Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans

Carlota R. Gazulla [1,2*] Adrià Auladell [2]
Clara Ruiz-González [2] Pedro C. Junger [3]
Marta Royo-Llonch [2] Carlos M. Duarte [4]
Josep M. Gasol [2,6] Olga Sánchez [1**] and
Isabel Ferrera [5***]

[1]*Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Catalunya, 08193, Spain.*

[2]*Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, ICM-CSIC, Barcelona, Catalunya, 08003, Spain.*

[3]*Department of Hydrobiology (DHB), Laboratory of Microbial Processes and Biodiversity (LMPB), Universidade Federal de São Carlos (UFSCar), São Carlos, SP, 13565-905, Brazil.*

[4]*Red Sea Research Center (RSRC) and Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.*

[5]*Centro Oceanográfico de Málaga, Instituto Español de Oceanografía, IEO-CSIC, 29640 Fuengirola, Málaga, Spain.*

[6]*Centre for Marine Ecosystems Research, School of Sciences, Edith Cowan University, Joondalup, WA, Australia.*

Summary

The aerobic anoxygenic phototrophic (AAP) bacteria are common in most marine environments but their global diversity and biogeography remain poorly characterized. Here, we analyzed AAP communities across 113 globally-distributed surface ocean stations sampled during the Malaspina Expedition in the tropical and subtropical ocean. By means of amplicon sequencing of the *puf*M gene, a genetic marker for this functional group, we show that AAP communities along the surface ocean were mainly composed of members of the Halieaceae (Gammaproteobacteria), which were adapted to a large range of environmental conditions, and of different clades of the Alphaproteobacteria, which seemed to dominate under particular circumstances, such as in the oligotrophic gyres. AAP taxa were spatially structured within each of the studied oceans, with communities from adjacent stations sharing more taxonomic similarities. AAP communities were composed of a large pool of rare members and several habitat specialists. When compared to the surface ocean prokaryotic and picoeukaryotic communities, it appears that AAP communities display an idiosyncratic global biogeographical pattern, dominated by selection processes and less influenced by dispersal limitation. Our study contributes to the understanding of how AAP communities are distributed in the horizontal dimension and the mechanisms underlying their distribution across the global surface ocean.

Introduction

The discovery of marine aerobic photoheterotrophs (i.e., aerobic anoxygenic phototrophic (AAP) bacteria and proteorhodopsin-containing bacteria) (Béjà *et al*., 2000; Kolber *et al*., 2000) challenged the classic view of bacterioplankton being composed of photoautotrophic microorganisms as primary producers and of chemoheterotrophs as consumers. Since then, many studies have investigated their abundance, diversity and distribution in the ocean, and ultimately tried to understand their role in the marine ecosystem (DeLong and Béjà, 2010; Kirchman and Hanson, 2013; Koblížek, 2015; Pinhassi *et al*., 2016). AAP bacteria are photoheterotrophs that use dissolved organic matter but harvest solar energy using bacteriochlorophyll *a* (Bchla) to supplement their metabolism. In the marine environment, these organisms can typically constitute up to 10% of total prokaryotes (Schwalbach and Fuhrman, 2005; Sieracki *et al*., 2006; Jiao *et al*., 2007; Hojerová *et al*., 2011), and are an active part of the community because they consist of large cells that display higher growth rates and receive higher grazing pressure than most bacteria (Sieracki *et al*., 2006;

Koblížek *et al*., 2007; Ferrera *et al*., 2011, 2017). It has thus been hypothesized that this functional group plays a remarkably important role in the processing of organic matter and, as a consequence, in the global carbon cycle (see review by Koblížek, 2015).

Phylogenetically, marine AAP bacteria belong mainly to the Alpha- and Gammaproteobacteria classes. The *puf*M gene, which encodes the M subunit of the photosynthetic reaction centre, is commonly used to screen the diversity of AAPs in environmental samples and to describe their distribution patterns. The first studies showed AAP communities as being mainly affiliated to the alphaproteobacterial *Roseobacter*-like clade (Béjà *et al*., 2002; Oz *et al*., 2005) but the global ocean sampling (GOS), based on metagenomic data, unveiled that an important fraction of marine AAP bacteria were associated to phylogroups without cultured representatives (Yutin *et al*., 2007). The later study also showed that, while the *Roseobacter*-like AAPs were the most ubiquitous clade, unidentified uncultured groups dominated in open ocean areas, while Gammaproteobacteria dominated in coastal sites (Yutin *et al*., 2007). Later investigations showed that Gammaproteobacteria have in fact a widespread distribution and can constitute an important fraction of AAP communities in diverse sites of contrasting trophic status (Mašín *et al*., 2006; Lehours *et al*., 2010; Ferrera *et al*., 2014; Lehours and Jeanthon, 2015; Auladell *et al*., 2019). In contrast, AAPs from the Betaproteobacteria clade – that currently belongs to the Gammaproteobacteria class, according to the Genome Taxonomy Database, GTDB (Parks *et al*., 2018) – are rarely prevalent in marine environments and they seem to prefer low-salinity waters (Waidner and Kirchman, 2008; Cottrell and Kirchman, 2009; Boeuf *et al*., 2013).

Although most AAP diversity studies have been restricted to particular areas of the world's ocean, a few studies have already compared communities across different oceanic regions. The pioneering metagenomic study by Yutin *et al*. (2007), which covered a transect between 45°N in the Atlantic Ocean and 15°S in the Pacific Ocean, showed that the composition of AAP communities varied between different biogeographical regions. By constructing clone libraries in a limited number of samples (N = 10) from the Pacific, Atlantic and Indian oceans, Jiao *et al*. (2007) reported diversity patterns linked to the trophic regime of the oceanic region. Later, another study compared clone libraries of different seas encompassing a very large environmental variability (Mediterranean Sea, North Pacific Ocean, Western Beaufort Sea, Barents Sea and Norwegian Sea), and found that deterministic processes largely influenced the structuring of AAP assemblages (Lehours *et al*., 2018). This study further concluded that diverse AAP lineages showed some habitat preference, suggesting the existence of a certain degree of ecological cohesiveness for AAP clades, at least when comparing contrasting biomes. Besides, a study applying high-throughput sequencing to coastal Australian waters concluded that AAP communities exhibited niche partitioning whereas others shared their preferred niches (Bibiloni-Isaksson *et al*., 2016). Altogether, these results indicate that AAP assemblages – and the taxa within them – display complex spatial patterns (Jiao *et al*., 2007; Yutin *et al*., 2007; Lehours *et al*., 2010; Jeanthon *et al*., 2011; Boeuf *et al*., 2013; Lehours and Jeanthon, 2015; Bibiloni-Isaksson *et al*., 2016), probably driven by environmental selection (Lehours *et al*., 2018). Nevertheless, these conclusions are drawn from studies performed at different scales, using various methodologies and biassed towards particular – and often coastal – ocean regions, so a coherent global assessment is still lacking.

The exploration of the worldwide distribution of marine microorganisms, and thus, the definition of global biogeographical patterns, has become feasible in the last decade thanks to contemporary global oceanographic circumnavigations like the Malaspina Circumnavigation Expedition (Duarte, 2015) or the *Tara* Oceans Expedition (Karsenti *et al*., 2011), that used standardized procedures in a large collection of samples, coupled with recent advances in sequencing methodologies. Large scale surveys have also been key in the definition of the underlying ecological mechanisms in bulk prokaryotic and small eukaryotic communities (de Vargas *et al*., 2015; Salazar *et al*., 2015; Sunagawa *et al*., 2015; Ruiz-González *et al*., 2019; Logares *et al*., 2020; Obiol *et al*., 2020). Data generated from large sequencing initiatives have also been used to retrieve new diversity (Tully *et al*., 2018; Nayfach *et al*., 2020), including that within the AAPs (from the *Tara* Oceans expedition, Graham *et al*., 2018). Hence, a comprehensive study defining the global ocean biogeography of AAP assemblages and the mechanisms underlying their patterns is now feasible, but yet to be performed.

Here, we present a global assessment of AAP bacteria communities across the global tropical and subtropical ocean based on the Malaspina Circumnavigation Expedition. In particular, we studied the diversity and biogeography of AAP communities at a fine scale in the surface ocean using Amplicon Sequence Variants (ASVs) of the *puf*M gene. Our objectives were three-fold: (i) to describe the diversity and biogeography of the surface AAP assemblages along the global tropical and subtropical ocean, (ii) to disentangle the factors driving global patterns of AAP communities, and (iii) to compare the trends observed in the AAP communities with those of the broader surface ocean microbiota (i.e., whole prokaryotic and picoeukaryotic communities). For this purpose, we

analyzed the composition of AAP communities based on the dominance or rarity of each individual taxa in an approach based on the spatial abundance distribution of each ASV. Furthermore, we estimated the role of different ecological processes shaping the structure of AAP communities. Since AAP bacteria, as a whole, display ecological traits that differentiate them from the rest of the bacterioplankton (i.e., photoheterotrophy, higher growth rates and higher susceptibility to predation than other prokaryotes), we hypothesize that their ecological patterns may deviate from those of the bulk communities.

Results and discussion

Oceanographic context

The 113 studied stations were representative of the tropical and subtropical regions of the three major oceans, the Pacific, the Atlantic and the Indian Ocean (Table 1, Fig. S1). The cruise track spanned across all five subtropical oceanic gyres, characterized by their oligotrophic conditions, as well as over relatively more productive areas such as the Equatorial Pacific, the Caribbean Sea, the Benguela Coastal province or the South Subtropical Convergence Zone, in the South Australian Bight (Estrada *et al*., 2016). The schedule of the cruise was planned so that most of the stations were sampled during spring and summer in order to avoid adverse weather conditions and allowing seasonal comparability.

Across this route, temperatures ranged between 15.8 and 29.3°C (mean 24.5°C), with the coldest waters found in the South Australian Bight and the warmest temperatures in stations located along the Equatorial Pacific and Atlantic Oceans (Fig. S2). Salinity ranged from 33.15 to 37.65, being the highest in stations from the Atlantic Ocean and lowest in certain stations from the Indian and Pacific Oceans. Chlorophyll *a* (Chla) ranged between 0.034 (station 38 in the South Atlantic) and 0.647 mg·m^{-3} (station 45 in the Benguela Current Coastal province) with a mean value of 0.155 mg·m^{-3}. Phosphate, nitrate and silicate had higher concentrations in the Equatorial Pacific, in the South African stations and in the South of Australia (Fig. S2). Water mass properties and productivity regimes for the stations sampled in the Malaspina Circumnavigation Expedition have been previously described in detail (e.g., Estrada *et al*., 2016; Regaudie-de-Gioux *et al*., 2019; Teira *et al*., 2019 and Villamaña *et al*., 2019).

