

DR. CLARA RUIZ-GONZALEZ (Orcid ID : 0000-0003-3568-4943)

MISS MIREIA MESTRE (Orcid ID : 0000-0003-0986-633X)

Article type : Original Article

Corresponding author mail id :- clara.ruiz.glez@gmail.com

Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean

Clara Ruiz-González¹, Ramiro Logares¹, Marta Sebastián^{1,2}, Mireia Mestre^{1,3}, Raquel Rodríguez-Martínez^{1,4}, Martí Galí⁵, Maria Montserrat Sala¹, Silvia G. Acinas¹, Carlos M. Duarte⁶ and Josep M. Gasol^{1,7}

(1) Institut de Ciències del Mar (ICM-CSIC), Barcelona, Catalonia, Spain

(2) Instituto de Oceanografía y Cambio Global, IOCG, Universidad de Las Palmas de Gran Canaria, ULPGC, Telde, Spain.

(3) Centro FONDAP de Investigación en Dinámica de Ecosistemas Marinos de Altas Latitudes (IDEAL), Universidad Austral de Chile, Valdivia, Chile

(4) Laboratorio de Complejidad Microbiana y Ecología Funcional, Instituto Antofagasta, Universidad de Antofagasta, Antofagasta, Chile

(5) Unité Mixte Internationale Takuvik (Université Laval-CNRS) and Québec-Océan, Département de Biologie, Université Laval, Québec, Québec, Canada G1V 0A6.

(6) King Abdullah University of Science and Technology (KAUST), Red Sea Research Center (RSRC), Thuwal, Saudi Arabia

(7) Centre for Marine Ecosystem Research, Edith Cowan University, Joondalup, WA, Australia

Running title: Bacterial dominance and rarity in the ocean

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.15026

This article is protected by copyright. All rights reserved.

Keywords:

Spatial abundance distribution; rare bacteria; marine prokaryotic communities; global surface ocean; Malaspina expedition; environmental transitions; dispersal

ABSTRACT

Microbial taxa range from being ubiquitous and abundant across space to extremely rare and endemic, depending on their ecophysiology and on different processes acting locally or regionally. However, little is known about how cosmopolitan or rare taxa combine to constitute communities and whether environmental variations promote changes in their relative abundances. Here we identified the Spatial Abundance Distribution (SpAD) of individual prokaryotic taxa (16S rDNA-defined Operational Taxonomic Units, OTUs) across 108 globally-distributed surface ocean stations. We grouped taxa based on their SpAD shape ('normal-like' - abundant and ubiquitous; 'logistic' - globally rare, present in few sites; and 'bimodal' - abundant only in certain oceanic regions), and investigated how the abundance of these three categories relates to environmental gradients. Most surface assemblages were numerically dominated by a few cosmopolitan 'normal-like' OTUs, yet there was a gradual shift towards assemblages dominated by 'logistic' taxa in specific areas with productivity and temperature differing most from the average conditions in the sampled stations. When we performed the SpAD categorization including additional habitats (deeper layers and suspended particles), the SpAD of many OTUs changed towards fewer 'normal-like' shapes, and OTUs categorized as globally rare in the surface ocean became abundant. This suggests that understanding the mechanisms behind microbial rarity and dominance requires expanding the context of study beyond local communities and single habitats. We show that marine bacterial communities comprise taxa displaying a continuum of SpADs, and that variations in their abundances can be linked to habitat transitions or barriers that delimit the distribution of community members.

INTRODUCTION

Microorganisms, comprising much of the living diversity across terrestrial and aquatic ecosystems, are the engines driving global biogeochemical cycles (Falkowski et al. 2008). A decade of biogeographic studies has established that the composition of microbial assemblages is determined by the influence of local (environmental conditions, ecological interactions) or regional factors (dispersal, Hanson et al. 2012, Martiny 2015), as well as by the different ecological strategies of taxa (Livermore et al. 2014; Salcher et al. 2014; Kirchman 2016). In the case of prokaryotes, this frequently results in a pattern wherein a few taxa are remarkably abundant and ubiquitous across space, while the majority remains rare and restricted to few sites (Nemergut et al. 2011; Salazar et al. 2016), even across highly heterogeneous habitats (Zinger et al. 2011; Ruiz-González et al. 2017). The reasons behind these patterns are not fully understood. Some studies have linked the ubiquity of taxa to their genomic content and metabolic versatility, showing that whereas the most ubiquitous bacteria in the surface oligotrophic ocean have simplified and streamlined genomes (Lauro et al. 2009, Swan et al. 2013), those from soils show large genomes and high metabolic versatility (Barberán et al. 2014). Also, the widespread distribution of certain prokaryotes has been associated with their capacity to persist in unfavorable conditions upon dispersal (i.e., as dormant or spores, as suggested by Hubert et al. 2009; Ruiz-González et al. 2015, 2017, Mestre et al. 2018). Still, however, little is known about how cosmopolitan or rare taxa assemble to constitute communities, and whether environmental transitions or dispersal barriers promote shifts in their abundances.

Given that closely related microbial taxa may respond differently to the same environmental drivers (Teeling et al. 2012; Jezberová et al. 2017), grouping them based on their spatio-temporal patterns may allow partitioning communities into groups of species that share ecological strategies or similar biogeography. For example, the exploration of the temporal dynamics of individual taxa allowed identifying specific strategies shared among taxonomically different bacteria, such as opportunists that bloom under favorable conditions (i.e., Shade et al. 2014; Alonso-Sáez et al. 2015; Lindh et al. 2015) or sequential responses to perturbations like soil rewetting or starvation (Aanderud et al. 2015; Sebastián et al. 2018). Conversely, knowledge of the spatial distribution of individual taxa

may provide insight into their persistence capacities and niche breadth, as well as into different dispersal-related processes governing their presence. This is particularly relevant in the case of prokaryotes, which are less affected by geographical barriers (e.g., Mayol et al. 2017; Reche et al. 2018; Mestre et al. 2018; Logares et al. 2018) than other organisms, and are able to persist out of their suitable niches for longer times until favorable conditions are encountered (Lennon and Jones 2011; Hervàs et al. 2009, Langenheder et al. 2016, Sebastián et al. 2018).

Delineating spatial patterns of individual prokaryotes thus requires considering the potential dispersal-driven connectivity between local communities, and a larger degree of environmental heterogeneity than for other organisms. For example, classifying individual bacterial taxa based on the shape of their spatial abundance distribution (SpADs) across *ca.* 200 lake communities, an approach that had been only applied to macroorganisms (Conlinsk et al. 2012), Niño-García et al. (2016a) were able to identify groups of bacteria displaying contrasting spatial patterns underlain by different assembly processes: A pool of ‘core’ abundant lake taxa structured by in-lake conditions, and a vast number of rare taxa passively dispersed by rivers into the lakes. This study exemplifies that the mechanism explaining rarity in lake bacteria (i.e. hydrologic dispersal of allochthonous taxa) can only be understood when considering the spatial distribution of bacteria beyond the lake habitat itself, i.e., in the associated river network. Moreover, it suggests that given the high dispersal and persistence of microbes, or understanding of microbial dominance and rarity might be largely influenced by the spatial context wherein we study them, an issue rarely contemplated in microbial biogeographic investigations.

Compared to freshwater and terrestrial ecosystems, defining dispersal routes, habitat transitions and geographic barriers of individual microbial taxa in the global ocean is less obvious, particularly in the relatively homogenous surface subtropical and tropical ocean. Despite the effort to identify oceanic provinces (Longhurst, 1998), surface planktonic bacterial communities do not clearly reflect these environmental divisions (Gibbons et al. 2013, Ladau et al. 2013, Sunagawa et al. 2015). In addition, compared to macroorganisms, our understanding of the environmental and spatial heterogeneity that individual microbes sense is much more limited (e.g. Stocker 2012). Exploring

local variations in the abundances of groups of taxa displaying different spatial distributions across the ocean (i.e. differing in their SpADs), may thus provide extra layers of information regarding taxa occurrence that cannot be derived from biogeographic patterns in taxonomic composition. For example, identifying whether taxonomic changes are caused by local increases in the contribution of cosmopolitan or endemic microbial taxa might allow delineating relevant ecotones or dispersal barriers in the ocean, as well as gaining insight into the mechanisms behind taxa persistence, ubiquity and rarity.

