle, jointly with the fact that marine particles are embedded into the water where the FL prokaryotes inhabit, suggests the predominance of selective pressures in the maintenance of the phylogenetic pattern described here. That is, reduced dispersion between the FL and PA habitat seems unlikely to have maintained the long-term isolation necessary for the phylogenetic conservation of the two lifestyles. Thus, the most likely mechanism maintaining such a pattern would be the existence of strong selection for the FL and PA populations within their respective habitats. The capacity of PA lifestyle, seems to depend on relatively complex metabolic machineries, and thus, would not be easily transferred by horizontal gene transfer. This combination would explain the maintenance of two pools of phylogenetically distant prokaryotes adapted to either a FL or a PA lifestyle.

Taxonomic affiliation of the PA and FL members in the bathypelagic ocean

The existence of 7 out of 15 Phyla with a clear and consistent association either to the FL or PA mode of life supports the conclusion that particle-associated lifestyle is a deeply conserved trait, as proved before. The archaeal domain was highly restricted to a FL lifestyle: both Thaumarchaeota and Euryarchaeota at a Phylum level had a significant association with the FL lifestyle and representative OTUs of Marine Group I and II dominated the indicator OTU list for the FL fraction. A lifestyle associated with particles should provide access to organic substrates supporting the microbe's requirements, compared to the diluted pool of

organic carbon that limits growth of free-living prokaryotes (Arrieta *et al.* 2015). The FL lifestyle of bathypelagic Archaea is consistent with their proven capacity to grow autotrophically, presumably linked to the oxidation of ammonia (Könneke *et al.* 2005; Swan *et al.* 2014) or through the incorporation of simple organic compounds such as amino acids (Ouverney & Fuhrman 2000) or urea (Alonso-Saez *et al.* 2012; Swan *et al.* 2014). Although the primers used in this study are known to underestimate some Thaumarchaeota lineages (Parada *et al.* 2015) we did not observe such underestimation when metagenomic data of these same samples (and thus, without primer biases) was used for comparison (Salazar *et al.* 2015). However all the analyses in the current study are based on abundance comparisons between FL and PA samples and the possible overestimation/underestimation of some lineage abundances should occur in both sets of samples and thus, would not affect our conclusions.

The bacterial OTUs associated with the FL fraction corresponded to those found to be abundant and specific of bathypelagic waters in the few studies where different sampling depths have been analyzed (but without size fractionation), such as representatives of the SAR86, SAR324, SAR406 and SAR202 clades (Agogué *et al.* 2011; Ghiglione *et al.* 2012). This suggests that FL prokaryotes would make up the bulk of the microbial populations in bathypelagic waters while cells attached to particles would constitute a minor fraction of total abundances, as reported for mesopelagic environments (Kirchman & Mitchell 1982; Turley & Mackie 1995; Ghiglione *et al.* 2007). However, PA prokaryotes seem to be highly active (Kirchman 1993; Ghiglione *et al.* 2007) and thus play a very relevant ecological role, in spite of their lower abundance. In the present study, Bacteroidetes, Firmicutes and Planctomycetes clearly exhibited a consistent PA lifestyle. The preference of Bacteroidetes for the degradation of polymers (Cottrell & Kirchman 2000) and for a PA lifestyle had been also reported before for surface waters (Delong *et al.* 1993).

The differentiation between the particle-associated and free-living lifestyles is consistent with the differences in the biochemical composition between deep particulate and dissolved organic materials. Marine particles are concentrated sources of polymeric material (Minor *et al.* 2003) while the deep oceanic dissolved organic carbon available to free-living bacteria consists mainly of small molecular-size, very diluted molecules (Arrieta *et al.* 2015; Hansman *et al.* 2015). Planctomycetes members are known to be specialized degraders of marine snow and thus play a key role in global carbon turnover (Woebken *et al.* 2007). In fact, most of the indicator OTUs from this Phylum belonged to the well-known *Rhodopirellula* genus, whose genome sequence revealed a large number of genes involved in the breakdown of sulfated polysaccharides (Glöckner *et al.* 2003). The preference for a PA lifestyle for these three Phyla had been previously detected in a single deep sample from the Mediterranean Sea (Crespo *et al.* 2013) and another sample from the Puerto Rico Trench (Eloe *et al.* 2011). Here we report that the association to particles of these three Phyla, jointly with the hitherto unknown association of Deltaproteobacteria clades OM27 and Desulfuromonadales with particles, seems to be a globally consistent feature of the bathypelagic ocean.