Contrasting patterns of alpha diversity in distinct biogeographical provinces

Our survey of the *puf*M gene allowed us to generate the largest dataset of Amplicon Sequence Variants of the *puf*M gene so far available. Partial sequencing of this marker resulted in 1119 distinct ASVs that clustered into 229 OTUs (Operational Taxonomic Units of 94% similarity). Rarefaction curves reached a plateau for all samples (Fig. S3A), indicating that we obtained a fair representation of the AAPs' surface ocean diversity for each individual sample, while the global sample-based rarefaction curve (Fig. S3B) suggested that the number of ASVs would rise had more stations been sampled. We observed a large variability in the richness estimates (Chao1 index) per community (Fig. 1), which varied between 14 and 132 ASVs (mean 61), while the Shannon diversity index ranged between 0.9 and 3.9 (mean 2.9). Overall, richness values were within the same range than those previously reported from the Mediterranean Sea or Australian coastal waters using similar methodologies (Bibiloni-Isaksson *et al*., 2016; Auladell *et al*., 2019).

Richness and diversity of AAP communities were highest in the North Atlantic (mean richness 81, mean Shannon diversity 3.1) compared to other regions (Tukey test, $P < 0.001$, Fig. 1). Taxonomic richness and diversity varied between and within some Longhurst provinces. In general, AAP bacteria diversity was lower in eutrophic areas (correlation between Shannon and Chla concentration, $N = 107$, $R = -0.33$, $P < 0.001$ and primary production, $N = 96$, $R = -0.38$, $P < 0.001$), consistent with previous observations (Jiao *et al*., 2007; Jeanthon *et al*., 2011). In contrast, AAP communities having higher richness values were associated with low concentrations of phosphate ($N = 89$, $R = -0.48$, $P < 0.0001$) and nitrate ($N = 89$, $R = -0.34$, $P = 0.001$), and correlated positively with temperature and salinity ($N = 113$, $R = 0.24$, $P = 0.011$; $R = 0.29$, $P = 0.002$, respectively, see Table S1). Temperature and salinity had been shown to influence AAP bacterial richness at local scales (Lehours and Jeanthon, 2015; Bibiloni-Isaksson *et al*., 2016). Our results demonstrate that temperature, salinity and trophic status govern patterns of AAP bacterial alpha diversity at the global scale.

We also explored whether the patterns of AAP diversity were similar to the trends observed for other picoplanktonic groups. To this end, we compared the Shannon diversity of prokaryotes and picoeukaryotes (previously determined in the same sample set, Ruiz-González *et al*., 2019; Logares *et al*., 2020) with the values obtained for AAP bacteria. We observed a significant negative correlation between the Shannon diversity index of AAP communities and that of total prokaryotes ($N = 104$, $R = -0.32$, $P = 0.001$, Fig. S4), while no significant correlation was found with the picoeukaryotic community values. However, the low diversity values observed for AAP bacteria in some eutrophic regions (PEQD, PNEC and SSTC provinces, cf. Table 1 for

Table 1. Provinces covered during the Malaspina Circumnavigation Expedition and values (average ± standard deviation, minimum–maximum) of temperature, salinity and chlorophyll *a* concentration measured in each province.

Provinces	Province abbreviations	N	Temperature (°C)	Salinity	Chlorophyll *a* (mg·m^{-3})
East Australian Coast	AUSE	2	21.39 ± 0.35 (21.14–21.64)	35.52 ± 0.08 (35.47–35.58)	0.34 ± 0.02 (0.32–0.36)
Western Australian and Indonesian Coast	AUSW	4	23.05 ± 1.41 (21.36–24.8)	35.48 ± 0.09 (35.34–35.54)	0.13 ± 0.03 (0.1–0.16)
Benguela Current Coastal	BENG	2	20.55 ± 0.16 (20.44–20.66)	35.52 ± 0.06 (35.48–35.56)	0.14 ± 0.11 (0.06–0.22)
Caribbean	CARB	4	28.73 ± 0.29 (28.38–29.09)	35.6 ± 0.08 (35.54–35.71)	0.14 ± 0.04 (0.09–0.19)
East Africa Coastal	EAFR	3	23.94 ± 1.82 (22.56–26)	35.41 ± 0.12 (35.31–35.54)	0.3 ± 0.3 (0.09–0.65)
Indian Ocean South Subtropical Gyre	ISSG	14	23.57 ± 1.36 (21.74–25.92)	35.65 ± 0.25 (35.23–36.14)	0.07 ± 0.03 (0.04–0.14)
Northeast Atlantic Subtropical Gyre	NASE	10	21.35 ± 1.88 (18.45–24.31)	37.03 ± 0.39 (36.43–37.65)	0.1 ± 0.07 (0.04–0.25)
North Atlantic Tropical Gyre	NATR	11	26.82 ± 1.32 (24.83–28.85)	36.68 ± 0.67 (35.53–37.57)	0.14 ± 0.1 (0.05–0.31)
North Pacific Tropical Gyre	NPTG	13	23.86 ± 1.49 (21.65–26.35)	34.66 ± 0.2 (34.2–34.94)	0.17 ± 0.09 (0.08–0.44)
Pacific Equatorial Divergence	PEQD	3	27.5 ± 0.62 (26.89–28.13)	35.21 ± 0.31 (34.85–35.39)	0.24 ± 0.05 (0.19–0.29)
North Pacific Equatorial Countercurrent	PNEC	8	28.37 ± 0.58 (27.61–29.28)	33.84 ± 0.4 (33.15–34.28)	0.34 ± 0.11 (0.18–0.52)
South Atlantic Tropical Gyre	SATL	19	24.7 ± 2.26 (20.9–27.33)	36.49 ± 0.45 (35.79–37.25)	0.07 ± 0.03 (0.03–0.12)
South Pacific Subtropical Gyre	SPSG	7	28 ± 2 (23.99–29.31)	35.04 ± 0.41 (34.43–35.59)	0.11 ± 0.05 (0.06–0.18)
South Subtropical Convergence	SSTC	7	17.29 ± 1.35 (15.75–19.55)	35.3 ± 0.2 (34.99–35.61)	0.25 ± 0.14 (0.1–0.52)
Western Tropical Atlantic Province	WTRA	6	27.6 ± 0.29 (27.27–28.05)	35.77 ± 0.38 (35.42–36.33)	0.23 ± 0.11 (0.11–0.44)

Names and abbreviations according to Longhurst (2007). N = number of stations visited in each Longhurst province.

complete names) were not observed in the whole pro-karyotic dataset, suggesting that trophic status may exert a strongest role in shaping the diversity of the AAP sub-community than of the bulk prokaryotic assemblage.

### Spatially structured communities dominated by distinct taxonomic groups

We classified all the ASVs into seven broad taxonomic groups based on their placement in a reference phylo-genetic tree (Fig. S5). One group contained sequences assigned to the family Halieaceae of the Gammaproteobacteria (here-after 'Gamma-Halieaceae' group), while the 'Gamma-Burkholderiales' group included sequences from Burkholderiales order (for-merly affiliated to the Betaproteobacteria class, but now reassigned within the Gammaproteobacteria class, according to the GTDB; Parks *et al.*, 2018). Members of the Alphaproteobacteria were distinguished into four sub-groups: 'Methylobacteriaceae' (sequences from order Rhizobiales, family Methylobacteriaceae), 'Rhodobacterales' (order Rhodobacterales), 'Sphingomondales' (order Sphingomonadales), and 'Alpha-Others', which grouped other members of the Alphaproteobacteria that could not

be further assigned. Finally, sequences that did not belong to any of these groups were classified as 'Others'.

Most of the studied communities (75 out of 113 sam-pled stations, Fig. 2 and Fig. S6) were dominated by Gamma-Halieaceae, followed by 24 stations dominated by Alpha-Rhodobacterales. The overall dominance of these groups is in agreement with previous studies from the Mediterranean Sea (Lehours *et al.*, 2010; Jeanthon *et al.*, 2011; Ferrera *et al.*, 2014; Auladell *et al.*, 2019), the Baltic Sea (Mašín *et al.*, 2006), the Arctic Ocean (Lehours and Jeanthon, 2015), and Australian waters (Bibiloni-Isaksson *et al.*, 2016).

The large dominance of gammaproteobacterial clades in marine AAP communities has been a matter of debate; it has been argued that it could be due to possible primer biases in amplicon-based studies (Lehours *et al.*, 2010; Ferrera *et al.*, 2014). In fact, PCR-based approaches can suffer from amplification biases that could result in mis-representation of the relative abundances of various taxa as well as in low phylogenetic coverage. Nevertheless, a recent comparison of AAP assemblages in the Mediterra-nean Sea using metagenomics and *puf*M amplicon sequencing showed that, despite there were some dis-crepancies in the relative abundance of certain taxa,

Richness (Chao1 index)