We thus hypothesize that bacterial communities are composed of taxa displaying a continuum of spatial distribution patterns, ranging from highly ubiquitous to rare and endemic, and that variations in their contribution to communities will reflect shifts in fundamental structuring processes linked to habitat transitions or dispersal barriers in the ocean (e.g. Baltar et al. 2016). To test this hypothesis, we characterized the SpAD-based structure of the prokaryotic communities across the global subtropical and tropical surface ocean. We i) identified the large-scale SpADs of 9,210 individual taxa across 108 surface free-living prokaryotic communities sampled from the Atlantic, Indian and Pacific oceans during the Malaspina 2010 Circumnavigation Expedition (Duarte 2015), ii) grouped these taxa based on their SpADs, and iii) explored whether the SpAD-based community structure varied along environmental gradients. We further assessed how the individual spatial patterns of taxa (i.e., their SpAD categorization) changed when including additional oceanic habitats (deeper ocean layers and particles of different sizes). Since these pattern-based categorizations should be conditioned by the nature of the spatial context considered, we expected that this information would allow delineating environmental breadths and persistence capacities of individual taxa, as well as inferring the spatial limits to their cosmopolitanism or rarity.

MATERIALS AND METHODS

Study area, sampling design and basic parameters

The Malaspina 2010 circumnavigation expedition (Duarte 2015) was conducted between December 2010 and July 2011 on board R/V Hespérides through the tropical and subtropical global ocean. A total of 108 stations located along the tropical and subtropical Atlantic, Indian, and Pacific oceans (Fig. 1) were sampled. At each station, surface water (3 m) was sampled with a 20 L Niskin bottle. Additionally, in eight of these stations (see asterisks in Fig. 1) water from three other layers (Deep Chlorophyll Maximum (DCM, 48-150 m), the mesopelagic (250-670 m) and the bathypelagic (3105-4000 m)) was sampled with a rosette of Niskin bottles mounted on a conductivity–temperature–depth (CTD) profiler (Seabird SBE 911). Water was pre-filtered through a 200 μm to remove large plankton. Salinity, temperature and dissolved oxygen were recorded at each station with the CTD sensors installed in the rosette sampler. Concentration of nutrients (nitrate, phosphate, and silicate), as well as estimations of the mixed layer depth (MLD) and daily and monthly solar radiation (SWR), were determined as explained in Catalá et al. (2016).

Biological parameters

Prokaryotic abundances, cell sizes, and the number of cells with high nucleic acid content (HNA) were determined by flow cytometry as described in Gasol and del Giorgio (2000). Viral abundances were estimated by flow cytometry as explained in Lara et al. (2017). Prokaryotic heterotrophic production was estimated using the ^3H -leucine incorporation method (Kirchman et al. 1985) as described in Morán et al. (2017). Chlorophyll *a* (Chl) concentration was determined as detailed in Estrada et al. (2016).

Satellite-derived chlorophyll concentration and sea surface temperature

Satellite-retrieved chlorophyll concentration (Chl) and sea surface temperature (SST) were obtained from the Moderate resolution Imaging Spectrometer (MODIS) onboard the Aqua satellite. Eight-day global-ocean composites at 9.28 km resolution (level 3-mapped data, 2014.0 reprocessing) were downloaded from the NASA Ocean Color Web (<https://oceancolor.gsfc.nasa.gov/>). For each station and variable, an annual time series (46 eight-day periods) ending on the sampling date was extracted. To ensure good coverage and avoid data gaps due to clouds, each temporal point was represented by the average of a square of 10x10 pixels (93x93 km²) centered on the station coordinates. We then extracted the mean over different periods at each station: the time of sampling (8-day period), and the month and the year prior sampling. Finally, the departure of a given station from prevailing conditions in the global dataset was calculated as the ratio between the local mean Chl (or SST) and the global median Chl (or SST) in the entire dataset.

Prokaryotic community composition

At each of the 108 surface stations, 6-15 L of seawater were sequentially filtered through a 20 µm mesh, a 47-mm polycarbonate membrane filter of 3 µm (Merck Millipore, Darmstadt, Germany, Isopore polycarbonate) and 0.2 µm (Merck Millipore, Express Plus) pore size with a peristaltic pump. We focused on the 0.2–3 µm fraction, which represents mostly free-living bacteria. In addition, at 8 of these stations (asterisks in Fig. 1), the prokaryotic communities inhabiting deeper layers and larger size-fractions were also characterized by sequentially filtering 10 L of water from the rosette through 20, 5.0, 3.0, 0.8 and 0.2 µm 47-mm pore-size filters (20 µm pore-size filter from GE Water and Process Technologies (Trevose, PA, USA) and the rest from Millipore (Billerica, MA, USA)), resulting in size fractions that represent free-living prokaryotes (0.2-0.8 µm), as well as various particle-attached fractions (0.8-3; 3-5; 5-20 and 20-200 µm, see Mestre et al. (2018) for details on this particular dataset). These different fractions were collected from the surface but also from the DCM, mesopelagic, and bathypelagic waters (see above).

Genomic DNA was extracted using the standard phenol-chloroform protocol with slight modifications (Salazar et al. 2016). The V4-V5 region of the 16S rRNA gene was amplified with the primers 515F and 926R (Parada et al. 2016) and sequenced in an Illumina MiSeq platform using 2x250bp paired-end approach at the Research and Testing Laboratory facility (Lubbock, Texas, USA; <http://www.researchandtesting.com>). Amplicons were processed through a custom-made pipeline (Logares, 2017) based on UPARSE (Edgar 2013). Briefly, reads were merged with PEAR (Zhang et al. 2014), and sequences >200 bp were used for further processing. Quality check, dereplication, OTU clustering (99%), and reference-based chimera filtering (using SILVA v.119) were processed with USEARCH (Edgar 2010). Taxonomic assignment of OTU representative sequences was based on BLASTing against SILVA v123. Chloroplast and mitochondrial sequences were removed. To allow for comparisons between samples, the OTU table of the 108 free-living surface communities was randomly subsampled to the minimum number of reads/sample (n=10,904). An additional OTU table (rarefied at 5,390 reads/sample) was constructed with the eight stations where four depths and five size-fractions had been characterized. Finally, these two datasets were combined in order to generate three additional OTU tables that were used to assess the effect of increasing environmental heterogeneity (see below), which were rarefied at 7,061 reads/sample. Raw sequence data have been deposited in the European Nucleotide Archive (acc. num. PRJEB25224; ERP109198, ERS2539749–ERS2539903). Computing analyses were run at the MARBITS bioinformatics platform at the Institut de Ciències del Mar (<http://marbits.icm.csic.es>; ICM, Barcelona, Spain).

Identifying individual prokaryotic spatial abundance distributions

The analysis of the individual spatial abundance distributions (SpADs) was done as described in Niño-García et al. (2016a) using the rarefied OTU table from the 108 surface stations. Briefly, the procedure chooses the statistical distribution that best fits the SpADs of each OTU across the 108 marine stations: First, all OTU abundance distributions ($\log_{10}(x + 1)$ transformed) that deviate from unimodality are identified by computing the Hartigans' dip test statistic (*dip test* R package, Maechler

2015): Any OTU with a P-value < 0.01 is categorized as having a bimodal distribution. Then, several potential model distributions are fit for unimodal OTUs (normal, Weibull, gamma, Cauchy, lognormal, logistic, and exponential) using maximum likelihood estimation with the function `distfit` (*fitdistrplus* R package, Delignette-Muller and Dutang 2014), which also provides the Akaike Information Criterion (AIC) used to select the best model for each OTU (i.e. the model showing the lowest AIC, see Niño-García et al. (2016a) for details). In total, 7 different types of model distributions fitted the different individual SpAD of marine OTUs. Among OTUs having unimodal distributions, we could distinguish between those with more symmetrical distributions with various degrees of dispersion (normal, Weibull, Cauchy), which were generally abundant and ubiquitous, and those with positively skewed distributions with various degrees of asymmetry with the mode corresponding to zero abundance (absences). We thus grouped OTUs into 3 SpAD categories: 1) OTUs with ‘normal-like’ spatial abundance distributions, which comprised those OTUs presenting normal, Weibull, gamma or Cauchy statistical distributions; 2) bimodal OTUs, i.e., those whose distribution had two clear density peaks and 3) logistic OTUs, which comprised both logistic and lognormal statistical distributions, characterized by a zero abundance mode generated by many absences. Even though these two categories of rare taxa (logistic and lognormal) had previously been shown to display somewhat different dynamics across boreal lakes (Niño-García et al. 2016a), here we observed that their abundances largely covaried throughout the surface ocean (Suppl. Fig. 1) and may thus represent a similar distribution. Therefore, for all subsequent analyses, logistic and lognormal OTUs were considered together and will be hereafter referred to as ‘logistic’ OTUs.

