Global Ecology and Biogeography

A Mac

Global beta diversity patterns of microbial communities in the surface and deep ocean

Journal:	*Global Ecology and Biogeography*
Manuscript ID	GEB-2021-0125.R3
Manuscript Type:	Research Article
Keywords:	oceanography, microbial ecology, B diversity, community, dispersal, plankton

Global beta diversity patterns of microbial communities in the surface and deep ocean

ABSTRACT

Aim: Dispersal and environmental gradients shape marine microbial communities, yet the relative importance of these factors across taxa with distinct sizes and dispersal capacity in different ocean layers is unknown. Here, we report a comparative analysis of surface and deep ocean microbial beta diversity and examine how these patterns are tied to oceanic distance and environmental gradients.

Methods: Beta diversity was calculated from metabarcoding data on prokaryotic and picoeukaryotic microbes (eukaryotes between 0.2 and 3 μm) collected during the Malaspina expedition across the tropical and subtropical oceans. Mantel correlations were used to determine the relative contribution of environment and oceanic distance in driving community beta diversity.

Location: Tropical and subtropical regions of the ocean (30 °N - 40 °S).

Results: Mean community similarity across all marine sites for prokaryotes was 38.9% in the surface and 51.4% in the deep ocean, compared to mean similarity of 25.8% and 12.1% in the surface and deep ocean, respectively, for picoeukaryotes. Higher dispersal rates and smaller body sizes of prokaryotes relative to picoeukaryotes likely contributed to the significantly higher community similarity for prokaryotes compared with picoeukaryotes. The ecological mechanisms determining the biogeography of microbes varies across depth. In the surface, the environmental differences in space were a more important factor driving microbial distribution compared with the oceanic distance between samples, defined as the shortest path between two sites avoiding land. In the deep ocean, picoeukaryote communities were slightly more structured by the oceanic distance, while prokaryotes were shaped by the combined action of oceanic distance and environmental filtering.

Main conclusions: Horizontal gradients in microbial community assembly differed across ocean depths, as did mechanisms shaping them. In the deep ocean, the oceanic

distance and environment played significant roles driving microbial spatial distribution, while in the surface the influence of the environment was stronger than oceanic distance.

Keywords: oceanography, microbial ecology, beta diversity, community, dispersal, plankton.

INTRODUCTION

Marine microbes play important roles in marine food webs, carbon cycling, and climate (Azam & Malfatti, 2007; Falkowski, Barber, & Smetacek, 1998; Guidi et al., 2016). Elucidating the processes underlying spatial variations in marine microbial community composition is essential to understand the structure and function of the marine microbiome (De Vargas et al., 2015; Lima-Mendez et al., 2015; Logares et al., 2020; Sunagawa et al., 2015).

Marine microbes exhibit varying vertical and horizontal mobility (Durham et al., 2013; Villareal, Altabet, & Culver-Rymsza, 1993), but due to their small size, how the move and where is mainly determined by ocean currents (Richter et al., 2019; Sommeria-Klein et al., 2021; Villarino et al., 2018). Over large spatial scales, the dispersal of microbes and the resulting connectivity, or lack thereof, between oceanic regions influence species biogeography and biodiversity (Jonsson & Watson, 2016; Villarino et al., 2018; Watson et al., 2011). However, as in terrestrial ecosystems (Condit et al., 2002; Zellweger, Roth, Bugmann, & Bollmann, 2017), marine microbial beta diversity (or β diversity)–the shift in species composition among sites (Whittaker, 1960) – is not only determined by dispersal, but also by speciation, selection, and ecological drift (Grilli, 2020; Hanson, Fuhrman, Horner-Devine, & Martiny, 2012; Martiny et al., 2006). Over timescales that are long compared to organism generation times, speciation enhances local diversity by adding new species to a community (Casteleyn et al., 2010), while selection retains species best adapted to local environmental conditions (Hellweger, Van Sebille, & Fredrick, 2014; Walworth, Zakem, Dunne, Collins, & Levine, 2020). Dispersal reduces compositional differences among sites (Soininen, Lennon, & Hillebrand, 2007), and ecological drift (Hubbell, 2001), or the demographic

stochasticity that creates differences among taxa, enhances beta diversity, similar to selection. These processes interact together to drive a decline in community similarity over increasing geographical distance, commonly called the distance-decay relationship (Nekola & White, 1999).