Shannon diversity index
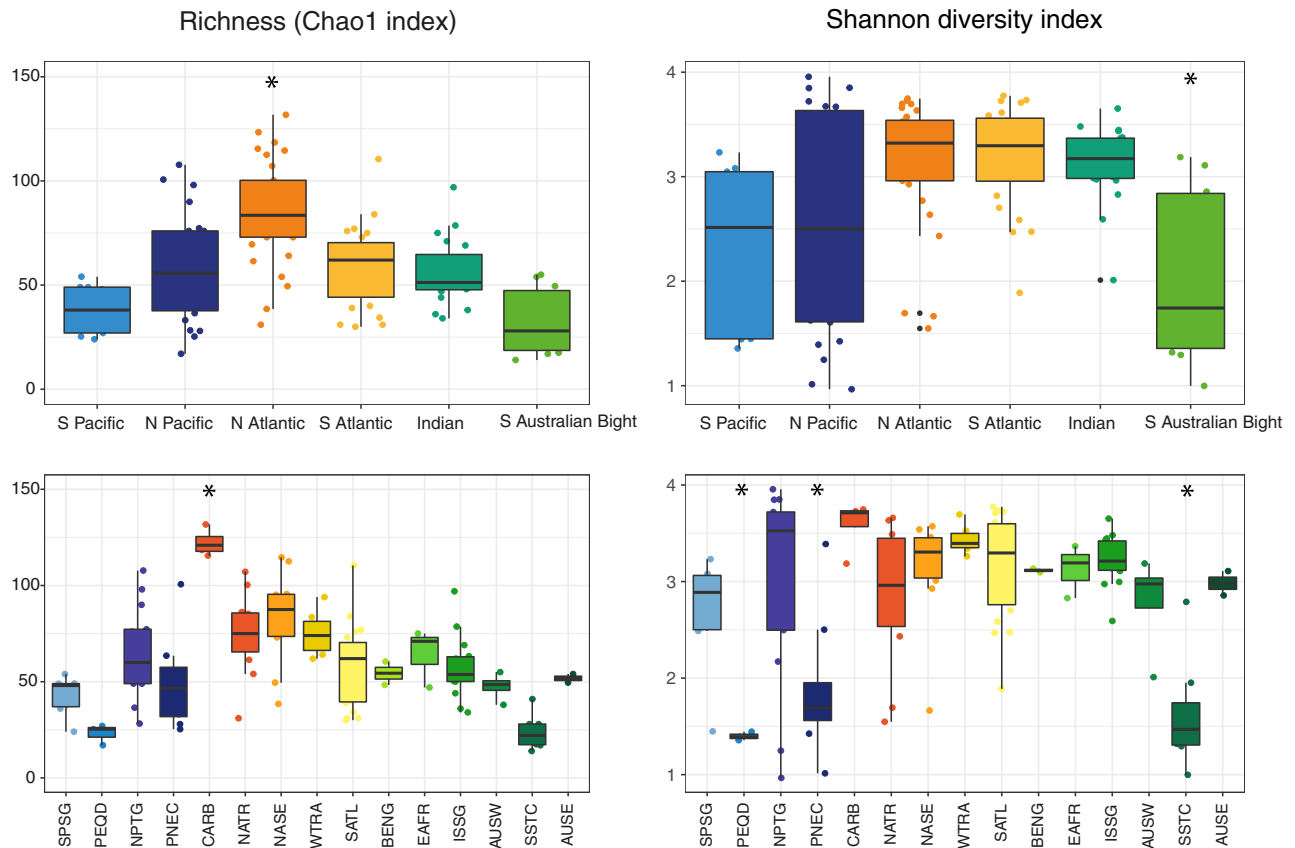
Fig. 1. AAP alpha diversity measured as richness (Chao1 index) and Shannon diversity index within each oceanic region (top panels) and each Longhurst province (bottom panels) sampled during the Malaspina Circumnavigation Expedition. The complete names of the Longhurst provinces are listed in Table 1 and Fig. S1. Asterisks indicate regions or provinces that are statistically different from the others, after a post-hoc Tukey test (*P* < 0.001).

Gammaproteobacteria were abundant in both the amplicon and metagenomic datasets, which showed comparable patterns of diversity and community structure (Auladell *et al.*, 2019). This study also showed that despite that the amplicon approach – identical to the one used here – missed some phylogenetic groups, it allowed the identification of other groups that were overlooked by metagenomics because they were present in low abundances, as well as the retrieval of more variants, enabling the definition of distinct ecotypes among very similar sequences (Auladell *et al.*, 2019). Metagenomics is often considered the least biassed approach for functional gene analysis, but it is limited in its capacity to retrieve the least abundant members of the communities, and AAP taxa are generally present at relatively low abundances in natural samples (often below 10%). Although technically possible, the cost of conducting a high-resolution global ecological study based on a specific functional gene using metagenomics would be prohibitive and unfeasible for most researchers because, among other reasons, metagenomes retrieve less copies of specific marker genes for a given sequencing investment. Given that the goal of this work was to establish the global

ecological patterns of AAP communities, and to understand how these are assembled at the fine scale, we consider that using *puf*M amplicon sequencing, despite not free of biases, was the most suitable approach to address our questions.

Interestingly, Gamma-Halieaceae and Alpha-Rhodobacterales-dominated communities were not randomly distributed but appeared to be spatially structured, with a marked succession of samples dominated by either one or the other group (Fig. 2A). Gamma-Halieaceae contributed between 0.1% and 99.7% of total community *puf*M sequences (mean 58.42%). In locations where they were not abundant, the contribution of Alpha-Rhodobacterales was high, suggesting a replacement of the dominant taxonomic groups across space (Fig. 2). Both groups also showed high intragroup diversity across samples (Fig. 2), yet we observed that in stations 114, 115, 118 and 119, in the North Pacific region, this intragroup diversity decreased, and one single sequence assigned to the Gamma-Halieaceae (ASV217) represented abundances over 50% coinciding with a decrease in salinity. The relative contribution of Alpha-Rhodobacterales increased towards ultraoligotrophic gyre
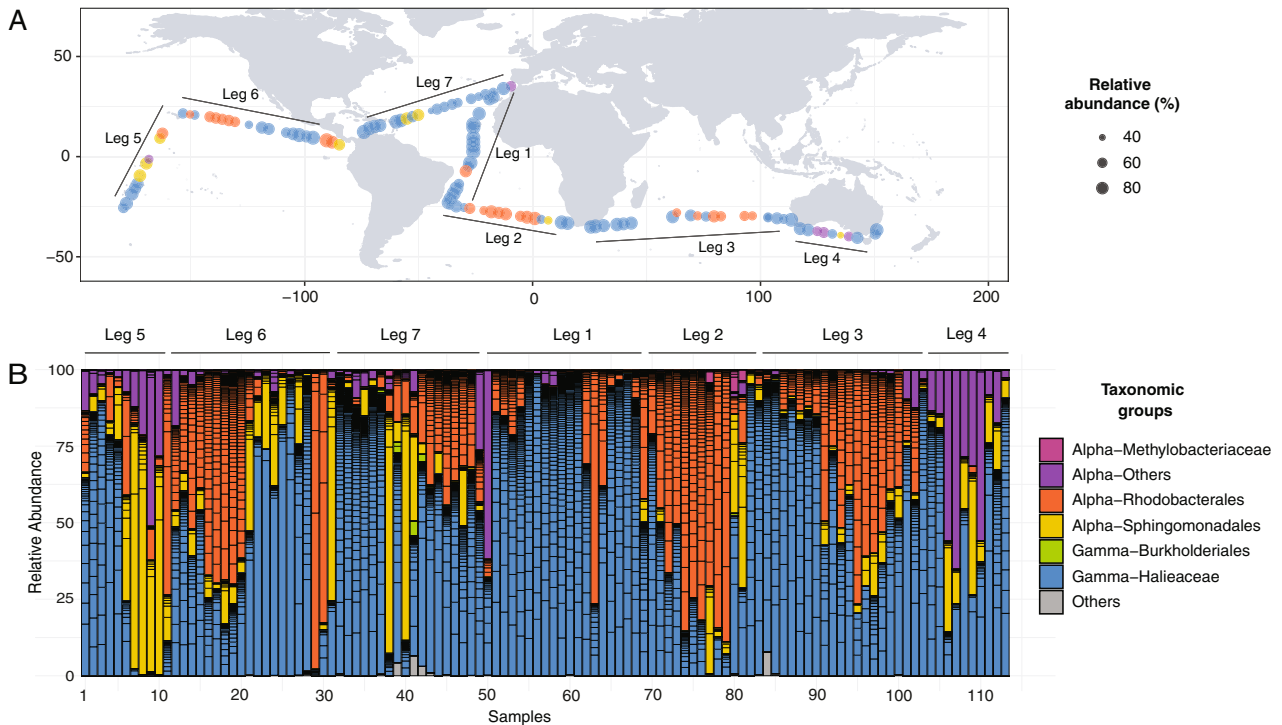
Fig. 2. A. Dominant AAP taxonomic groups across the global tropical and subtropical surface ocean. Each station is color-coded by the most abundant taxonomic group in the sample (see taxonomy legend in panel B), and the size of the dot is proportional to the relative abundance of the dominant taxon. The Malaspina Circumnavigation Expedition legs are indicated to help visualize the cruise track.
B. Community composition at each station, expressed as the relative contribution of each *puf*M sequence color-coded by its taxonomic affiliation. Samples are ordered following the cruise path as in panel A.

waters, characterized by low Chla concentrations (N = 107, R = −0.42, P < 0.001) and deeper chlorophyll maxima (N = 113, R = 0.42, P < 0.001). While the negative correlation between the contribution of Gamma-Halieaceae and Alpha-Rhodobacterales was observed before (Ferrera *et al*., 2014; Bibiloni-Isaksson *et al*., 2016; Auladell *et al*., 2019), in those cases, higher relative abundances of Alpha-Rhodobacterales were linked to higher concentrations of Chla and, in general, to higher nutrient levels. While those studies were conducted in coastal stations, the Malaspina Circumnavigation Expedition occupied open-ocean stations, yet covered contrasting regions, from some relatively eutrophic areas (such as the Equator, South African provinces or the South Australian Bight) to the oligotrophic open ocean gyres. Just like seasonal ecotypes have been defined within the Alpha-Rhodobacterales based on 16S rRNA gene sequencing (Mena *et al*., 2020), one possible explanation for the observed contrasting results is that closely related, but ecologically different Alpha-Rhodobacterales could be divided into an ecotype with a preference for productive regions such as coastal areas and an ecotype dominant in oligotrophic environments like the oceanic gyres.

Representatives from the Alpha-Sphingomonadales and 'Other Alpha' were scarce across the surface ocean with some localized exceptions (Fig. 2B). The relative abundances of Alpha-Sphingomonadales members correlated positively with prokaryotic heterotrophic production (N = 113, R = 0.44, P < 0.005), prokaryotic cell volume and total biomass (N = 113, R = 0.52, p < 0.001 and R = 0.46; P < 0.005 respectively). Interestingly, in stations where Sphingomonadales dominated, this dominance was due to a single ASV (ASV512), which contributed up to 50% of the total AAP community reads. This ASV is related to an uncultured bacterial sequence (96% of identity in the *puf*M nucleotide sequence) detected in the Beaufort Sea (Boeuf *et al*., 2013), but does not resemble any cultured AAP bacteria. Thus, information on the physiology of the organism behind this sequence is missing. In any case, its widespread distribution from the Arctic to the tropical oceans and its ability to dominate communities under different conditions are remarkable. Other ASVs that could only be assigned to the Alpha-Proteobacteria level (and were thus grouped as 'Alpha-Others') dominated communities in some stations across the whole transect (Fig. 2), such as station 93 in the Pacific Ocean, station 1 in the Atlantic Ocean, adjacent to the Strait of Gibraltar, and stations 71, 72 and

75 in the South Australian Bight, coinciding with the South Subtropical Convergence Zone (SSTC Longhurst province). In these stations, two ASVs were dominant, ASV860 in the Atlantic Ocean and ASV1102 in the Pacific and in the South of Australia. Although we could not classify them further and they do not have close cultured representatives, they are very similar to sequences from previous studies. In particular, ASV860 is very similar (99.5% identity) to a sequence retrieved from the Atlantic Ocean (OTU SPIT34 in Lehours and Jeanthon (2015), GenBank accession number KM654597) and ASV1102 is identical to an uncultured bacterium found in the East coast of Tasmania (Biboni-Isaksson et al., 2016). This ASV appears to be associated to low water temperature (correlation with temperature, $N = 113$, $R = -0.40$, $P < 0.001$) and higher concentrations of nitrate ($N = 89$, NO_3^-, $R = 0.47$, $P < 0.001$). Finally, Gamma-Burkholderiales representatives were scarce along the dataset (only 11 sequences with very low abundances) as expected, since this group is mostly absent in the marine environment (Ferrera et al., 2014; Biboni-Isaksson et al., 2016; Lehours et al., 2018; Auladell et al., 2019).