We decided to run the SpAD analysis on the rarefied OTU table to avoid sampling biases associated to very different sequencing depths across our samples (range 8000 to 147,700 reads), yet we acknowledge that by rarefying we may be losing many rare taxa, probably underestimating the total number of ‘logistic’ OTUs. Since most of the logistic OTUs (95%) remained permanently rare (see Results), we believe that losing some rare taxa does not invalidate our results, because it is highly likely that in terms of sequences they do not represent a high fraction of the community.

Effect of habitat heterogeneity on the SpAD categorization of taxa

To test how the SpAD categorization of an OTU changes depending on the environmental context where it is studied, we repeated the SpAD analysis using three datasets of increasing habitat (and spatial) heterogeneity, which were constructed by combining the 108 surface free-living communities with different subsets of communities sampled at the eight selected stations (see above, Fig. 1). The first dataset included 24 additional surface particle-attached communities associated with the three largest size-fractions (3-200 μm), whereas the second dataset incorporated 29 free-living communities (0.2-0.8 μm) sampled from the surface, DCM, mesopelagic and bathypelagic waters. The third and most heterogeneous dataset included, besides all these, all other particle-attached fractions collected from the four depths, resulting in 258 samples in total.

Statistical analyses

Communities were clustered using non-metric multidimensional scaling (NMDS) analysis based on Bray-Curtis distances. The k-means classification algorithm using the best ‘simple structure index (‘ssi’) value (i.e., a diagnostic measure to determine the best partition) was run as implemented in the vegan R package. The Vegan envfit function (Oksanen et al., 2015) was used to fit environmental vectors onto the ordination space. All analyses were performed with R 3.0.0 software (R Development Core Team, 2013).

RESULTS

Sequencing of the 16S rRNA gene from the 108 surface stations resulted in 1,177,632 quality-checked sequences after rarefaction, which clustered into 9,210 OTUs (at 99% identity). The OTU number per sample ranged between 578 and 1,383 (mean \pm SD = 1,113 \pm 145), and the Shannon Index of OTU diversity varied between 2.5 and 5.2 (mean \pm SD = 4.2 \pm 0.6). Altogether, 23 bacterial phyla were detected, yet the total sequences were mainly dominated by Cyanobacteria (49% of the

total number of sequences), followed by Proteobacteria (35%), Bacteroidetes (8%), and Actinobacteria (5%). The global OTU accumulation curve approached a plateau, suggesting that we obtained good but not complete coverage of the surface ocean prokaryotic richness (Suppl. Fig. 2). Overall, free-living prokaryotic communities did not show any clear segregation depending on the oceanic region, province, or sampling transect (Suppl. Fig. 3a,b), and surface water temperature, dissolved oxygen (Suppl. Fig. 3c,d, envfit analysis $p < 0.001$) and salinity (not shown) were the only factors showing a significant association to the observed distribution of assemblages.

Distribution of SpAD categories throughout the surface ocean

All analysed OTU spatial-distributions could be assigned to one of three basic model distributions: (1) 'normal-like'; (2) 'bimodal' and (3) 'logistic' (see examples of distribution shapes in Fig. 1b–d). Briefly, the normal-like category harbored OTUs showing a unimodal symmetrical or slightly positively skewed shape, varying around a central value that was always $> 0.1\%$ of the local abundance, and thus represented regionally abundant and ubiquitous OTUs (Fig. 1b, Fig. 2a,b). These OTUs might comprise habitat generalists, as they showed the broadest environmental tolerances (Suppl. Fig. 4). The bimodal category included OTUs showing two density peaks, one of which corresponded to zero in most cases (Fig. 1c), resulting in OTUs with lower average abundances and occurrences (Fig. 2a,b), and which could represent habitat specialists (i.e., taxa with more restricted environmental tolerances, Suppl. Fig. 4). Finally, the SpAD of the rarest OTUs followed either a logistic or a lognormal distribution, showing a characteristic zero abundance mode due to many absences (Fig. 1d), and included OTUs that are globally rare and restricted to few sites (Fig. 2a,b).

Only 68 OTUs (0.7% of total OTUs) displayed a 'normal-like' pattern, yet in terms of their abundances they represented the dominant SpAD category, accounting for 55% of the total sequences (Fig. 2c). 1,788 OTUs (19%) were categorized as bimodal and made up 34% of total sequences. Finally, the logistic category comprised most of the OTUs identified (80%) but represented only 12% of the sequences (Fig. 2c). There were no obvious changes in terms of taxonomic composition

between the three SpAD categories, but in general most normal-like OTUs were associated to *Prochlorococcus* and SAR11, whereas other groups such as *Synechococcus*, Flavobacteriia, Oceanospirillales, Rhodospirillales or Alteromonadales showed higher contributions among bimodal or logistic OTUs (Suppl. Fig. 5).

The analysis of the distribution of the SPADs across the 108 surface stations unveiled that most communities were numerically dominated by normal-like OTUs (Fig. 2d), followed by bimodal OTUs and finally by the logistic category of globally rare taxa. However, there were deviations from this overall structure in several stations where either bimodal or logistic OTUs showed marked increases in their relative abundance with respect to other communities. In most cases these deviations were concentrated in particular regions. For example, several contiguous stations in the Equatorial Pacific (leg 5), South Atlantic (leg 2) or Indian Ocean (leg 4) showed high proportions of logistic or bimodal OTUs, and changes in these proportions varied gradually along the ocean transects (Fig. 2d).

Links between the SpAD-based community structure and *in situ* environmental gradients

To explore whether the observed deviations in SpAD-based community structure (Fig. 2d) could be associated to variations in environmental or biological factors, we classified prokaryotic communities based on the contribution of the SpAD categories. Using a standard k-means classification algorithm and the 'ssi' criterion to determine the best partition, we grouped communities into four clusters. These differed in the relative abundances of the different SpAD categories, which were named after their dominant SpAD category (NL: normal-like, BM: bimodal, LG: logistic). Most communities (77%) fell into clusters NL1 (n=54) and NL2 (n=29), and represented assemblages numerically dominated by normal-like OTUs (avg. 67% and 52% of community sequences, respectively, Figs 3a-c), although NL2 had higher contributions of bimodal and logistic categories than NL1. Cluster BM comprised 11 communities showing the highest proportions of bimodal OTUs, and Cluster LG represented 14 communities characterized by the highest contribution of logistic OTUs to total community sequences (Figs. 3a-c).

We observed that, except for a few cases, communities from clusters BM and LG appeared located together in certain areas (Fig. 3d), such as the Equatorial Pacific, the South Australian Bight and the Benguela Coastal province in the South Atlantic. In each of these three areas, we observed gradual increases in the proportion of logistic OTUs towards colder temperatures, whereas that of normal-like OTUs decreased (Fig. 4 a,c). Bimodal OTUs often showed the maximum contribution somewhere in between the temperature gradient (Fig. 4b). This resulted in a gradual change in the SpAD-based community structure along the different temperature gradients of each of these three regions, which shifted from normal-like (NL) dominated communities towards those with the highest proportions of logistic OTUs (LG stations). When all stations were pooled together, however, this temperature-driven pattern was no longer apparent, as LG communities occupied both the warmest (7 Equatorial Pacific sites, 'LGpac' stations in Table1) and coldest waters (rest of LG stations). In fact, no *in situ* variable appeared to explain the gradual changes in the relative abundance of SpAD categories from NL1 to LG (Table 1). Remarkably, though, all LG communities consistently appeared in the most productive sites, showing significantly higher chlorophyll concentrations, % of high nucleic acid content (HNA) bacteria, and abundances of viruses (Table 1). Interestingly, the increase in relative abundance of logistic taxa was due to just a few OTUs (7 to 19) that became abundant (i.e., surpassed at least once the 1% of local relative abundance, Suppl. Fig. 6), rather than an overall increase in the number of logistic OTUs. Among these specialized 'bloomers', some were associated to Flavobacteriia, Rhodobacterales, Oceanospirillales and presumably different ecotypes of picocyanobacteria, but their composition differed between the three geographic areas (Suppl. Fig. 6). Conversely, most logistic OTUs remained permanently rare across all surface stations.