Conclusions

In summary, a high proportion of the bathypelagic prokaryotes seem to be either adapted to a particle-attached lifestyle or to a free-living lifestyle but rarely dominate both niches. Moreover, these two lifestyles are highly conserved in the phylogeny of deep-ocean prokaryotes suggesting that transitions from one to the other have been rare at an evolutionary time-scale. Consequently, PA and FL communities present both beta-diversity and alpha-diversity differences and are dominated by distantly related prokaryotic lineages,

identified in the present work. These results provide a first step towards the organization of the deep-ocean prokaryotic diversity into an parsimonious set of ecologically-coherent assemblages with distinct biogeochemical roles, processing carbon associated with suspended particles and the highly diluted pool of dissolved organic carbon in the dark ocean, each requiring specific sets of capacities and metabolic potentials that our results suggest evolved early in the phylogenetic history of prokaryotes.

Acknowledgements

This research was funded by the Spanish Ministry of Economy and Competitiveness Science and Innovation through the Consolider-Ingenio program (project Malaspina 2010 Expedition, ref. CSD2008 – 00077). We thank our fellow scientists and the crew and chief scientists of the different cruise legs for the smooth operation. Sequencing at the JGI was supported by U.S. Department of Energy (DOE) JGI 2011 Microbes Program grant CSP 387 grant to S.G.A. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We also thank R. Logares and the Barcelona Supercomputer Center (BSC) for providing access to the MareNostrum Supercomputer (grants BCV- 2011-2-0003/3-0005, 2012-1-0006/2-0002 to R.Logares) and for his advice in phylogeny construction. Additional funding was provided by the Spanish Ministry of Science and Innovation grant CGL2011-26848/BOS MicroOcean PANGENOMICS and the Spanish Ministry of Economy and Competitivity grant MALASPINOMICS (CTM2011-15461-E). GS and FMCC held a Ph.D. JAE-Predoc (CSIC) and an FPI (MICINN) fellowships respectively SGA was supported by a project FP7-OCEAN-2011 "Micro B3".

References

Acinas SG, Antón J, Rodríguez-Valera F (1999) Diversity of Free-Living and Attached Bacteria in Offshore Western Mediterranean Waters as Depicted by Analysis of Genes Encoding 16S rRNA. *Applied and Environmental Microbiology*, **65**, 514–522.

Agogué H, Lamy D, Neal PR, Sogin ML, Herndl GJ (2011) Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Molecular Ecology*, **20**, 258–274.

Allen LZ, Allen EE, Badger JH *et al.* (2012) Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *The ISME Journal*, **6**, 1403–1414.

Alonso-Saez L, Waller AS, Mende DR *et al.* (2012) Role for urea in nitrification by polar marine Archaea. *Proceedings of the National Academy of Sciences*, **109**, 17989–17994.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology*, **215**, 403–10.

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.

Andersson AF, Riemann L, Bertilsson S (2010) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *The ISME journal*, **4**, 171–81.

Apprill a, McNally S, Parsons R, Weber L (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, **75**, 129–137.

Arístegui J, Gasol JM, Duarte CM, Herndl GJ (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography*, **54**, 1501–1529.

Ar´stegui J, Duarte CM, Agust´ S *et al.* (2002) Dissolved Organic Carbon Support of Respiration in the Dark Ocean. *Science* , **298** , 1967.

Arrieta JM, Mayol E, Hansman RL *et al.* (2015) Dilution limits dissolved organic carbon utilization in the deep ocean. *Science*, **348**, 331–333.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic biology*, **60**, 291–302.

Brazelton WJ, Morrill PL, Szponar N, Schrenk MO (2013) Bacterial Communities Associated with Subsurface Geochemical Processes in Continental Serpentinite Springs. *Applied and Environmental Microbiology*, **79**, 3906–3916.

De Cáceres M, Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology*, **90**, 3566–74.

Caporaso JG, Lauber CL, Walters W a *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, **108 Suppl** , 4516–22.

Caron DA, Davis PG, Madin LP, Sieburth, McN. J (1982) Heterotrophic Bacteria and Bacterivorous Protozoa in Oceanic Macroaggregates. *Science*, **218**, 795–797.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **345**, 101–18.

Cottrell MT, Kirchman DL (2000) Natural Assemblages of Marine Proteobacteria and Members of the Cytophaga-Flavobacter Cluster Consuming Low- and High-Molecular-Weight Dissolved Organic Matter Natural Assemblages of Marine Proteobacteria and Members of the Cytophaga-Flavobacter Cluster C.

Crespo BG, Pommier T, Fernández-Gómez B, Pedrós-Alió C (2013) Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. *Microbiology Open*, **2**, 541–52.

Crump BC, Armbrust E V, Baross JA (1999) Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. *Applied and environmental microbiology*, **65**, 3192–204.