The taxonomic composition hitherto described here is based on the relative abundances of ASVs. In order to obtain data on the absolute abundance of AAP bacteria, we quantified them by epifluorescence microscopy. Unfortunately, we were not able to quantify AAP abundance along the entire transect, but only in a subset of 21 stations (samples for other stations were either not available or of insufficient quality). Yet, the stations for which the abundance was quantified were uniformly distributed along the transect (except for the Indian Ocean for which samples were not available) and should provide a good representation of the abundance variation along the tropical and subtropical oceans. Abundances ranged between $5.52 \cdot 10^2$ and $6.2 \cdot 10^4$ cells· ml^{-1} and the percentage of AAP bacteria within the prokaryotic community varied between 0.1% and 10% (Fig. S7). Although we estimated AAP abundances for a subgroup of samples, their absolute and relative cell abundances are in line with the abundances reported in previous studies using the same methodology (see data reviewed in Koblížek, 2015). We observed higher AAP bacteria concentrations at lower latitudes (correlation between latitude and %AAP, $N = 21$, $R = 0.50$, $P = 0.024$, Fig. S7), and interestingly, we did not find any relationship between the abundance of AAP bacteria and the taxonomic composition of the AAP communities (Fig. S7). This observation indicates that despite several communities were dominated by different ASVs, there was not a single dominant taxonomic group associated to the increase in absolute AAP bacterial abundances.

ASVs displaying bimodal and lognormal spatial abundance distributions (SpADs) dominate AAP assemblages

We explored the spatial patterns of AAPs and found that most of the individual taxa (64%) were only found in one oceanic region, and these sequences represented only around 10% of the total number of reads. On the contrary, very few sequences (30 ASVs) appeared in all sampled areas, and they represented almost 50% of the total number of reads. Within this group of prevalent sequences, we found representatives of all the taxonomic groups defined above (data not shown), and thus, dominance or rarity of individual sequences does not seem to be linked to taxonomy. For this reason, to better understand the ecological behaviour of AAP taxa, we went beyond their taxonomic affiliations by analyzing the Spatial Abundance Distribution (SpAD) of the individual taxa, an approach that has proven as a useful tool for identifying groups of bacteria sharing similar spatial patterns regardless of their identity (Niño-García et al., 2016; Ruiz-González et al., 2019). In particular, the SpADs analysis classifies individual taxa into different categories according to the shape of their abundance distribution (see Experimental Procedures). The different shapes can be interpreted as ecological traits because the abundance distribution of a given taxon will be the result of the combination of its physiological capacities, environmental tolerances or ability to persist under unfavourable conditions, but also of the external factors controlling its abundance. This approach has been previously used to explore the mechanisms behind the ubiquity or rarity of taxa within aquatic prokaryotic or picoeukaryotic communities (Niño-García et al., 2016; Mangot et al., 2018; Ruiz-González et al., 2019; LaBrie et al., 2021), but to our knowledge, this is the first time that it is restrictively applied to a functional group.

We only found two ASVs displaying normal-like distributions presenting high abundances and broad environmental tolerances (Fig. S8A,C); the bimodal category ($N = 15$ ASVs) included ASVs with lower average abundances and occurrence, likely representing less generalist taxa whose presence is restricted to specific areas, while lognormal ($N = 228$) and logistic ($N = 872$) distributions, which represented the majority of cases, were characteristic of globally rare and endemic AAPs (Fig. S8). AAP assemblages in the surface ocean were dominated by bimodal and lognormal ASVs (Fig. 3C), mostly associated to the Gamma-Halieaceae, Alpha-Rhodobacterales and Alpha-Sphingomonadales groups (Fig. S8), and only few communities were dominated by either normal-like or logistic taxa. The two normal-like ASVs were Sphingomonadales-like (Fig. S8), suggesting large environmental tolerances for this category, regardless of its relatively low contribution in most stations (Fig. 2B).

Communities dominated by logistic ASVs in our study appeared spatially clustered and coincided with productive

regions such as the Benguela coastal province in the South Atlantic, the Caribbean Sea, the Equatorial Pacific and the station nearest to the Strait of Gibraltar (Fig. 3C). In fact, the relative abundances of logistic ASVs showed a significant positive correlation with the mean Chla concentration across stations (N = 107, *R* = 0.43, *P* < 0.0001), pointing to local selection of globally-rare opportunistic AAP bacteria in nutrient-rich areas, as shown for prokaryotic and picoeukaryotic bloomers (Ruiz-González *et al*., 2019; Logares *et al*., 2020).

Yet, the overall distribution of SpADs in our study differs from that reported by Ruiz-González *et al*. (2019) for the whole prokaryotic communities from the same Malaspina

Circumnavigation Expedition surface samples. Whereas bimodal and lognormal ASVs were prevalent in AAP communities, bulk prokaryotic assemblages were dominated by a few cosmopolitan normal-like OTUs. For the bulk community, bimodal and logistic OTUs increased in stations with anomalies in temperature and productivity with respect to the average values. This different distribution suggests that AAP bacteria are less homogeneously distributed than the bulk bacterioplankton (or at least than their dominant members) and that changes in the environment have a large effect on the AAP communities, promoting larger compositional shifts across environmental gradients and the increase of habitat specialists within this functional group.
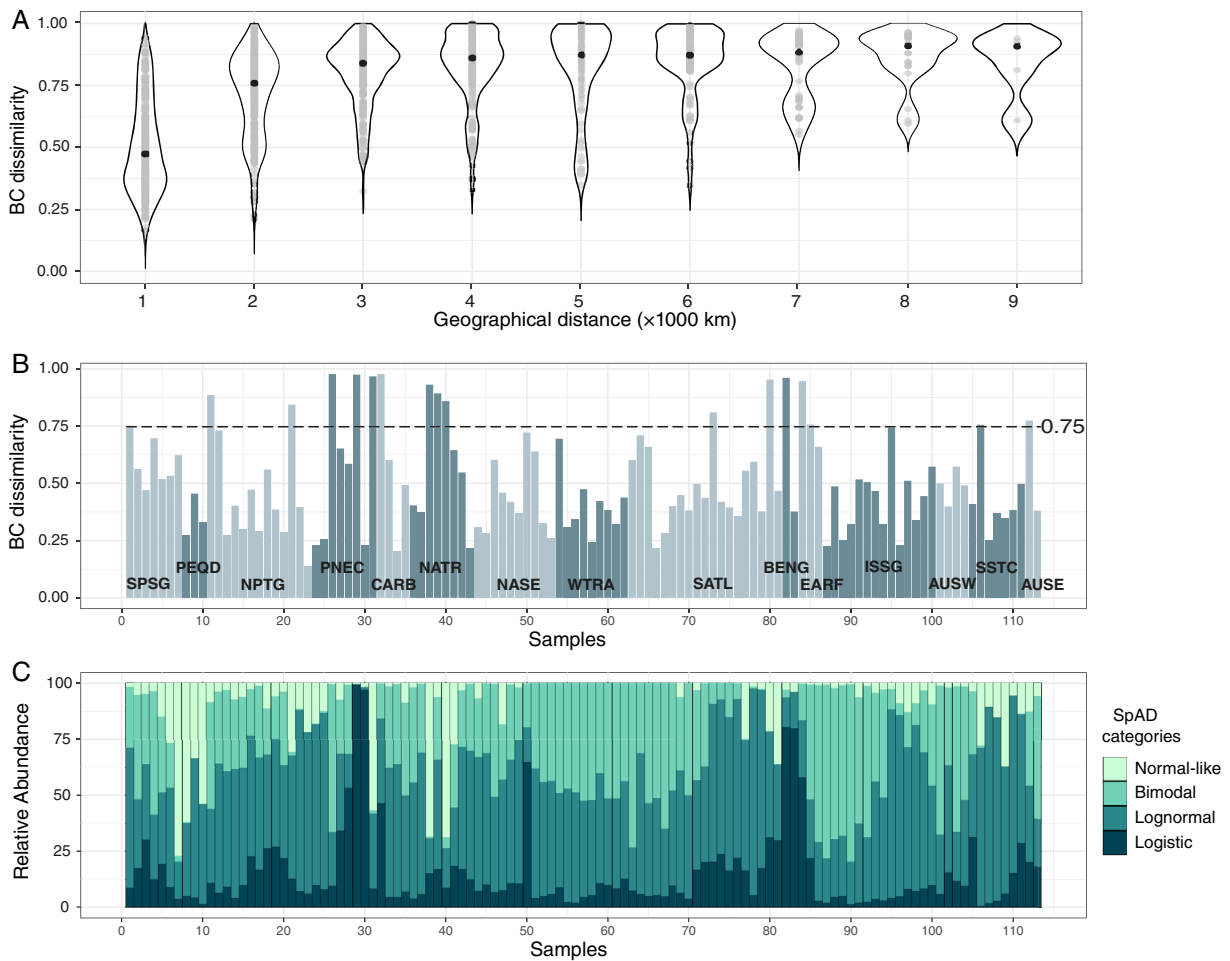