Changes in SpAD-based community structure due to past environmental conditions

Given that the different SpAD categories may reflect to some extent differences in the growth rate of taxa (e.g., normal-like could be related to oligotrophic taxa, logistic to copiotrophic bloomers), it is possible that the contemporary community composition results from a mixture of fast and slow

Accepted Article

responses of taxa to environmental conditions, remnants of past proliferations, and even allochthonous inactive taxa. Hence, the *in situ* environmental conditions at the time of sampling may not reflect the conditions affecting all the different community members. We thus explored whether changes in SpAD-based structure would be more clearly explained if the past environmental history of communities were considered, for which we used satellite derived surface seawater temperature (SST) and chlorophyll concentration data. For each station, we calculated the average SST and chlorophyll during i) eight days, ii) one month prior to the sampling date, and iii) during the whole year prior to sampling (see Methods) and compared these between the four clusters (Fig. 5). Interestingly, only chlorophyll concentrations averaged during the whole year prior to sampling (Fig. 5c) showed a gradual increase from NL1 to LG stations, whereas this was not apparent with the *in situ* chlorophyll values (Table 1) nor the values averaged over shorter periods (8 days or 1 month, Fig. 5a,b). Changes in SST across clusters over these different time spans were similar with changes in *in situ* water temperature (Fig. 5a-c, Table 1), but did not reveal a progressive change from NL1 to LG assemblages because communities within the LG cluster were exposed to largely contrasting temperatures.

Interestingly, however, we observed that stations classified as BM and LG consistently comprised environments less represented in our dataset; when we compared the local mean of satellite-derived chlorophyll and SST with the median of those same values across the whole dataset (horizontal lines in Fig. 5), we found that in both cases BM and LG communities differed more from the global median than those from clusters NL1 and NL2. These deviations involved highly productive sites and temperature extremes relative to the global median (Fig. 5), implying that the inclusion of such stations in the dataset resulted in the addition of habitats that were distinct enough to cause drastic shifts in the SpAD-based community structure, increasing the proportion of less cosmopolitan (bimodal and logistic) OTUs.

Increasing habitat heterogeneity leads to changes in SpAD-based community structure

The above observations suggest that covering higher habitat diversity may influence the categorization of an OTU as normal-like, bimodal or logistic (and thus our perception of their distribution). To explore so, we repeated the SpAD analysis increasing the microbial niches covered. We constructed three additional datasets that included, besides the 108 free-living surface communities: i) communities attached to three different size particles (3-5; 5-20 and 20-200 μm , see Methods) from surface waters ($n=24$), to test the effect of increasing habitat heterogeneity without expanding the spatial scale, given that these other size fractions were sampled from the same stations, ii) free-living communities from three other layers (DCM, meso- and bathypelagic, $n=29$) which increased both habitat and spatial coverage (i.e., vertical distance) and finally iii) communities associated with five particle size-fractions collected at the four layers ($n=150$, see Methods), which included a large diversity of microbial niches.

Interestingly, including just 24 additional surface communities associated with larger size-fractions (3-200 μm) already resulted in a fundamental change in the outcome of the SpAD analysis, with the number of OTUs categorized as normal-like decreasing from 68 to 19 and the number of logistic OTUs increasing (Fig. 6a,b). Including a comparable number (29) of free-living communities from deeper layers (Fig. 6c) resulted in a similar decrease in the number of normal-like OTUs (68 to 19) but in a more dramatic change in the average contribution of the different categories, with bimodal and logistic OTUs increasing up to 57% and 24% of the sequences, respectively (Fig. 6c). Finally, when all particle-attached communities from all depth layers were added (150 additional samples), only 2 OTUs remained categorized as normal-like and most sequences belonged to bimodal OTUs (69%, Fig. 6d). The taxonomic identity of the OTUs that remained categorized as 'normal-like' after each step is shown in Suppl. Table S1.

We observed that 90% of the sequences detected across deeper layers and larger size-fractions belonged to OTUs also found in the 108 surface free-living communities (pie chart in Suppl. Fig. 7), suggesting that the above-mentioned changes in OTU categorization were mostly due to shifts in the

SpAD of taxa shared between the two datasets, not to the appearance of new taxa. In order to check this, we explored how the contribution of the SpAD categories defined based on the surface free-living dataset changed across deeper oceanic layers and particles of varying size classes, using the samples from the 8 stations where all size classes were available. This analysis differs from the previous one (Fig. 6) in which we did not re-run the SpAD analysis, but rather traced the abundances of the original surface-derived SpAD categories (Fig. 2) across these other habitats. We observed that the dominance of surface normal-like OTUs was mostly restricted to the free-living fraction and surface waters, whereas the relative abundance of surface bimodal and logistic OTUs largely increased in communities attached to larger particles and in deeper oceanic layers (Suppl. Fig. 7). However, most rare OTUs remained permanently rare, so that this increase in their abundances was again due to just a few OTUs (87 out of 1,736 in the case of bimodal, and 186 out of 4,963 in the case of logistic OTUs) that became abundant (i.e. >1% of the local sequences) at least once in deeper waters or in larger size fractions.

DISCUSSION

Our analysis of the spatial abundance distributions of 9,210 surface ocean prokaryotic OTUs allowed partitioning microbial communities into three groups of taxa showing different global distributions independently of their taxonomy. More importantly, this approach unveiled variations in the proportions of the different groups across local communities. Although many studies have focused on the biogeography of rare and abundant prokaryotes (Galand et al. 2009; Logares et al. 2013; Vergin et al. 2013; Liu et al. 2015; Ruiz-González et al. 2017), to our knowledge this is the first study exploring how taxa with different spatial distributions assemble to constitute marine prokaryotic assemblages, as well as the potential causes promoting variations in the resulting pattern-based community structure.

Accepted Article

Compared to macroorganisms, delineating spatial distribution areas in microorganisms is much more challenging, as well as inferring life strategies (e.g., habitat generalist/specialist, copiotrophs/oligotrophs, Newton and Shade 2016) from those spatial patterns. First, how microbes perceive their environment at the microscale remains poorly understood (Stocker 2012), and we do not even have an accurate estimate of the total microbial diversity present in a sample (Locey and Lennon, 2016). Second, prokaryotes can be dispersed much longer distances and can persist out of their suitable niches (dormant or inactive but capable of resuming growth, Lennon and Jones 2011) for longer times than larger organisms (Langenheder et al. 2016, Sebastián et al. 2018). This, coupled to the fact that our perception of taxa occurrence is largely dependent on the sequencing resolution (Gibbons et al. 2013; Ruiz-González et al. 2017), explains why our understanding of the ecology and mechanisms associated with microbial cosmopolitanism, dominance or rarity is far behind that of macroorganisms. In this context, the identification of bacterial spatial ‘behaviors’ across large spatial scales is emerging as a promising way to gain insight into microbial biogeography drivers (Mestre et al. 2018, Ruiz-González et al. 2015, 2017, Niño-García et al. 2016a). For example, the SpAD approach allowed discovering that hydrologic transport from rivers explained the presence of most rare bacteria in boreal lakes, but at the same time inoculated taxa able to thrive in lake conditions (Niño-García et al., 2016a). Still, however, the origin and the causes underlying the huge reservoir of rare diversity and its role in the ocean are poorly understood (Logares et al. 2015, Ser-Giacomi et al. 2018) and we are only starting to comprehend how processes such as dispersal may shape the biogeography of marine microorganisms (Wilkins et al. 2013; Mestre et al. 2018).