Delong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached marine bacterial assemblages. , **38**, 924–934.

Duarte CM (2015) Seafaring in the 21St Century: The Malaspina 2010 Circumnavigation Expedition. *Limnology and Oceanography Bulletin*, **24**, 11–14.

Ducklow HW, Steinberg DK, Buesseler KO (2001) Upper ocean carbon export and the biological pump. *Oceanography*, **14**, 50–58.

Dufrene M, Legendre P (1997) Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach. *Ecological Monographs*, **67**, 345–366.

Dunthorn M, Otto J, Berger SA *et al.* (2014) Placing Environmental Next-Generation Sequencing Amplicons from Microbial Eukaryotes into a Phylogenetic Context. *Molecular Biology and Evolution*, **31**, 993–1009.

Eloe EA, Shulse CN, Fadrosh DW *et al.* (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environmental microbiology reports*, **3**, 449–58.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.

Fernández-Gómez B, Richter M, Schüler M *et al.* (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *The ISME journal*, **7**, 1026–37.

Ghiglione J-F, Conan P, Pujo-Pay M (2009) Diversity of total and active free-living vs. particle-attached bacteria in the euphotic zone of the NW Mediterranean Sea. *FEMS microbiology letters*, **299**, 9–21.

Ghiglione J-F, Galand PE, Pommier T *et al.* (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 17633–8.

Ghiglione JF, Mevel G, Pujo-Pay M *et al.* (2007) Diel and seasonal variations in abundance, activity, and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. *Microbial ecology*, **54**, 217–31.

Glöckner FO, Kube M, Bauer M *et al.* (2003) Complete genome sequence of the marine planctomycete Pirellula sp. strain 1. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8298–303.

González JM, Fernández-Gómez B, Fernàndez-Guerra A *et al.* (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8724–9.

Hansman RL, Dittmar T, Herndl GJ (2015) Conservation of dissolved organic matter molecular composition during mixing of the deep water masses of the northeast Atlantic Ocean. *Marine Chemistry*.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics (Oxford, England)*, **24**, 129–31.

Herndl GJ, Reinthaler T (2013) Microbial control of the dark end of the biological pump. *Nature Geoscience*, **6**, 718–724.

Hollibaugh JT, Wong PS (2000) Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. *Aquatic Microbial Ecology*, **21**, 103–114.

Ivars-Martinez E, Martin-Cuadrado A-B, D'Auria G *et al.* (2008) Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *The ISME journal*, **2**, 1194–212.

Kellogg C, Deming J (2009) Comparison of free-living, suspended particle, and aggregate-associated Bacterial and Archaeal communities in the Laptev Sea. *Aquatic Microbial Ecology*, **57**, 1–18.

Kembel SW, Cowan PD, Helmus MR *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics (Oxford, England)*, **26**, 1463–4.

Kirchman DL (1993) Leucine incorporation as a measure of biomass production by heterotrophic bacteria. In: *Handbook of methods in aquatic microbial ecology. Lewis* , pp. 509–512.

Kirchman D, Mitchell R (1982) Contribution of Particle-Bound Bacteria to Total Microheterotrophic Activity in Five Ponds and Two Marshes Contribution of Particle-Bound Bacteria to Total Microheterotrophic Activity in Five Ponds and Two Marshes.

Koeppel AF, Wu M (2012) Lineage-dependent ecological coherence in bacteria. *FEMS microbiology ecology*, **81**, 574–82.

Könneke M, Bernhard AE, de la Torre JR *et al.* (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, **437**, 543–546.

Lande R, Wood AM (1987) Suspension times of particles in the upper ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, **34**, 61–72.

Larsson J, Celepli N, Ininbergs K *et al.* (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *The ISME Journal*, **8**, 1892–1903.

Lauro FM, McDougald D, Thomas T *et al.* (2009) The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 15527–33.

Logares R, Bråte J, Bertilsson S *et al.* (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends in microbiology*, **17**, 414–22.

Logares R, Sunagawa S, Salazar G *et al.* (2013) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology*, **16**, 2659–71.

Martiny AC, Treseder K, Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *The ISME journal*, **7**, 830–8.

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

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, **82**, 290–297.

MEDOC Group (1970) Observation of Formation of Deep Water in the Mediterranean Sea. *Nature*, **227**, 1037–1040.

Mengual-Chuliá B, García-Pérez R, Gottschling M, Nindl I, Bravo IG (2012) Novel animal papillomavirus sequences and accurate phylogenetic placement. *Molecular phylogenetics and evolution*, **65**, 883–91.