Fig. 3. A. Changes in community dissimilarity between AAP assemblages, measured as Bray–Curtis (BC) dissimilarity with regard to the geographical distance among samples. All comparisons are represented by grey dots while black dots indicate the median value of dissimilarity at each distance. We only considered pairwise comparisons between samples located in the same ocean.
B. Sequential change in community composition across space (sequential beta diversity). Bars represent BC dissimilarity between each community and the one sampled immediately before (e.g., the first bar represents BC dissimilarity between samples 113 and 1, the second bar represents BC dissimilarity between samples 1 and 2, and so on, up to samples 112 and 113). Samples are ordered following the cruise path as in Fig. 2 for comparison. Alternating light and dark color represent a change in Longhurst provinces along the transect and the provinces are indicated according to Longhurst (2007) abbreviations.
C. Relative contribution in each community of the four Spatial Abundance Distributions (SpADs) categories of ASVs throughout the surface ocean, displayed using the same sample order as in panel B.
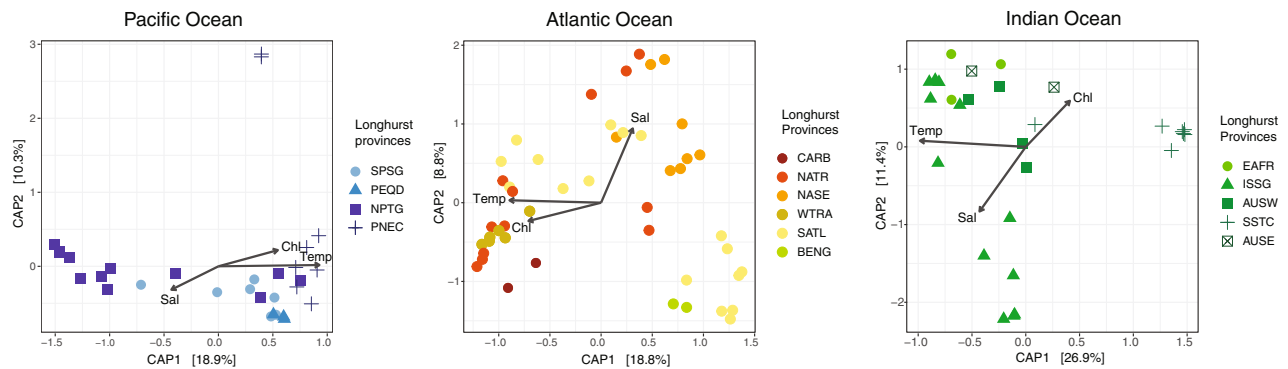
Fig. 4. Distance-based redundancy analysis (dbRDA) performed separately for the Pacific, Atlantic and Indian Ocean stations. Samples are color-coded according to the Longhurst provinces to which they belong. Temperature, salinity and chlorophyll *a* were the three variables that explained the largest fraction of community variance, and they are represented by arrows, where 'Temp' is temperature, 'Sal' is salinity, and 'Chl' is chlorophyll *a* concentration.

Environmental setting drives marked differences in community structure among oceanic regions

We further analyzed the AAP community structure along the Malaspina transect using Bray–Curtis dissimilarity metrics. The overall Bray–Curtis dissimilarity (mean 0.85 ± 0.15) was significantly higher than that described for prokaryotic and picoeukaryotic assemblages in the same transect (prokaryotes mean = 0.61 ± 0.19; picoeukaryotes mean = 0.74 ± 0.08, Logares *et al*., 2020), meaning that changes in the species composition and abundance distributions across AAP communities are larger than across bulk microbial groups. The higher beta diversity observed is consistent with these results, showing that AAP communities are mainly composed of habitat specialists (ASVs with a bimodal distribution) and rare taxa (lognormal distribution), while bulk prokaryotic communities were dominated by few abundant and ubiquitous species (Ruiz-González *et al*., 2019).

Moreover, we explored which abiotic and biotic variables influenced AAP community structure across the global ocean through PERMANOVA ($P < 0.001$), and temperature, salinity and Chla emerged as the most important variables (Table S2). When we pulled all samples together in a distance-based redundancy analysis (dbRDA, Fig. S9), the first two axis explained only 16% of the variation, and there was no obvious clustering of communities based on region or province, even though simple analyses of variance showed statistical differences ($P = 0.001$, Table S3). Thus, we further analyzed the samples for each ocean separately (Fig. 4). Higher percentages of variation were explained by the two first axis (28.9%, Pacific Ocean, 29.1%, Atlantic Ocean and 40.2%, Indian Ocean), and the main variables associated were temperature (with the first axis) and salinity (with the second axis) (Figs. 4 and S10). Stations from the same Longhurst province clustered together in most cases and, in general, we observed that communities from adjacent locations where more similar to each other than communities from distant stations, pointing to gradual changes

in community structure along areas of the surface ocean (Fig. 4). Previous studies restricted to specific areas of different ocean basins observed that the composition of AAP bacteria varied with the trophic conditions (Jiao *et al*., 2007; Yutin *et al*., 2007), while studies from the Arctic Sea showed that the hydrological context of the water masses was also relevant (Boeuf *et al*., 2013; Lehours and Jeanthon, 2015). Our results indicate that temperature, salinity and the general environmental context (as defined by the Longhurst provinces) largely structure AAP surface communities in the global tropical and subtropical ocean.

Community dissimilarity increases with increasing geographic distance

To visualize the turnover of AAP communities along the Malaspina track, we plotted taxonomic community dissimilarities versus geographic distance (Fig. 3A) – considering only pairwise comparisons within the same ocean – which unveiled a strong pattern of biogeography, that is, a remarkable increase of community dissimilarity with increasing distance within each ocean. To further explore community turnover at a fine scale, we explored the sequential changes of beta diversity across the whole sampling transect and found 17 stations displaying Bray-Curtis (BC) dissimilarity values > 0.75 which can be interpreted as sites of abrupt changes in community structure (Fig. 3B). In general, the pattern of sequential beta diversity followed the changes shown through the SpADs analysis (Fig. 3B and C). Stations showing the highest dissimilarity values (BC > 0.9) were located in the South African Atlantic Coast (BENG and EARF) and the Costa Rica Dome (PNEC), where some sequences – belonging to logistic ASVs – presented remarkably high relative abundances, associated to an increase in Chla concentration (see above). Other sites (BC values 0.75–0.9) were in the borders of several Longhurst provinces, such as the South

Subtropical Convergence (SSTC), the Pacific Equatorial Divergence (PEQD), the North Atlantic tropical gyre (NATR) or provinces in the South Atlantic (SATL, BENG and EARF) (Fig. 3B). The partition of the surface ocean into biogeographical provinces was proposed by Longhurst (2007) based on changes in environmental oceanic variables and their annual dynamics. This subdivision has been extensively used in several studies analyzing the surface ocean microbiota and has been proven to explain their biogeographic structure (see for example Friedline *et al.*, 2012; Frank *et al.*, 2016; Milici *et al.*, 2016; Logares *et al.*, 2020; Ruiz-González *et al.*, 2020). We indeed observed that different Longhurst provinces harboured distinct AAP communities but it should also be considered that the borders between these provinces are dynamic and change seasonally (Reygondeau *et al.*, 2013). For example, during the boreal summer, the Northwest Atlantic subtropical gyre (NASW, not included in this sampling) and North Atlantic tropical gyre (NATR) provinces tend to become mixed and an infiltration from the NASW province into the NATR province has clearly been observed (Fig. 4 in Reygondeau *et al.*, 2013). In this same area (NATR province) and timing (during June and July), we observed two stations (133 and 135) that differed largely from the rest, as seen by their different taxonomic composition (Leg 7 in Fig. 2) and high BC sequential dissimilarities (Fig. 3B). This difference could not be attributed to any measured environmental variable. Although this is speculative, the infiltration of water from a different province or some other physical oceanographic feature (Baltar *et al.*, 2010, 2016; Bagnaro *et al.*, 2020) could explain the abrupt changes seen in the North Atlantic in our study. Overall, we observed that AAP communities displayed strong biogeographic patterns, with large dissimilarities across the surface ocean which surpassed in magnitude those described for the bulk surface ocean microbiota.

Selection has a prominent role in structuring AAP communities

Our analyses of AAP community turnover clearly showed a biogeographical pattern across the surface ocean. The different patterns of diversity and species composition across spatial scales result from the combination of different ecological processes, such as selection, dispersal, or drift (Vellend, 2016). Changes in microbial species composition across space could be related to selection processes driven by changes in environmental variables (Fig. 4). Nevertheless, we observed that environmental conditions at adjacent stations were generally comparable, so these changes could also arise from dispersal limitation imposed by physical oceanographic features (Baltar *et al.*, 2010, 2016, Bagnaro *et al.*, 2020). In fact,

previous studies have shown that oceanic features such as boundaries between different ocean regions can act as strong barriers and delimit the distribution of microbes in the ocean (Baltar *et al.*, 2016; Raes *et al.*, 2018).