Here we show that most (77%) of the studied free-living bacterial assemblages in the subtropical and tropical surface ocean are numerically dominated by a few abundant and cosmopolitan OTUs (68 normal-like OTUs), followed by bimodal OTUs of intermediate abundance and occurrences, and finally by a large number of rare (logistic) OTUs present at a few stations (Fig. 1, 2). This partition agrees with previous observations based on physiological and genomic approaches showing that free-living marine bacterial communities are dominated by a few cosmopolitan oligotrophs, different

copiotrophs able to persist and bloom upon specific conditions (Lauro et al. 2009; Yooseph et al. 2010, Swan et al. 2013, Kirchman 2016), and a large pool of highly diverse rare taxa (Pedrós-Alió et al. 2012). Accordingly, most normal-like sequences belonged to *Prochlorococcus* and SAR11 (Suppl. Fig. 5), oligotrophs with simplified and streamlined genomes that dominate communities in the relatively homogeneous tropical and subtropical oligotrophic ocean (Morris et al. 2002; Partensky and Garczarek, 2010; Giovannoni 2017). Typical copiotrophic groups such as Flavobacteriia, Oceanospirillales, Rhodospirillales or Alteromonadales (e.g. Teeling et al., 2012, Landa et al. 2016) showed higher presence among bimodal or logistic OTUs (Suppl. Fig. 5). The observation that only ~1 % of the OTUs fitted the normal-like distribution also agrees with the fact that most abundant microbial taxa are usually a small fraction of the total diversity being also the most widely distributed (Nemergut et al., 2011; Martiny, 2015; Salazar et al., 2016, Niño-García et al. 2016a).

The distribution of SpAD categories was quite uniform across communities from the global tropical and subtropical surface ocean (Fig. 2), yet there were some clear deviations in which either bimodal or logistic OTUs increased their abundances. Most of these BM (dominated by bimodal) or LG (dominated by logistic) communities were localized in three areas potentially subjected to coastal (South Australian Bight, Benguela region in the South Atlantic) or Equatorial upwelling (Equatorial Pacific), which generally showed higher chlorophyll concentrations and enhanced fluxes of nitrate towards the euphotic layer (Estrada et al. 2016). Remarkably, regardless of the temperature range within each of these regions, in all three cases communities shifted gradually from a dominance of normal-like OTUs (NL1 stations) to a dominance of logistic OTUs (LG stations) along a gradient of decreasing temperature (Fig. 4). Such regional shifts in SPAD-based community structure may reflect fundamental niche transitions for microbes along these temperature gradients, likely linked to the influence of colder nutrient-rich upwelled waters. Moreover, abrupt transitions in community structure were also detected, like one LG community in the Costa Rica dome (Fig. 3), an upwelling zone caused by a persistent cyclonic eddy (Fiedler 2002), or that in front of Rio de Janeiro (Brazil), which coincided with a change in water masses due to salt finger intrusions (Fernández-Castro et al. 2015). This agrees with previous smaller-scale studies showing that oceanic fronts and mesoscale

features delimit the distribution of bacterioplankton in the ocean (Baltar et al. 2010, 2016, Flaviani et al. 2018). Our results further suggest that the surface tropical and subtropical global ocean represents a relatively homogeneous habitat that results in similar SpAD-based community structures featuring small pools of dominant taxa with cosmopolitan distributions.

The correlation between temperature and the abundances of the different SpAD categories disappeared when all communities were pooled together, and the only variable explaining a sequential change in SpAD-based community structure (from NL1 to LG) was the satellite-derived chlorophyll concentration averaged during the year prior to sampling, which gradually increased from NL1 (dominated by normal-like) to LG (dominated by logistic) stations (Fig. 5c). This suggests that the SpAD-based structure reflects to some extent the environmental history, or average conditions, experienced by communities. Experimental and field studies have shown that environmental conditions occurring several days before sampling can be better predictors of spatial variation between communities than the present environment, suggesting that current taxonomic composition reflects selection by both present and past conditions because not all community components react equally to environmental changes (Berga et al. 2012, Andersson et al. 2014). However, the fact that only the satellite-derived chlorophyll values averaged over a year (in contrast to shorter time periods) explained the gradual shift in SpAD structure (Fig. 5), might indicate that the variations in annual chlorophyll between clusters (Fig. 5c) represent the average niche availability that those communities experience. Hence, higher annual chlorophyll values might reflect a greater diversity of microbial niches (e.g., different phytoplankton blooms along the year) that allow the coexistence and/or the sequential establishment of specialist taxa unable to thrive in more oligotrophic conditions, some of which may be currently present as remnants of past proliferations (Comte et al. 2017). Interestingly, only a few logistic OTUs (7 to 19 OTUs per area, Supp. Fig.6) became abundant in these more productive sites, reflecting local selection of taxa rather than massive immigration events (i.e., mass effects, Niño-García et al. 2016b), even though upward transport of deeper ocean communities can happen in some of these areas (Estrada et al. 2016). Along these lines, the differences between the two LG sub-clusters are also illustrative. The Equatorial Pacific sub-cluster (LGpac) featured lower Chl

concentration, over the year, than the remaining LG stations (Fig. 5c) . However, LGpac had the deepest mixed layer (Table 1), meaning easier entrainment of deep-water bacteria to the surface.

These differences suggest that different processes may simultaneously contribute to higher proportions of globally rare OTUs.

These observations might indicate that among logistic OTUs there are a few that are globally rare, specialist opportunistic taxa. However, the spatial categorization of OTUs as habitat generalists and specialists, or ubiquitous and rare, is not necessarily intrinsic to the OTU, but to its prevalence within the environmental context considered. In fact, higher proportions of bimodal and logistic OTUs (BM and LG stations) coincided with stations that presented temperature and chlorophyll regimes deviating furthest from the global median value (i.e., less represented) in the dataset (Fig. 5). This agrees with previous findings showing higher numbers of specialist taxa at environmental extremes (Logares et al. 2013), yet raises the question of whether these habitat specialists would be regarded as cosmopolitan (normal-like) if e.g. only upwelling regions had been sampled. The fact that the identity of these bloomers differed between the three productive regions (Suppl. Fig. 6), though, suggests that these areas represent contrasting and spatially fragmented habitats that select for different globally rare taxa (e.g., Landa et al. 2016), and thus that they comprise hotspots of specialization or endemism in the global surface ocean.

The proportion of normal-like OTUs dramatically decreased when the SpAD analysis was conducted on datasets covering higher habitat and spatial heterogeneity (Fig. 6). This suggests that regardless the spatial coverage, homogeneous environments will be dominated by normal-like OTUs. To further confirm the homogeneous nature of the surface dataset we rerun the SpAD analysis on 10 random subsamples of 50 surface stations each, and consistently found that surface communities were dominated by normal-like OTUs regardless of the subset of stations considered (see Suppl. Fig. 8). Our observation of higher proportion of bimodal and logistic OTUs with increasing habitat (i.e., niche) heterogeneity is in agreement with the dominance of bimodal OTUs found by Niño-García et al. (2016a) across the much more heterogeneous boreal lakes. Interestingly, the inclusion of larger size fractions from the surface resulted in a less dramatic change in the distribution of the different

SpAD groups (Fig. 6b) than the inclusion of free-living communities from deeper waters (Fig. 6c). This might imply that, despite the well-known differences in taxonomic composition between free-living and particle-attached communities (Acinas et al 1999; Crump et al, 1999; Salazar et al. 2015, Mestre et al. 2017, 2018) and across depths (Zinger et al, 2011, Sunagawa et al. 2015, Walsh et al. 2015), the transition from suspended to attached is a weaker barrier than moving along the water column as a free-living cell, where besides a much stronger environmental filtering, dispersal limitation likely plays a role (Mestre et al. 2018). The environmental context of individual studies thus constrains the apparent spatial patterns, environmental ranges, and even the inferred life strategies of the studied taxa (e.g., Newton and Shade 2016). Indeed, we observed that OTUs originally categorized as bimodal or logistic based on their presence in surface free-living communities, largely increased their relative abundances at other depth layers and size-fractions (Suppl. Fig. 7). This indicates that globally rare surface taxa may become abundant in these other habitats, in agreement with the recently reported strong connectivity between surface and deep microbial communities via sinking particles or water masses movement (Mestre et al. 2018, Wilkins et al. 2013, Luna et al. 2016). Exploring taxa distributions across different environments may thus help determine how differently environmental transitions or boundaries affect the distribution of each taxon.