Distance-decay patterns have been previously observed for marine microbes in the surface ocean at a range of spatial scales and taxonomic resolutions using traditional geographical distance metrics, i.e., the length of a segment connecting two points (Chust, Irigoien, Chave, & Harris, 2013; Salazar et al., 2015; Zinger et al., 2011). In contrast to the surface layers of the ocean, where horizontal environmental gradients are strong, deep ocean environmental gradients are characteristically smooth (Reid, 1981), due to relatively limited mixing and advection. Weaker advection in the deep ocean likely implies weaker microbial dispersal there (Agogue, Lamy, Neal, Sogin, & Herndl, 2011; Hamdan et al., 2013).

Assessment of distance-decay patterns and the relative contribution of selection, dispersal, and drift in shaping the spatial structure of the ocean microbiome remains limited and confined to the upper (Ibarbalz et al., 2019; Logares et al., 2020; Richter et al., 2019; Sunagawa et al., 2015; Villarino et al., 2018) or deep ocean (Cermeno & Falkowski, 2009; Pernice et al., 2015; Salazar et al., 2015). A few studies have compared the microbial community structure across ocean layers, however they were either restricted to a single microbial group (prokaryotes or picoeukaryotes), (Cordier et al., 2022; Giner et al., 2019; Ward, Cael, Collins, & Robert Young, 2021; Zinger et al., 2011), temporally resolved but limited in space (Brown et al., 2009; Yeh & Fuhrman, 2022), or interpreted primarily by considering vertical transfer of sinking particles between surface and deep ocean (Guidi et al., 2016; Mestre et al., 2018; Ruiz-González et al., 2020).

These previous efforts suggested that the microbial horizontal diversity patterns appeared to be spatially structured by different ecological mechanisms globally in the pelagic (Logares et al., 2020) and benthic oceans (Cordier et al., 2022; Zinger et al., 2011). For example, in the surface oceans, picoeukaryote communities were predominantly structured by dispersal-limitation, while prokaryotic counterparts

appeared to be shaped by the combined action of dispersal-limitation, selection and drift (Logares et al., 2020). In the deep ocean, dispersal-limitation derived from differences in physical mixing and water mass played a fundamental role in the distribution patterns of marine bacteria and eukaryotes, as benthic communities showed a higher dissimilarity with increasing distance than pelagic communities (Cordier et al., 2022; Zinger et al., 2011). Across depth, there is a remarkably low overlap in microbial composition with well-defined microbial assemblages corresponding to the sunlit and dark ocean in prokaryotes and picoeukaryotes (Brown et al., 2009; Giner et al., 2019), while Mestre et al. (2018) found a strong vertical connectivity in prokaryotes as most abundant groups in the deep ocean also present in surface waters. However, a thorough comparison of horizontal patterns of microbial beta diversity including taxa of varying dispersal capabilities across upper and deep global ocean has yet to be conducted and can provide important insights into how microbial beta diversity is maintained in pelagic ecosystems.

Here, we report a comparative analysis of surface and deep ocean microbial beta diversity, defined as the difference in community composition between two sites. We analyzed the relationship between microbial community beta diversity, derived from metabarcoding data on prokaryotes and picoeukaryotes sampled during the Malaspina Circumnavigation Expedition in the upper and deep ocean (Duarte, 2015), with oceanic distance and environmental differences among sites. We tested the hypothesis that there will be greater similarity among microbial communities in the surface compared to deep ocean due to stronger horizontal water transport among sites. We also expect that there will be differences in the relative contribution of environmental filtering and oceanic distance on microbial community structure between the surface and deep ocean, and between prokaryotes and picoeukaryotes, due to different dispersal abilities and body sizes of microbes.