Whether the pattern observed is the result of environmental selection and/or dispersal limitation cannot be determined based on our previous analysis (see also Hanson *et al.*, 2012). Thus, to further investigate the ecological processes shaping AAP communities across the global surface ocean, we applied the approach proposed by Stegen *et al.* (2013), which quantitatively estimates the influence of selection, dispersal and drift based on the phylogenetic turnover of communities. Since this method relies solely on the phylogeny of the *puf*M gene and on null models (randomization), it avoids the problem of unmeasured environmental variables that can potentially be associated with selection or dispersal (Stegen *et al.*, 2013). The influence of selection was estimated using the β-nearest taxon index (βNTI), which is the difference between the observed phylogenetic turnover for a given pair of communities and the null model after 999 randomizations (see Experimental Procedures). The values of βNTI were calculated for all the pairwise community comparisons possible in the dataset. We found that approximately 23% of the pairwise comparisons had values of $\beta NTI < -2$, which implies that there is a shorter phylogenetic distance within these pairs of communities, than expected by chance (Stegen *et al.*, 2012). Lower turnover between communities is expected when environmental conditions are very similar and there is a homogeneous selection of closely related taxa in these communities. Likewise, approximately 27% of the pairwise comparisons had $\beta NTI > 2$, which is associated with a greater phylogenetic distance than the expected under a null model and can be interpreted as different environmental conditions heterogeneously selecting distantly related taxa (Stegen *et al.*, 2012). Overall, approximately 50% of the observed turnover could be explained by selection, with homogeneous and heterogeneous selection being almost equally important at a global scale (Fig. 5). Within samples located in the same Longhurst province, homogeneous selection had an important role, as the main ecological process in most provinces (Fig. S11). In turn, heterogeneous selection had a modest role within Longhurst provinces, and only operated in some provinces. Based on βNTI values of comparisons between adjacent stations, heterogeneous selection was high in areas where logistic ASVs dominated (data not shown), pointing towards the selection of rare taxa in productive areas. These results are in line with previous studies that already pointed to selection as a major ecological process structuring AAP communities in both spatial (Lehours *et al.*, 2018) and temporal studies (Auladell *et al.*, 2019).
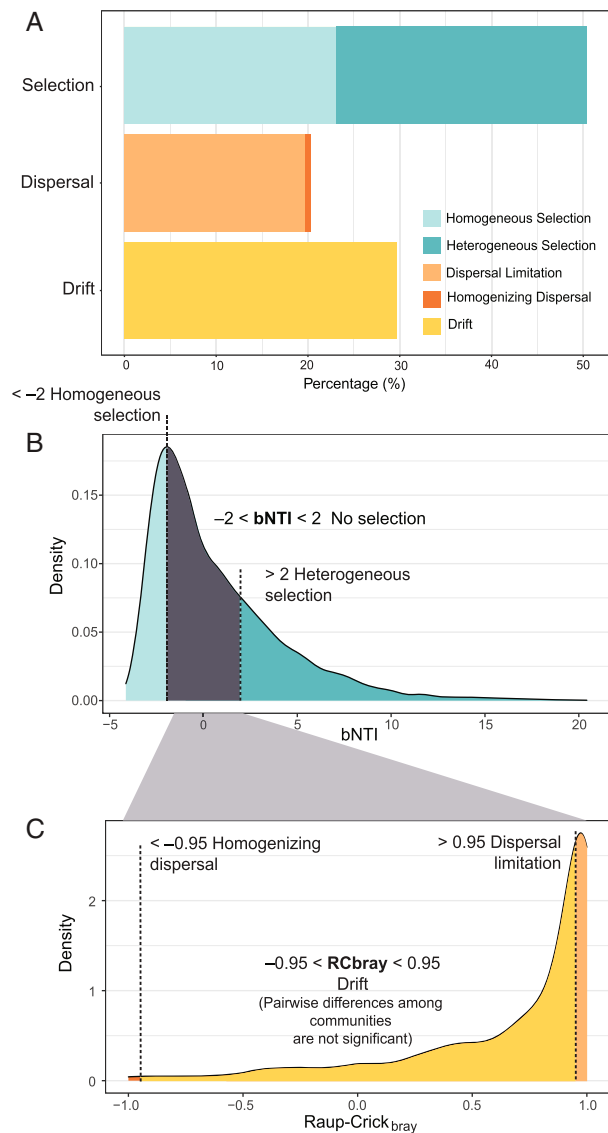
Fig. 5. A. Percentage of the AAP bacterial community turnover associated to each ecological process in the tropical and subtropical surface ocean.
B. Distribution of βNTI estimates for the total number of comparisons between all samples in the dataset. Absolute values of βNTI above two are considered as significant departures from random phylogenetic turnover and are associated to homogeneous and heterogeneous selection (blue areas). The grey area represents the fraction of nonsignificant βNTI values. To disentangle whether drift or dispersal are the main ecological processes shaping the turnover between these communities, Bray–Curtis-based Raup–Crick (RC$_{bray}$) was calculated.
C. Distribution of RC$_{bray}$ for the pairwise community comparisons that are not structured by selection. RC$_{bray}$ values between −0.95 and + 0.95 point to a community assembly governed by drift (yellow area). On the contrary, RC$_{bray}$ values > +0.95 or < −0.95 indicate that community turnover is driven by dispersal limitation or homogenizing dispersal respectively (orange areas).

For the remaining pairwise comparisons, the value of Bray–Curtis-based Raup–Crick (RC$_{bray}$) characterized the magnitude of deviation between the observed BC

and the null BC. RC$_{bray}$ distribution varied between −1 and 1, and only values |RC$_{bray}$| > 0.95 were considered as significant departures from drift (see Experimental Procedures and Fig. 5). Dispersal limitation explained approximately 20% of the community turnover (RCbray > 0.95) while homogenizing dispersal was observed only for 18 pairwise comparisons (0.7%). The limited role of dispersal limitation structuring AAP communities could be expected, since distant microbial communities are known to be connected on a global scale, under what has been described as the 'Microbial Conveyor Belt' (Mestre and Höfer, 2021). Finally, almost 30% of the community turnover was explained by drift (stochastic processes), as the differences between the null and the observed beta diversity were not significant. Stochastic processes are difficult to predict and to distinguish from other ecological processes (Zhou and Ning, 2017); however, they play an important role in microbial community assembly (Evans et al., 2017, Graham and Stegen, 2017) and their importance increases under high selection and low dispersal (Fodelianakis et al., 2021), as it happens in AAP communities across the surface ocean.

Remarkably, the observed pattern is different from that reported for whole prokaryotic communities along the same transect (Logares et al., 2020), which appeared to be structured to a similar extent by both selection and dispersal (representing each process approximately 25% of the community turnover). In contrast, dispersal limitation had a much more important role in structuring picoeukaryotic communities (approximately 65%), likely due to their larger cell sizes and lower abundances (Logares et al., 2020). The relatively higher importance of selection mechanisms in AAP communities suggests that AAP bacteria are more affected by small changes in environmental conditions than the prokaryotic community as a whole. As we have shown above, the community turnover measured as Bray–Curtis dissimilarity is higher in this functional group than in the bulk picoplankton, pointing to higher changes in the composition and structure of AAP communities over short distances. Besides, while prokaryotic assemblages are dominated by few cosmopolitan and very abundant taxa, AAP assemblages are mainly composed by taxa classified as rare or habitat specialists, with more restricted environmental tolerances.

Concluding remarks

In this study, we describe the global diversity and community structure patterns of marine AAP bacteria in the tropical and subtropical oceans. Alpha diversity varied across biogeographical provinces mainly related to temperature, salinity and trophic status and showed remarkably low values in the more productive Longhurst

provinces. AAP communities along the surface ocean were mainly composed of members of the Halieaceae (Gammaproteobacteria), which were adapted to a large range of environmental conditions, and by different clades of the Alphaproteobacteria, that seemed to dominate under particular circumstances, such as in the oligotrophic gyres. These taxa were not randomly distributed but appeared to be spatially structured, with a marked succession of samples dominated by either one or the other class. Communities from adjacent stations shared more taxonomic similarities, that is, community dissimilarity increased with increasing distance, which resulted in a remarkable biogeographical pattern. However, this pattern was to a large extent the result of homogeneous and heterogeneous selection of individual taxa, while dispersal and drift had less of a role in shaping the structure of AAP bacterial communities. While the seasonal patterns of AAPs have been shown to be notably comparable to those of the bulk bacterioplankton, at a large-scale, AAP communities seem to have their own spatial patterns that do not mimic those of the bulk picoplankton. Of the measured environmental variables, temperature, salinity and Chla were found to influence AAP community structure. Small changes in environmental conditions translated into significant changes in AAP communities, and therefore, several habitat specialists and many rare species dominated their communities. The photoheterotrophic metabolism, high growth rates and high predation pressure on AAP bacteria, among other attributable traits to this functional group, could explain the stronger role of selection for this group as compared to the bulk surface ocean microbiota. Overall, our results represent the most comprehensive study investigating the global biogeography of AAP communities and shows how different ecological processes explain these patterns.

Experimental procedures

Sample collection

The Malaspina Circumnavigation Expedition took place between December 2010 and July 2011 (Duarte, 2015). Samples were collected in 113 stations across the tropical and subtropical waters of the Pacific, Atlantic and Indian Oceans. At each station, about 12 L of surface seawater (3 m depth) were collected with large (30 L) oceanographic bottles. Simultaneously, a CTD profiler was used to profile temperature, salinity, conductivity, fluorescence and dissolved oxygen. Seawater was pre-filtered through a 20 μm nylon mesh and a 3 μm filter onto a 0.2 μm Millipore polycarbonate filter. Samples were conserved at −80°C until further processing. Samples for enumerating AAP cells were pre-filtered through

a 200 μm mesh and 10 ml of each sample were filtered onto 0.2 μm polycarbonate filters. Cells were enumerated by infra-red epifluorescence microscopy in 21 stations as described in Ferrera *et al*. (2014). The environmental biotic and abiotic parameters used in this study were determined as reported in Estrada *et al*. (2016) and Ruiz-González *et al*. (2019).

DNA extraction, pufM amplification, sequencing and ASV generation

DNA was extracted from the 0.2 μm filter using the phenol–chloroform protocol as described in Massana *et al*., (1997). Partial amplification of the *puf*M gene (~245 bp fragment) was performed in 50 μl reactions using primers *puf*M forward (5′-TACGGSAACC TGTWCTAC-3′, (Béjà *et al*., 2002) and pufM_WAW reverse (5′-AYNGCRAACCACCANGCCCA- 3′ (Yutin *et al*., 2005)) as described in Auladell *et al*. (2019). DNA was sequenced in an Illumina MiSeq sequencer (2 × 250 bp, Research and Testing Laboratory; http://rtlgenomics.com/). After sequencing, we used cutadapt v1.16 (Martin, 2013) to remove primers and DADA2 v1.10 (Callahan *et al*., 2016) to infer amplicon sequence variants (ASV) with the following parameters: maxEE = c (2,6) and trunclen = c(210,150). After filtering chimeras and spurious sequences, we kept 82% of the initial number of reads (mean 24,173, min 4503, max 79,968). To be able to compare our data with previous studies that used OTUs, we clustered the ASVs with UCLUST v10.0 (Edgar, 2010) at 94% similarity, the threshold usually employed for the *puf*M gene (Zeng *et al*., 2007).

Phylogenetic classification

We used phylogenetic placement for predicting the taxonomic assignment of the *puf*M gene short sequences. Due to the lack of comprehensive public databases for AAP bacteria, we built a custom made *puf*M database retrieving more than 750 sequences longer than 600 bp from the GTDB and GenBank, as well as from metagenomic datasets from the Tara *Oceans* Expedition (Sunagawa *et al*., 2015), the Malaspina Expedition (unpublished), the Global Ocean Survey (GOS, Yutin *et al*., 2007; Cuadrat *et al*., 2016), and the Blanes Bay Microbial Observatory (BBMO, Auladell *et al*., 2019). The alignment was done using the *Decipher* R package (Wright, 2016) and MAFFT v7 (Katoh and Standley, 2013). After a manual curation using AliView v1.26 (Larsson, 2014), we kept 673 sequences. A phylogenetic tree was constructed using RAxML v8.2 (Stamakis, 2014) (GTRGAMMA model, 100 bootstraps), and visualized using iTOL (Letunic and Bork, 2019),

(Fig. S5). Finally, to infer the phylogeny of the amplicon sequence variants, we applied the Evolutionary Placement Algorithm v0.3.5 (Barbera *et al*., 2019).