It is notable that two OTUs were still categorized as normal-like even when all depths and size fractions were included (Fig. 6d), suggesting a remarkable versatility of a few taxa within communities (Suppl. Table 1). Alternatively, the cosmopolitanism shown by some community members might be the result of hidden diversity that cannot be resolved with the 16S rRNA marker, since identical 16S rRNA sequences can represent different ecotypes (e.g., Hahn et al. 2016; Farrant et al. 2016). Indeed, 22 out of our 68 surface normal-like OTUs matched more than one publicly available genome at $\geq 99\%$ similarity level (Suppl. Fig. 9), although it remains unknown whether these cosmopolitan OTUs represented one or several genomes. Using single-cell genomics, Kashtan et al. (2014) showed that the globally abundant *Prochlorococcus* comprises hundreds of coexisting subpopulations adapted to different micro-habitats, and proposed that this may be a general feature of abundant taxa. Whereas resolving these adaptations to micro-habitats would require using full

genomes, which are currently unavailable for most prokaryotes, amplicon-based studies provide useful information about biogeographic patterns at the species or genus levels. In our study, for example, the predominance of normal-like OTUs indicates that single entities or evolutionarily closely related species dominate the tropical and subtropical surface ocean.

In summary, we show that marine bacterial communities are composed of taxa displaying a continuum of spatial distribution patterns, and that variations in their contribution to communities can be linked to transitions between microbial habitats, hotspots of specialization or endemism or spatial or environmental barriers that differentially delimit the distribution of community members. This may have implications for our understanding of microbial community functioning and resilience, since the proportion of generalist versus specialist species seems to influence the productivity and stability of communities (Gravel et al. 2011; Matías et al. 2013) and their susceptibility to environmental alterations (e.g. Clavel et al. 2011). Finally, our results show that our estimates of taxa distributions can be largely influenced by the degree of environmental or habitat heterogeneity covered by the studied spatial context.

ACKNOWLEDGEMENTS

This work was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) through the Consolider-Ingenio program (Malaspina 2010 Expedition, ref. CSD2008-00077), with contributions from grant CTM2015-70340R, CTM2015-65720-R, CTM2015-69936-P and King Abdullah University of Science and Technology (KAUST). We thank all scientists and crew involved in the Malaspina expedition, particularly M. Pernice, G. Salazar, F.M. Cornejo, E. Borrull, C. Diez-Vives, E. Lara, M. and D. Vaqué for help with DNA collection and extractions, and X. A. Álvarez-Salgado for providing environmental data. This is a contribution of Grup Consolidat de Recerca of the Catalan Government 2014SGR/1179. CRG was supported by a Juan de la Cierva (IJCI-2015-23505, MINECO, Spain) fellowship and RL by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). MS was supported by grant EcoRARE (CTM2014-60467-JIN), funded by the Spanish

This article is protected by copyright. All rights reserved.

Government and the European Regional Development *Fund* (ERDF), REMEI (CTM2015-70340-Rf) funded by the Spanish Government, and a Viera y Clavijo contract funded by the ACIISI and the ULPGC. MM was supported by CONICYT (FONDAP-IDEAL 15150003). SGA was funded by BIOSENSOMICS through the ‘Convocatoria 2015 de ayudas Fundación BBVA a investigadores y creadores culturales’. RRM is supported by CONICYT FONDECYT 11170748.

Author Contributions

J.M.G., C.M.D., M.M.S. and S.G.A participated in the design of the expedition and designed the sampling scheme. R.R., M.S., M.M., J.M.G. prepared the samples on board or extracted the DNA. R.L. analyzed genomic sequences. M.G. provided and processed satellite data. C.R.G. conceived this work, analyzed the data and wrote this article. All authors made comments and relevant editions to the text.

Data Accessibility

- DNA sequences and associated metadata: ENA (European Nucleotide Archive) accession numbers PRJEB25224; ERP109198; ERS2539749–ERS2539903

REFERENCES

- Aanderud ZT, Jones SE, Fierer N & Lennon JF (2015) Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. *Frontiers in Microbiology* **6**: 24.
- Acinas SG & Rodriguez-Valera AJ (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.
- Alonso-Sáez L, Díaz-Pérez L & Morán XA (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environmental Microbiology* **17**: 3766–3780.

- Andersson MGI, Berga M, Lindstrom ES & Langenheder S (2014) The spatial structure of bacterial communities is influenced by historical environmental conditions. *Ecology* **95**: 1134-1140.
- Baltar F, Arístegui J, Gasol JM, Lekunberri I & Herndl GJ (2010) Mesoscale eddies: hotspots of prokaryotic activity and differential community structure in the ocean. *The ISME Journal* **4**: 975-988.
- Baltar F, Currie K, Stuck E, Roosa S & Morales SE (2016) Oceanic fronts: transition zones for bacterioplankton community composition. *Environmental Microbiology Reports* **8**: 132-138.
- Barberán A, Ramirez KS, Left JW, Bradford MA, Wall DH & Fierer N (2014) Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters* **17**: 794-802.
- Berga M, Szekely AJ & Langenheder S (2012) Effects of disturbance intensity and frequency on bacterial community composition and function. *PLoS One* **7**: e36959.
- Caporaso JG, Paszkiewicz K, Field D, Knight R & Gilbert JA (2012) The Western English Channel contains a persistent microbial seed bank. *The ISME Journal* **6**: 1089-1093.
- Catalá TS, Alvarez-Salgado XA, Otero J, *et al.* (2016) Drivers of fluorescent dissolved organic matter in the global epipelagic ocean. *Limnology and Oceanography* **61**: 1101-1119.
- Clarke KR (1993) Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* **18**: 117-143.
- Clavel J, Julliard R & Devictor V (2011) Worldwide decline of specialist species: toward a global functional homogenization? *Front Ecol Environ* **9**: 222-228.
- Comte J, Lindström ES, Eiler A & Langenheder S (2014) Can marine bacteria be recruited from freshwater sources and the air? *The ISME Journal* **8**: 2423-2430.
- Comte J, Berga M, Severin I, Logue JB & Lindstrom ES (2017) Contribution of different bacterial dispersal sources to lakes - population and community effects in different seasons: Bacterial dispersal in lakes. *Environmental Microbiology* **19**: 2391-2404.
- Conlisk J, Conlisk E, Kassim R, Billick I & Harte J (2012) The shape of a species' spatial abundance distribution. *Global Ecology and Biogeography* **21**: 1167-1178.
- Crump BC, Armbrust EV & 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**: 3192-3204.
- Crump BC, Amaral-Zettler LA & Kling GW (2012) Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME Journal* **6**: 1629-1639.
- Delignette-Muller ML & Dutang C (2014) fitdistrplus: an R Package for fitting distributions. *Journal of Statistical Software* **64**: 1-34.
- Duarte CM (2015) Seafaring in the 21st century: The Malaspina 2010 Circumnavigation Expedition. *Limnology and Oceanography Bulletin* **24**: 11-14.

- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**: 996-998.
- Estrada M, Delgado M, Blasco D, *et al.* (2016) Phytoplankton across tropical and subtropical regions of the Atlantic, Indian and Pacific oceans. *Plos One* **11**: e0151699.
- Falkowski PG, Fenchel T & Delong EF (2008) The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**: 1034.
- Farrant GK, Doré H, Cornejo-Castillo FM, *et al.* (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci USA* **113**: E3365-3374.
- Fernández-Castro B, Carballido-Mouriño B, Maraño E, *et al.* (2015) Importance of salt fingering for new nitrogen supply in the oligotrophic ocean. *Nature Communications* **6**: 8002.
- Fiedler PC (2002) The annual cycle and biological effects of the Costa Rica Dome. *Deep Sea Research I* **49**: 321-338.
- Flaviani F, Schroeder DC, Le Bret K, Balestreri C, Highfield AC, Schroeder JL & *al. e* (2018) Distinct oceanic microbiomes from viruses to protists located near the Antarctic Circumpolar Current. *Frontiers in Microbiology* **9**.
- Galand PE, Casamayor EO, Kirchman DL & Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Science of the United States of America* **106**: 22427–22432.
- Gasol JM & Del Giorgio PA (2000) Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Scientia Marina* **64**: 197-224.
- Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R & Gilbert JA (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Science of the United States of America* **110**: 4651-4655.
- Giovannoni SJ (2017) SAR11 bacteria: The most abundant plankton in the oceans. *Annual Review of Marine Science* **9**: 231-255.
- Gravel D, Bell T, Barbera C, Bouvier T, Pommier T, Venail P & Mouquet N (2011) Experimental niche evolution alters the strength of the diversity–productivity relationship. *Nature* **469**: 89.
- Hahn MW, Jezberová J, Koll U, Saueressig-Beck T & Schmidt J (2016) Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. *The ISME Journal* **10**: 1642-1655.
- Hanson CA, Fuhrman JA, Horner-Devine MC & Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **10**: 497-506.
- Hervàs A, Camarero L, Reche I & Casamayor EO (2009) Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environmental Microbiology* **11**: 1612-1623.