Miklós I, Podani J (2004) Randomization of presence–absence matrices: comments and new algorithms. *Ecology*, **85**, 86–92.

Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, **69**, 89–107.

Minor EC, Wakeham SG, Lee C (2003) Changes in the molecular-level characteristics of sinking marine particles with water column depth. *Geochimica et Cosmochimica Acta*, **67**, 4277–4288.

Monier A, Comte J, Babin M *et al.* (2014) Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *The ISME Journal*, **9**, 990–1002.

Moret BE, Roshan U, Warnow T (2002) Sequence-Length Requirements for Phylogenetic Methods. In: *Algorithms in Bioinformatics SE - 26* Lecture Notes in Computer Science. (eds Guigó R, Gusfield D), pp. 343–356. Springer Berlin Heidelberg.

Münkemüller T, Lavergne S, Bzeznik B *et al.* (2012) How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**, 743–756.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: Community Ecology Package.

Ouverney CC, Fuhrman JA (2000) Marine Planktonic Archaea Take Up Amino Acids. *Applied and Environmental Microbiology*, **66**, 4829–4833.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Parada A, Needham DM, Fuhrman JA (2015) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environmental microbiology*.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Pedrós-Alió C, Brock TD (1983) The importance of attachment to particles for planktonic bacteria. *Archiv für Hydrobiologie*, **98**, 354–379.

Philippot L, Andersson SGE, Battin TJ *et al.* (2010) The ecological coherence of high bacterial taxonomic ranks. *Nature reviews. Microbiology*, **8**, 523–9.

R Core Team (2015) R: A language and environment for statistical computing.

Salazar G, Cornejo-Castillo FM, Benítez-Barrios V *et al.* (2015) Global diversity and biogeography of deep-sea pelagic prokaryotes. *The ISME Journal*. doi: 10.1038/ismej.2015.137

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, **75**, 7537–41.

Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**, 379–423.

Smith MW, Zeigler Allen L, Allen AE, Herfort L, Simon HM (2013) Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. *Frontiers in Microbiology*, **4**, 1–20.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, **30**, 1312–3.

Stegen JC, Lin X, Fredrickson JK *et al.* (2013) Quantifying community assembly processes and identifying features that impose them. *The ISME Journal*, **7**, 2069–2079.

Stegen JC, Lin X, Konopka AE, Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME journal*, **6**, 1653–64.

Swan BK, Chaffin MD, Martinez-Garcia M *et al.* (2014) Genomic and Metabolic Diversity of Marine Group I Thaumarchaeota in the Mesopelagic of Two Subtropical Gyres (L Randau, Ed,). *PLoS ONE*, **9**, e95380.

Turley CM, Mackie PJ (1995) Biogeochemical significance of attached and free-living bacteria and the flux of particles in the NE Atlantic Ocean. , **115**, 191–203.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, New York.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, **73**, 5261–7.

Wang J, Shen J, Wu Y *et al.* (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *The ISME journal*, **1**–12.

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics (Oxford, England)*, **24**, 2098–100.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Woebken D, Teeling H, Wecker P *et al.* (2007) Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes. *The ISME journal*, **1**, 419–35.

Data Accessibility

- All raw sequences used in this study are publicly available at the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) under accession ID SRP031469.
- The OTU abundance table, the alignment and the phylogenetic tree used for the analyses are available at Dryad under DOI. 10.5061/dryad.t6hh8
- The R script used for the whole set of analyses in this work is publically available at https://github.com/GuillemSalazar

Author Contributtions

J.M.A., J.M.G., C.D. and S.G.A participated in the design of the expedition and designed the sampling scheme. G.S., F.M.C., E.B., C.D., E.L. and D.V prepared the samples on board. G.S. and F.M.C. extracted the DNA. G.S., J.M.G. and S.G.A. conceived this work. G.S. analyzed the data and wrote this article. All authors made comments and suggestions to the text.