Data analyses

All statistical analyses were performed using R v3.6.3 (R Core Team 2021). The ASV table was rarefied down to 4500 reads per sample using the *vegan* package. Alpha diversity was estimated using Chao1 and Shannon diversity indices (Chao and Lee, 1992), with the *phyloseq* package. Post-Hoc Tukey tests were employed to see if there were statistically significant differences between the diversity of different regions. To test whether diversity was influenced by environmental conditions, we performed Pearson correlations between a selection of environmental variables and the diversity indices and applied false discovery rate (FDR)-correction to all *P*-values. We also compared the diversity of AAP bacteria with those of the bulk prokaryotic communities using the 16S rRNA gene data presented in Ruiz-González *et al*. (2019) and the picoeukaryotic community data (18S rRNA gene) presented in Logares *et al*. (2020), both from the same samples taken during the Malaspina Circumnavigation Expedition. Community composition was analyzed and described using the *phyloseq* package in R.

In order to explore the spatial patterns of individual AAP bacteria across space, we analyzed the abundance distribution of each *puf*M sequence across all samples. We used the rarefied table of reads ($\log_{10}(x + 1)$ transformed) to select the statistical distribution that best fitted the spatial abundance distribution (SpAD) of each ASV, as described in Niño-García *et al*. (2016) and Ruiz-González *et al*. (2019). We could classify all ASVs into four SpAD categories: 'normal-like' ASVs showed a normal statistical distribution, which has previously been associated with globally abundant and widespread taxa, and which might represent habitat generalists (Niño-García *et al*., 2016; Ruiz-González *et al*., 2019). The distribution of 'bimodal' ASVs is characterized by two density peaks, with the first one commonly corresponding to zero cases. This group could be considered less generalists because they are only detected in certain regions and their average abundances are also lower. Finally, ASVs classified as 'logistic' and 'lognormal' present a distribution with a zero-abundance mode, and they have been shown to comprise mostly rare sequences (for more details on the analysis see Niño-García *et al*., 2016 and Ruiz-González *et al*., 2019). For each SpAD category, we calculated the mean relative abundance of the constituent ASVs across all stations, as well as their occurrence. We also estimated the individual environmental breath as the range of temperature, salinity, Chla, and dissolved oxygen concentration in which each of the ASVs within the different categories were detected.

The exploration of the main environmental drivers explaining the structure of AAP communities was done using a Bray–Curtis dissimilarity matrix, built with the *vegdist* function from the *vegan* package and visualized in a distance-base redundancy analysis (dbRDA), with a previous selection of significant environmental variables (PERMANOVA *P* < 0.01). Permutation tests (*adonis* function from *vegan* package) were employed to examine community differences among the six oceanic regions (South Pacific, North Pacific, North Atlantic, South Atlantic, Indian and South Australian Bight) and Longhurst oceanographic provinces (Longhurst, 2007). We used Mantel tests (1000 permutations) to compare the changes in the structure of AAP communities between stations with differences in temperature, salinity and Chla. In addition, we performed partial Mantel tests to compare the community structures of AAP, prokaryotes and picoeukaryotes, removing the effect of temperature, salinity and Chla. The Bray–Curtis dissimilarity matrix was also used to analyze the spatial community structure turnover, and to explore sequential changes along the Malaspina transect, by comparing each sample with the one sampled immediately before.

Finally, to quantify the relative importance of selection, dispersal and drift as processes structuring the communities of AAP bacteria, we followed the framework developed by Stegen *et al*. (2013). This approach assumes that there is a phylogenetic signal (Cavender-Bares *et al*. 2009) in the ASVs optimal habitat conditions (i.e., the habitat preferences of closely related taxa are more similar than the preferences of distantly related taxa). To confirm this assumption, we firstly compared ASVs niche distances (using temperature, salinity and Chla) and ASVs phylogenetic distances using a Mantel correlogram test. We detected phylogenetic signal in the *puf*M gene marker over relatively short phylogenetic distances (Fig. S12), as previously shown with other marker genes (e.g., Stegen *et al*., 2013; Dini-Andreote *et al*., 2015; Huber *et al*., 2020; Logares *et al*., 2020).

Then, to analyze the influence of selection, we calculated the β-mean nearest taxon distance (βMNT) metric, which quantifies the mean phylogenetic distances between two communities, and compared them to a random expectation (999 randomizations). The difference between the observed phylogenetic turnover (or βMNT) and the values obtained with the null model are denoted as β-nearest taxon index (βNTI). Absolute βNTI values above 2 (|βNTI| > 2) indicate that coexisting taxa are more closely related than expected by chance; thus, pointing to the action of selection. Afterwards, to differentiate whether drift or dispersal were the main structuring processes, we calculated the Raup–Crick metric (Chase

et al., 2011) using Bray–Curtis dissimilarities (RC$_{bray}$) (Chase *et al*., 2011; Stegen *et al*., 2013). RC$_{bray}$ compares the measured beta diversity to the beta diversity obtained by the null model (999 randomizations) that would be obtained under random community assembly (drift). RC$_{bray}$ values between -0.95 and $+0.95$ point to a community assembly governed by drift. On the contrary, RC$_{bray}$ values ≥ 0.95 or ≤ 0.95 indicate that community turnover is driven by dispersal limitation or homogenizing dispersal, respectively (Stegen *et al*., 2013). For this analysis, raw ASV sequences were aligned with AliView v1.26 (Larsson, 2014), aligned sequences were visually curated with Seaview (Gouy *et al*., 2010) and the phylogenic tree was constructed using FastTree v2.1.9 (Price *et al*., 2010). The βMNTD and βNTI metrics were calculated using the R package *Picante* (Kembel *et al*., 2010) and the RC$_{bray}$ was calculated with the *raup_crick_abundance* function following Stegen *et al*. (2013). These analyses were performed in R v3.6.3 (R Core Team 2021) and codes are available in Github (https://gitlab.com/crgazulla/malaspina_aaps). Sequence data have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA736051.

Acknowledgements

References

Auladell, A., Sánchez, P., Sánchez, O., Gasol, J.M., and Ferrera, I. (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME J* **13**: 1975–1987. https://doi.org/10.1038/s41396-019-0401-4

Bagnaro, A., Baltar, F., Brownstein, G., Lee, W.G., Morales, S.E., Pritchard, D.W., and Hepburn, C.D. (2020) Reducing the arbitrary: fuzzy detection of microbial ecotones and ecosystems – focus on the pelagic environment. *Environmental Microbiome* **15**: 16. https://doi.org/10.1186/s40793-020-00363-w.

Baltar, F., Arístegui, J., Gasol, J., Lekunberri, I., and Herndl, G.J. (2010) Mesoscale eddies: hotspots of prokaryotic activity and differential community structure in the ocean. *ISME J* **4**: 975–988. https://doi.org/10.1038/ismej.2010.33.

Baltar, F., Currie, K., Stuck, E., Roosa, S., and Morales, S.E. (2016) Oceanic fronts: transition zones for bacterioplankton community composition. *Environ Microbiol Rep* **8**: 132–138. https://doi.org/10.1111/1758-2229.12362

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019) EPA-ng: massively parallel evolutionary placement of genetic sequences. *Syst Biol* **68**: 365–369. https://doi.org/10.1093/sysbio/syy054

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al*. (2000) Bacterial rhodopsin: evidence for a new type of Phototrophy in the sea. *Science* **289**: 1902–1906. https://doi.org/10.1126/science.289.5486.1902

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., *et al*. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**: 630–633. https://doi.org/10.1038/415630a

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., and Brown, M.V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the East coast of Australia. *Environ Microbiol* **18**: 4485–4500. https://doi.org/10.1111/1462-2920.13436

Boeuf, D., Cottrell, M.T., Kirchman, D.L., Lebaron, P., and Jeanthon, C. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* **85**: 417–432. https://doi.org/10.1111/1574-6941.12130

Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583. https://doi.org/10.1038/nmeth.3869.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A., and Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* **12**: 693–715. https://doi.org/10.1111/j.1461-0248.2009.01314.x

Chao, A., and Lee, S.M. (1992) Estimating the number of classes via sample coverage. *J Am Stat Assoc* **87**: 210–217. https://doi.org/10.1080/01621459.1992.10475194

Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., and Inouye, B.D. (2011) Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* **2**: 1–11. https://doi.org/10.1890/ES10-00117.1.

Cottrell, M.T., and Kirchman, D.L. (2009) Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl*

Environ Microbiol **75**: 4958–4966. https://doi.org/10.1128/AEM.00117-09

Cuadrat, R., Ferrera, I., Grossart, H., and Dávila, A. (2016) Picoplankton bloom in global south? a high fraction of aerobic anoxygenic phototrophic bacteria in metagenomes from a coastal bay (Arraial do Cabo--Brazil). *Omi A J Integr Biol* **20**: 76–87. https://doi.org/10.1089/omi.2015.0142

DeLong, E.F., and Béjà, O. (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8**: 1–5. https://doi.org/10.1371/journal.pbio.1000359

Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., and Salles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *PNAS* **112**: E1326–E1332. https://doi.org/10.1073/pnas.1414261112.

Duarte, C.M. (2015) Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin* **24**: 11–14. https://doi.org/10.1002/lob.10008

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Estrada, M., Delgado, M., Blasco, D., Latasa, M., Cabello, A.M., Benítez-Barrios, V., *et al.* (2016) Phytoplankton across tropical and subtropical regions of the Atlantic, Indian and Pacific oceans. *PLoS One* **11**: 1–29. https://doi.org/10.1371/journal.pone.0151699

Evans, S., Martiny, J., and Allison, S. (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME* **11**: 176–185. https://doi.org/10.1038/ismej.2016.96.

Ferrera, I., Borrego, C.M., Salazar, G., and Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol* **16**: 2953–2965. https://doi.org/10.1038/s41598-018-22413-7

Ferrera, I., Gasol, J.M., Sebastián, M., Hojerova, E., and Koblížek, M. (2011) Comparison of growth rates of aerobic Anoxygenic phototrophic bacteria and other Bacterioplankton groups in coastal Mediterranean waters. *Appl Environ Microbiol* **77**: 7451–7458. https://doi.org/10.1128/AEM.00208-11

Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M., and Gasol, J.M. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME J* **11**: 2391–2393. https://doi.org/10.1038/ismej.2017.79

Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A., and Daffonchio, D. (2021) Direct quantification of ecological drift at the population level in synthetic bacterial communities. *ISME J* **15**: 55–66. https://doi.org/10.1038/s41396-020-00754-4.