Hubert C, Nickel LA, Baranyi AC, *et al.* (2009) A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. *Science* **325**: 1541-1544.

Jezberová J, Jezbera J, Znachor P, Nedoma J, Kasalicky V & Simek K (2017) The Limnohabitans genus harbors generalistic and opportunistic subtypes: evidence from spatiotemporal succession in a canyon-shaped reservoir. *Applied and Environmental Microbiology* **83**: e01530.

Kashtan N, Roggensack SE, Rodrigue S, *et al.* (2014) Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus. *Science* **344**: 416-420.

Kirchman D, Knees E & Hodson R (1985) Leucine incorporation and its potential as a measure of protein-synthesis by bacteria in natural aquatic systems. *Applied and Environmental Microbiology* **49**: 599-607.

Kirchman DL (2016) Growth rates of microbes in the oceans. *Annual Review of Marine Science* **8**: 285-309.

Ladau J, Sharpton TJ, Finucane MM, *et al.* (2013) Global marine bacterial diversity peaks at high latitudes in winter. *The ISME Journal* **7**: 1669-1677.

Landa M, Blain S, Christaki U, Monchy S & Obernosterer I (2016) Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. *The ISME Journal* **10**: 39-50.

Langenheder S, Comte J, Langenheder S, *et al.* (2016) Remnants of marine bacterial communities can be retrieved from deep sediments in lakes of marine origin. *Environmental Microbiology Reports* **8**: 479-485.

Lara E, Vaqué D, Sa EL, *et al.* (2017) Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science Advances* **3**: e1602565.

Lauro FM, McDougald D, Thomas T, *et al.* (2009) The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527-15533.

Lindh MV, Sjöstedt J, Andersson AF, *et al.* (2015) Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environmental Microbiology* **17**: 2459-2476.

Liu L, Yang J, Yu Z & Wilkinson DM (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *The ISME Journal* **9**: 2068-2077.

Livermore JA, Emrich SJ, Tan J & Jones SE (2014) Freshwater bacterial lifestyles inferred from comparative genomics. *Environmental Microbiology* **16**: 746-758.

Locey KJ & Lennon JT (2016) Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA* **113**: 5970-5975.

Logares R, Lindström ES, Langenheder S, *et al.* (2013) Biogeography of bacterial communities exposed to progressive long-term environmental change. *The ISME Journal* **7**: 937-948.

Logares R, Deutschmann IM, Giner CR, *et al.* (2018) Different processes shape prokaryotic and picoeukaryotic assemblages in the sunlit ocean microbiome. *bioRxiv*
<https://doi.org/10.1101/374298>.

Logares R, Audic S, Bass D, *et al.* (2015) Patterns of rare and abundant marine microbial eukaryotes. *Current Biology* **24**: 813-821.

Longhurst A (1998) *Ecological geography of the sea*. Academic, San Diego, CA.

Luna GM, Chiggiato J, Quero GM, Schroeder K, Bongiorno L, Kalenitchenko D & Galand PE (2016) Dense water plumes modulate richness and productivity of deep sea microbes. *Environmental Microbiology* **18**: 4537–4548.

Martiny JBH (2015) Dispersal and the Microbiome: Learning how fast and how far microorganisms move will help us better understand the diversity of microbial communities. *Microbe* **10**: 191-196

Matias MG, Combe M, Barbera C & Mouquet N (2013) Ecological strategies shape the insurance potential of biodiversity. *Frontiers in Microbiology* **3**.

Mayol E, Arrieta JM, Jimenez MA, *et al.* (2017) Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nature Communications* **8**: 201.

Mestre M, Borrull E, Sala MM & Gasol JM (2017) Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME Journal* **11**: 999–1010.

Mestre M, Ruiz-González C, Logares R, Duarte CM, Gasol JM & Sala MM (2018) Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Science of the United States of America*. *115*(29): 6799-6807.

Morán XA, Gasol JM, Pernice MC, *et al.* (2017) Temperature regulation of marine heterotrophic prokaryotes increases latitudinally as a breach between bottom-up and top-down controls. *Global Change Biology* **23**: 3956-3964.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA & Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806-810.

Nemergut DR, Costello EK, Hamady M, *et al.* (2011) Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology* **13**: 135-144.

Newton RJ & Shade A (2016) Lifestyles of rarity: understanding heterotrophic strategies to inform the ecology of the microbial

rare biosphere. *Aquatic Microbial Ecology* **78**: 51-63.

Niño-García JP, Ruiz-González C & del Giorgio PA (2016) Landscape-scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. *Ecology Letters* **19**: 1506–1515

Niño-García JP, Ruiz-González C & del Giorgio PA (2016) Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal* **10**: 1755–1766.

Niño-García JP, Ruiz-González C & del Giorgio PA (2017) Exploring the ecological coherence between the spatial and temporal patterns of bacterioplankton in boreal Lakes. *Frontiers in Microbiology* **8**: 636.

- Oksanen J, F., Blanchet G, Kindt R, *et al.* (2015) Vegan: Community Ecology Package. R package version 2.3-0.
- Parada AE, Needham DM & Fuhrman JA (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* **18**: 1403–1414.
- Partensky F & Garczarek L (2010) Prochlorococcus: Advantages and Limits of Minimalism. *Annual Review of Marine Science* **2**: 305-331.
- Pedrós-Alió C (2012) The rare bacterial biosphere. *Annual Review of Marine Science* **4**: 449-466.
- Pedrós-Alió C, Acinas SG, Logares R & Massana R (2018) Marine microbial diversity as seen by high throughput sequencing. *Microbial Ecology of the Oceans* (Gasol JM & Kirchman DL, eds.), pp. 592. Wiley-Blackwell, Hoboken, New Jersey.
- Reche I, D'Orta G, Mladenov N, Winget DM & Suttle CA (2018) Deposition rates of viruses and bacteria above the atmospheric boundary layer. *The ISME Journal*.
- Rieck A, Herlemann DPR, Jurgens K & Grossart HP (2015) Particle-associated differ from free-living bacteria in surface waters of the Baltic Sea. *Frontiers in Microbiology* **6**: 1297.
- Ruiz-González C, Niño-García JP & del Giorgio PA (2015) Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters* **18**: 1198-1206.
- Ruiz-González C, Niño-García JP, Kembel SW & del Giorgio PA (2017) Identifying the core seed bank of a complex boreal bacterial metacommunity. *The ISME Journal* **11**: 2012-2021.
- Salazar G, Cornejo-Castillo FM, Benitez-Barrios V, *et al.* (2015) Global diversity and biogeography of deep-sea pelagic prokaryotes. *The ISME Journal* **10** 596–608.
- Salazar G, Cornejo-Castillo FM, Borrull E, *et al.* (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Molecular Ecology* **24**.
- Salcher MM (2014) Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *Journal of Limnology* **73**: 74-87.
- Sebastián M, Auguet J-C, Restrepo-Ortiz CX, Sala MM, Marrasé C & Gasol JM (2018) Deep ocean prokaryotic communities are remarkably malleable when facing long-term starvation. *Environmental Microbiology* **20**: 713-723.
- Ser-Giacomi E, Zinger L, Malviya S, de Vargas C, Karsenti E, Bowler C & de Monte S (2018) Ubiquitous abundance distribution of non-dominant plankton across the world's ocean. *Nature Ecology and Evolution*.
- Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N & Gilbert JA (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* **5**: e01371-01314.
- Staley C, Unno T, Gould TJ, Jarvis B, Phillips JB, Cotner JB & Sadowsky MJ (2013) Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *Journal of Applied Microbiology* **115**: 1147-1158.