Figure's captions

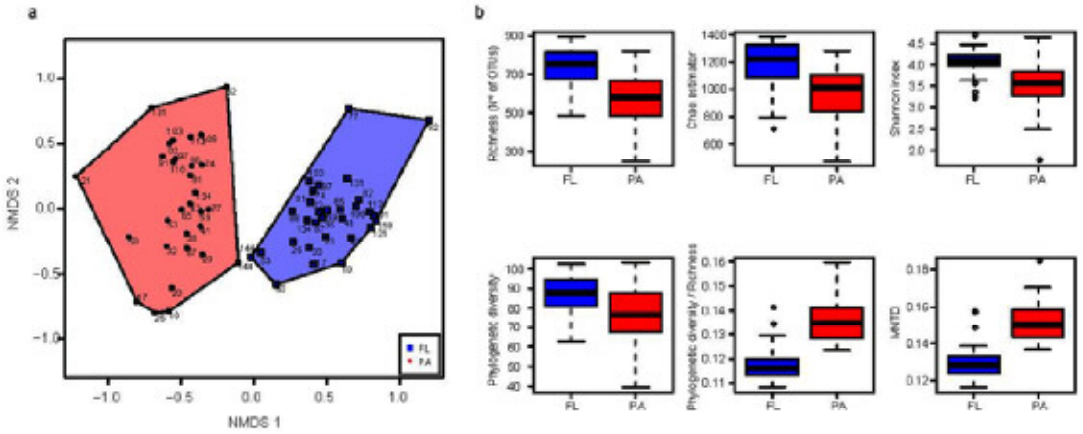
Figure 1: Beta-diversity and alpha-diversity of deep-sea prokaryotic communities. a) Beta-diversity visualized using Non-Metric Multidimensional Scaling (NMDS). Samples belonging to particle-attached (PA) and free-living (FL) are color-coded. The number close to each sample corresponds to the sampling station (see Fig. S1 and Table S1). b) Alpha-diversity measures using OTU-based (top panels) and phylogenetic (bottom panels) approaches.

Figure 2: Histogram of the distribution of real particle-associated niche index (PAN index) for each OTU compared to the null community model expectation (see *Materials and Methods* for details) based on 1,000 randomizations. a) Comparison done using the whole dataset and b) a reduced version using the OTUs with more than 10 reads. The lines correspond to kernel density estimates for each distribution: the null distribution (grey bars and the solid line) and the real distribution (hatched bars and dashed line).

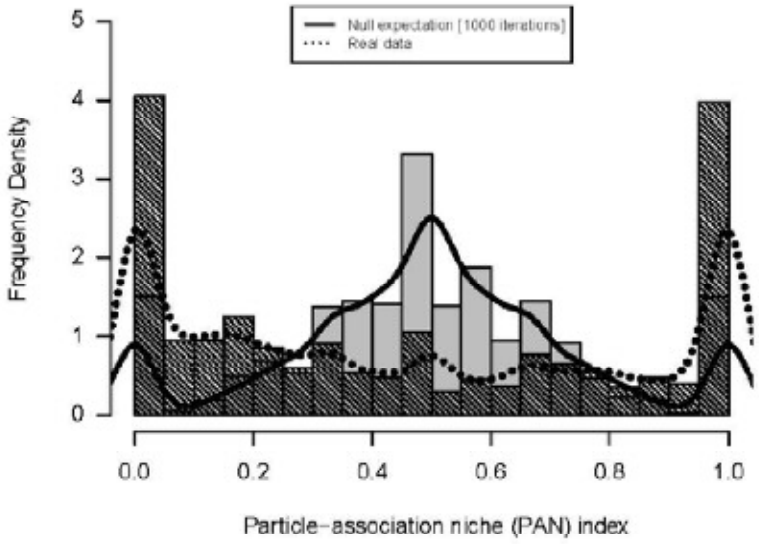
Figure 3: Phylogenetic signal of particle-associated niche index (PAN index). a) Mean PAN index differences between pairs of OTUs as a function of between-OTUs phylogenetic distance. Mean values are computed using 0.01 unit bins. The mean PAN index difference for all the OTUs is indicated with a horizontal line. b) Histograms of phylogenetic distances between OTUs belonging to the same Domain, Phylum, Class, Order, Family and Genus based on RDP fix-rank taxonomical annotation. Only taxonomic annotations with confidence values ≥90 were used. The phylogenetic distance scale in both panels is the same. Only the OTUs containing more than 10 reads were used for both panels.

Figure 4: Phylogenetic placement of PAN index. Phylogenetic tree representing the evolutionary history of all the OTUs with more than 10 reads. Abundance (log of number of reads) and PAN-index values are represented as bars for each OTU. Mean PAN index is additionally color-coded by a blue-red gradient for each of the main lineages (selected based on the SILVA-based taxonomical annotation). Indicator OTUs for both lifestyles are indicated using blue (free-living) and red (particle-attached) dots.

Figure 5: Boxplots of the particle-association niche index (PAN index) for a) the main Phyla and b) the main lineages within each Phylum. Only Phyla containing more than 40 OTUs are presented. Lineages within each Phylum were selected based on the SILVA-based taxonomical annotation. Boxplots with PAN-index values significantly different from 0.5 (P-value ≤ 0.05 using Wilcoxon signed rank test and after FDR correction, see *Material and Methods*) are in gray. The vertical broken line corresponds to a 0.5 PAN-index.



a



b