Frank, A.H., Garcia, J.A.L., Herndl, G.J., and Reinthaler, T. (2016) Connectivity between surface and deep waters determines prokaryotic diversity in the North Atlantic deep water. *Environ Microbiol* **18**: 2052–2063. https://doi.org/10.1111/1462-2920.13237

Friedline, C.J., Franklin, R.B., McCallister, S.L., and Rivera, M.C. (2012) Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. *Biogeosciences* **9**: 2177–2193. https://doi.org/10.5194/bg-9-2177-2012

Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221–224. https://doi.org/10.1093/molbev/msp259.

Graham, E.D., Heidelberg, J.F., and Tully, B.J. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* **12**: 1861–1866. https://doi.org/10.1038/s41396-018-0091-3

Graham, E.B., and Stegen, J.C. (2017) Dispersal-based microbial community assembly decreases biogeochemical function. *Processes* **5**: 65. https://doi.org/10.3390/pr5040065.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506. https://doi.org/10.1038/nrmicro2795

Hojerová, E., Mašín, M., Brunet, C., Ferrera, I., Gasol, J.M., and Koblížek, M. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol* **13**: 2717–2725. https://doi.org/10.1111/j.1462-2920.2011.02540.x

Huber, P., Metz, S., Unrein, F., Mayora, G., Sarmento, H., and Devercelli, M. (2020) Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. *ISME J* **14**: 2951–2966. https://doi.org/10.1038/s41396-020-0723-2.

Jeanthon, C., Boeuf, D., Dahan, O., le Gall, F., Garczarek, L., Bendif, E.M., and Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* **8**: 1955–1970. https://doi.org/10.5194/bg-8-1955-2011

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., and Wang, P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* **9**: 3091–3099. https://doi.org/10.1111/j.1462-2920.2007.01419.x

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., de Vargas, C., Raes, J., *et al.* (2011) A holistic approach to marine eco-systems biology. *PLoS Biol* **9**: 7–11. https://doi.org/10.1371/journal.pbio.1001177

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780. https://doi.org/10.1093/molbev/mst010

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Kirchman, D.L., and Hanson, T.E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* **5**: 188–199. https://doi.org/10.1111/j.1758-2229.2012.00367.x

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* **39**: 854–870. https://doi.org/10.1093/femsre/fuv032

Koblížek, M., Mašín, M., Ras, J., and Poulton, A.J. (2007) Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ Microbiol* **9**: 2401–2406. https://doi.org/10.1111/j.1462-2920.2007.01354.x

Kolber, Z.S., van Dover, C.L., Niederman, R.A., and Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**: 177–179. https://doi.org/10.1038/35025044

LaBrie, R., Bélanger, S., Benner, R., and Maranger, R. (2021) Spatial abundance distribution of prokaryotes is associated with dissolved organic matter composition and ecosystem function. *Limnol Oceanogr* **66**: 575–587. https://doi.org/10.1002/lno.11624

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**: 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lehours, A., and Jeanthon, C. (2015) The hydrological context determines the beta-diversity of aerobic anoxygenic phototrophic bacteria in European Arctic seas but does not favor endemism. *Front Microbiol* **6**: 638. https://doi.org/10.3389/fmicb.2015.00638

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., and Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* **74**: 397–409. https://doi.org/10.1111/j.1574-6941.2010.00954.x

Lehours, A.C., Enault, F., Boeuf, D., and Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci Rep* **8**: 1–10. https://doi.org/10.1038/s41598-018-22413-7

Letunic, I., and Bork, P. (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**: W256–W259. https://doi.org/10.1093/nar/gkz239.

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., *et al.* (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**: 1–17.

Longhurst, A. (2007) *Ecological Geography of the Sea*. 2nd Edition, San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-455521-1.X5000-1

Mangot, J.F., Forn, I., Obiol, A., and Massana, R. (2018) Constant abundances of ubiquitous uncultured protists in the open sea assessed by automated microscopy. *Environ Microbiol* **20**: 3876–3889. https://doi.org/10.1111/1462-2920.14408

Martin, M. (2013) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10. https://doi.org/10.14806/ej.17.1.200

Mašín, M., Zdun, A., Stoń-Egiert, J., Nausch, M., Labrenz, M., Moulisová, V., and Koblížek, M. (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquat Microb Ecol* **45**: 247–254. https://doi.org/10.3354/ame045247

Massana, R., Murray, A.E., Preston, C.M., & DeLong, EF (1997) Vertical distribution and phylogenetic characterization of marine planktonic. *Microbiology* **63**: 50–56. https://doi.org/10.1128/aem.63.1.50-56.1997

Mena, C., Reglero, P., Balbín, R., Martín, M., Santiago, R., and Sintes, E. (2020) Seasonal niche partitioning of surface temperate Open Ocean prokaryotic communities. *Front Microbiol* **11**: 1749. https://doi.org/10.3389/fmicb.2020.01749.

Mestre, M., and Höfer, J. (2021) The microbial Conveyor Belt: connecting the globe through dispersion and dormancy. *Trends Microbiol* **29**: 482–492. https://doi.org/10.1016/j.tim.2020.10.007.

Milici, M., Tomasch, J., Wos-Oxley, M.L., Decelle, J., Jáuregui, R., Wang, H., *et al.* (2016) Bacterioplankton biogeography of the Atlantic Ocean: a case study of the distance-decay relationship. *Front Microbiol* **7**: 1–15. https://doi.org/10.3389/fmicb.2016.00590

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., *et al.* (2020) A genomic catalog of Earth's microbiomes. *Nat Biotechnol* **39**: 499–509. https://doi.org/10.1038/s41587-020-0718-6.

Niño-García, J.P., Ruiz-González, C., and del Giorgio, P.A. (2016) Landscape-scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. *Ecol Lett* **19**: 1506–1515. https://doi.org/10.1111/ele.12704

Obiol, A., Giner, C.R., Sánchez, P., Duarte, C.M., Acinas, S. G., and Massana, R. (2020) A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Mol Ecol Resour* **20**: 718–731. https://doi.org/10.1111/1755-0998.13147

Oz, A., Sabehi, G., Koblízek, M., Massana, R., and Béjà, O. (2005) Roseobacter-like bacteria in red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl Environ Microbiol* **71**: 344–353. https://doi.org/10.1128/AEM.71.1.344-353.2005

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996–1004. https://doi.org/10.1038/nbt.4229

Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., and Pedrós-Alió, C. (2016) Marine bacterial and Archaeal ion-pumping Rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* **80**: 929–954. https://doi.org/10.1128/MMBR.00003-16

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490. https://doi.org/10.1371/journal.pone.0009490.

R Core Team (2021) *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/

Raes, E., Bodrossy, L., van de Kamp, J., Bissett, A., Ostrowski, M., Brown, M.V., *et al.* (2018) Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *PNAS* **115**: E8266–E8275. https://doi.org/10.1073/pnas.1719335115.

Regaudie-de-Gioux, A., Huete-Ortega, M., Sobrino, C., López-Sandoval, D.C., González, N., Fernández-Carrera, A., *et al.* (2019) Multi-model remote sensing

assessment of primary production in the subtropical gyres. *J Marine Systems* **196**: 97–106. https://doi.org/10.1016/j.jmarsys.2019.03.007.

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., and Maury, O. (2013) Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem Cycles* **27**: 1046–1058. https://doi.org/10.1002/gbc.20089

Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., *et al*. (2019) Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. *Mol Ecol* **28**: 1930–1945. https://doi.org/10.1111/mec.15026

Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M., Salazar, G., Agustí, S., *et al*. (2020) Major imprint of surface plankton on deep ocean prokaryotic structure and activity. *Mol Ecol* **29**: 1820–1838. https://doi.org/10.1111/mec.15454

Salazar, G., Cornejo-Castillo, F.M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., *et al*. (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Mol Ecol* **24**: 5692–5706. https://doi.org/10.1111/mec.13419

Schwalbach, M.S., and Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol Oceanogr* **50**: 620–628. https://doi.org/10.4319/lo.2005.50.2.0620

Sieracki, M.E., Gilg, I.C., Thier, E.C., Poulton, N.J., and Goericke, R. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the Northwest Atlantic. *Limnol Oceanogr* **51**: 38–46.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., *et al*. (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* **7**: 2069–2079. https://doi.org/10.1038/ismej.2013.93

Stegen, J.C., Lin, X., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664. https://doi.org/10.1038/ismej.2012.22

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., *et al*. (2015) Structure and function of the global ocean microbiome. *Science* **348**: 1–10. https://doi.org/10.1126/science.1261359

Teira, E., Logares, R., Gutiérrez-Barral, A., Ferrera, I., Varela, M.M., Morán, X.A.G., and Gasol, J.M. (2019) Impact of grazing, resource availability and light on prokaryotic growth and diversity in the oligotrophic surface global ocean. *Environ Microbiol* **21**: 1482–1496. https://doi.org/10.1111/1462-2920.14581

Tully, B.J., Graham, E.D., and Heidelberg, J.F. (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data* **5**: 1–8. https://doi.org/10.1038/sdata.2017.203

de Vargas, C., Audie, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al*. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605–1/11. https://doi.org/10.1126/science.1261605

Vellend, M. (2016) *The Theory of Ecological Communities*, 1st ed. Woodstock, UK: Princeton University Press.

Villamaña, M., Marañón, E., Cermeño, P., Estrada, M., Fernández-Castro, B., Figueiras, F.G., *et al*. (2019) The role of mixing in controlling resource availability and phytoplankton community composition. *Prog Oceanogr* **178**: 102181. https://doi.org/10.1016/j.pocean.2019.102181

Waidner, L.A., and Kirchman, D.L. (2008) Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene *pufM* in the Delaware estuary. *Appl Environ Microbiol* **74**: 4012–4021. https://doi.org/10.1128/AEM.02324-07

Wright, E.S. (2016) Using DECIPHER v2.0 to analyze big biological sequence data in R. *R Journal* **8**: 352–359. https://doi.org/10.32614/RJ-2016-025

Yutin, N., Suzuki, M.T., and Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol* **71**: 8958–8962. https://doi.org/10.1128/AEM.71.12.8958-8962.2005

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., and Béjà, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific oceans using the Global Ocean sampling expedition metagenomes. *Environ Microbiol* **9**: 1464–1475. https://doi.org/10.1111/j.1462-2920.2007.01265.x

Zeng, Y.H., Chen, X.H., and Jiao, N.Z. (2007) Genetic diversity assessment of anoxygenic photosynthetic bacteria by distance-based grouping analysis of *pufM* sequences. *Lett Appl Microbiol* **45**: 639–645. https://doi.org/10.1111/j.1472-765X.2007.02247.x

Zhou, J., and Ning, D. (2017) Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* **81**: e00002–e00017. https://doi.org/10.1128/MMBR.00002-17.

Supporting Information

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

Appendix S1: Supplementary Information.