- Stocker R (2012) Marine microbes see a sea of gradients. *Science* **338**: 628-633.
- Sunagawa S, Coelho LP, Chaffron S, *et al.* (2015) Structure and function of the ocean microbiome. *Science* **348**: 1261359-1261351.
- Swan BK, Tupper B, Sczyrba A, *et al.* (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci USA* **110**: 11463–11468.
- Teeling H, Fuchs BM, Becher D, *et al.* (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.
- Vergin KL, Done B, Carlson CA & Giovannoni SJ (2013) Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *The ISME Journal* **71**: 1-13.
- Wilhartitz I, Mach RL, Teira E, Reinthaler T, Herndl GJ & Farnleitner AH (2007) Prokaryotic community analysis with CARD-FISH in comparison with FISH in ultra-oligotrophic ground- and drinking water. *Journal of Applied Microbiology* **103**: 871-881.
- Wilkins D, Seville E, Rinthoul SR, Lauro FM & Cavicchioli R (2013) Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. *Nature Communications* **4**: 2457.
- Yooseph S, Neelson K, Rusch D, *et al.* (2010) Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60-67.
- Zhang J, Kobert K, Flouri T & Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614-620.
- Zinger L, Amaral-Zettler LA, Furrman JA, *et al.* (2011) Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *Plos One* **6**: e24570.

FIGURE LEGENDS

Figure 1. (a) World map showing the location of the surface stations sampled during the Malaspina 2010 expedition. The seven different transects performed are indicated with numbers. Asterisks (*) indicate 8 stations where prokaryotic communities from 3 additional depth layers (DCM, meso- and bathypelagic) associated to 5 different size-fractions were also analyzed (see Methods for further details). (b-d) Example histograms of abundance distributions of representative OTUs belonging to the 3 different spatial distribution categories across the 108 surface stations. The orange line represents the shape of the distribution as a density line. The height of the bars represents the number of stations for a given class of abundance ($\text{Log}_{10}(\text{number of sequences}+1)$) for each individual OTU.

Figure 2. Contribution of the SpAD categories to the surface microbial communities (n=108). a) Mean abundance (number of sequences), b) occurrence (number of stations where an OTU was present), c) contribution of each SpAD category to the total number of free-living OTUs and their associated sequences. d) contribution, at each station, of each SpAD category to the total community sequences. Stations are ordered following the cruise path as in Figure 1a for comparison.

Figure 3. (a-c) Relative contribution of the three types of SpAD categories to the communities contained within each of the four clusters (NL1,NL2, BM, LG) of stations (i.e. *k*-means classification at 4 groups partition). **d)** World map showing the distribution of the stations coloured by cluster. The clusters are coloured depending on the SpAD category that shows the highest contribution within them (for example, the green cluster (LG) comprises those stations with the highest percentages of ‘logistic’ OTU sequences, coloured in green in panel c).

Figure 4. Changes in the relative abundances of normal-like, bimodal and logistic taxa along temperature gradients in the three areas potentially influenced by upwelled waters, which were located in Leg 2 (South Atlantic), Leg 4 (South Australian Bight) and Leg 5 (Equatorial Pacific). The line in the case of bimodal OTUs is a best-fit smooth curve through the centre of the data calculated using weighted least squares, in order to show that their maximum abundances appeared somewhere in between the temperature gradient.

Figure 5. Inter-cluster variation of satellite-derived or chlorophyll concentration (Chl, $\mu\text{g L}^{-1}$, a-c) or sea surface temperature (SST, d-f) taken over different time periods (8 days, 1 month or the whole year prior to sampling, see Methods for details). LG stations were split into two groups (7 Equatorial Pacific stations –‘LGpac’-, and the rest) for visualizing their very different conditions. Note that a gradient of increasing chlorophyll concentrations from cluster NL1 to LG is only apparent when the annual variability is taken into account (c), but not when the *in situ* measurements (Table 1) or shorter periods are considered (a,b). The dotted lines indicate the global median of the values considered in each case, to visualize deviations from the prevailing subtropical and tropical environmental conditions in among the sampled stations.

Figure 6. Change in the average contribution of SpAD categories (% of community sequences) when conducting the SpAD analysis using datasets of increasing spatial or environmental heterogeneity (see Methods). a) Original SpADs defined based on the surface free-living dataset (108 stations), or b) including communities attached to the three largest size fractions (3-5, 5-20, and 20-200 μm) from surface waters of 8 selected stations, c) including free-living communities from surface, DCM, meso- and bathypelagic layers in 8 selected stations, and d) including communities associated with particles of the different sizes along the same 4 layers. Below is indicated the number of OTUs within each dataset that were categorized as normal-like, bimodal or logistic.

Table 1. Comparison of the environmental and biological properties characterizing the stations within each of the four clusters (NL1, NL2, BM, LG). Different letters in the same row represent values that are significantly different from each other (Tukey's post hoc test; $P < 0.001$). Values are means of 54 (NL1), 29 (NL2), 11 (BM) and 14 (LG) stations, respectively. Cluster LG included stations with contrasting environmental conditions, so we divided it into two groups of sites (seven Equatorial Pacific stations and the rest of LG stations) to show that in many cases the differences observed between the four clusters were due to higher values in the Equatorial Pacific LG stations but not in the rest. Areas shaded in grey show the only cases where all LG communities showed consistently higher values than the other clusters. *MLD*, Mixed layer depth; *dSWR*, *mSWR*, daily and monthly short wave solar radiation, respectively.

	Cluster NL1	Cluster NL2	Cluster BM	Cluster LG	Within Cluster LG	
					7 Eq. Pacific stations (LGpac)	Rest of LG stations
Prokaryotic community features						
Number of OTUs	1131	1122	1067	1063	1223	943
Taxonomic diversity (Shannon)	3.8 ^a	4.6 ^b	4.1 ^a	4.8 ^b	5.0	4.6
Measured site characteristics						
Temperature (°C)	25.9 ^a	23.8 ^b	20.5 ^c	23.3 ^{bc}	27.4	20.1
Dissolved oxygen (ml L ⁻¹)	4.4 ^a	4.5 ^b	4.8 ^c	4.6 ^{bc}	4.2	4.9
Turbidity (FTU)	0.07	0.07	0.05	0.12	0.04	0.19
Salinity (PSU)	35.5 ^{ab}	36.0 ^a	36.0 ^{ab}	35.2 ^b	35.1	35.3
MLD (m)	49.1	42.3	46.5	56.3	84.5	42.3
dSWR (MJ m ⁻² d ⁻¹)	251.2	252.2	254.1	237.3	274.5	218.8
mSWR (MJ m ⁻² d ⁻¹)	308.6	315.7	303.4	300.7	345.3	278.4
SiO ₄ (μmol L ⁻¹)	1.26	1.16	0.92	1.44	1.3	1.5
NO ₂ (μmol L ⁻¹)	0.01 ^a	0.01 ^a	0.01 ^a	0.05 ^b	0.11	0.02
PO ₄ (μmol L ⁻¹)	0.08 ^a	0.07 ^a	0.10 ^{ab}	0.16 ^b	0.20	0.15
NO ₃ (μmol L ⁻¹)	0.29 ^a	0.34 ^a	0.26 ^{ab}	1.04 ^b	1.6	0.7
Biological properties						
Prokaryotic abundances (10 ⁵ cell mL ⁻¹)	7.6	8.3	6.4	9.1	3.3	13.5
HNA prokaryotes (%)	32.6 ^a	32.2 ^a	28.5 ^a	44.2 ^b	49.9	39.9
Prokaryotic size (μm ³)	0.037 ^a	0.041 ^a	0.036 ^a	0.064 ^b	0.10	0.04
Prok. heterotrophic activity (pmol	43.3 ^a	51.7 ^a	12. ^a	156.3 ^b	323.5	30.9

Leu L ⁻¹ h ⁻¹)						
Virus abundance (10 ⁶ mL ⁻¹)	7.1 ^{ab}	5.0 ^a	8.9 ^{ab}	13.9 ^b	12.8	15.0
Chlorophyll concentration (μg L ⁻¹)	0.15 ^a	0.14 ^a	0.10 ^a	0.26 ^b	0.28	0.25