METHODS

Collection of biological data

Data on the relative abundance of various planktonic taxa were obtained from the Malaspina Expedition, which sailed the tropical and subtropical oceans in both hemispheres onboard the R/V Hespérides during December 2010 - July 2011 (Duarte, 2015). In this study, we focus on two members of the microbial community in the surface (0-3m) and deep ocean (4000m): the prokaryotes (0.2-3 µm and 0.2-0.8 µm, surface and deep, respectively) and the picoeukaryotes (0.2-3 µm) (Supplementary Table 1). The microbial community datasets used in this study are based upon previous works focusing on surface picoeukaryotes and prokaryotes (Logares et al., 2020), deep prokaryotes (Salazar et al., 2015), and deep picoeukaryotes (Giner et al., 2019). We included all publicly available data, including 118 sites for surface and deep prokaryotes, 30 sites for deep prokaryotes, and 12 sites for deep picoeukaryotes (Fig. 1). The prokaryotic group includes members of archaea and bacteria, including Thaumarchaeota and Proteobacteria as well as many other groups (Ruiz-González et al., 2019; Salazar et al., 2015). The picoeukaryotes include members of marine alveolates, radiolarians, and acanthareans, for example (Giner et al., 2019; Pernice et al., 2015) (Supplementary Table 1). Plankton samples were taken at the surface and deep ocean with Niskin bottles attached to a rosette sampler. To determine the composition of prokaryotes and picoeukaryotes present, ~12 L of seawater was used. Surface water samples were pre-filtered through a 200 µm nylon mesh to remove large plankton, and then by 3 and 0.2 µm pore-size polycarbonate filters of 142 mm diameter (Isopore, Millipore). The two filters were saved, including the 0.2 µm one that includes the 0.2 to 3 µm fraction, and the 3 µm one that includes the 3 to 20 µm fraction. Picoeukaryotes (Giner et al., 2019) and prokaryotes (Ruiz-González et al., 2019) were collected from the smaller of the two filters. Deep samples were pre-filtered through a 200 µm nylon mesh to remove large plankton, and then by 20, 0.8, and 0.2 µm pore-size polycarbonate filters of 142 mm diameter (Isopore, Millipore). The two filters were saved (i.e., the 0.2 µm one that includes the 0.2 to 0.8 µm fraction), and the 0.8 µm one (which includes the 0.8 to 20 µm fraction). The 0.2 to 0.8 µm fraction was used for observing free-living prokaryotes, (used in this study), and the 0.8-20 µm fraction for particle-attached prokaryotes (Salazar et al., 2015). For the deep picoeukaryotes, the 0.8-20 µm fraction was used (Pernice et al., 2015). In both the surface and deep samples, filtration time was 15-20 minutes. The filters were then flash-frozen in liquid N_2 and stored at -80 °C until DNA extraction. More complete details on sampling methodology are found in Giner et al. (2019), Logares et al. (2020), Ruiz-Gonzalez et al. (2019) and Salazar et al. (2015).

See supplementary Methods for full details on the sequencing and processing of the microbial community.

Collection of environmental data

We analyzed variations in environmental data that have relevance to microbial community structure and microbial physiology (Giner et al., 2019; Pernice et al., 2016; Salazar et al., 2015). Vertical profiles of salinity, temperature, dissolved oxygen, Chl-a, and PAR (Photosynthetic Active Radiation) were recorded continuously at each station with a Seabird 911 CTD probe, a Seabird SBE-43 polarographic membrane oxygen sensor, a Seapoint SCF fluorometer, and a Biospherical/Licor radiometer installed in the rosette sampler. The absorption coefficient of dissolved organic matter at 254 nm was used as a proxy of total dissolved organic carbon (Catalá et al., 2015). Samples were drawn from Niskin bottles at fourteen discrete depths throughout the water column, poured directly into acid cleaned 250mL glass bottles and immediately stored in dark conditions to allow equilibration with room temperature and to avoid photobleaching (Catalá et al. 2015). Primary productivity was estimated from Niskin bottles at five discrete depths across the epipelagic layer, but here we considered data from surface (3m) only. Data on surface and deep (~4000m) phosphate, nitrate and silicate were taken from the World Ocean Atlas (https://www.nodc.noaa.gov/OC5/SELECT/woaselect/woaselect.html). We used water mass age to trace the time elapsed since the water was last in contact with the atmosphere following Catalá et al. (2015), who interpolated the global gridded mean age estimates of Khatiwala et al. (2012) to the times, locations, and depths of the samples collected during the Malaspina circumnavigation. The difference in water mass age between surface and deep ocean is related to global thermohaline circulation, which may be an important factor for microbial communities because it constrains various environmental properties. Here, for surface and deep-water samples the average age of the epipelagic layer (0-200m) and the age at the corresponding sampling depth (about 4000 m) were considered, respectively. Though surface waters are relatively young compared with most deep-water samples, we found meaningful spatial variation (0-120 years) in water mass age between samples in the epipelagic layer. We summarize here the set of acronyms and units of the environmental variables used. PAR: Photosynthetic

Active Radiation (µE cm^{-2} s^{-1}), a254: absorption coefficient of dissolved organic matter at 254 nm (m^{-1}), PP: Primary Productivity (mg C m^{-3} h^{-1}), SiO_4 (µmol L^{-1}), PO_4 (µmol L^{-1}), NO_3 (µmol L^{-1}), Water mass age (years), T: Temperature (ºC), S: Salinity, O: Oxygen (µmol kg^{-1}).

Dissimilarity and distance matrices

We calculated the prokaryotic and picoeukaryotic community dissimilarity between all pairs of sites in the surface (0-3m) and deep ocean (4000m) using the Bray-Curtis dissimilarity (BC) index (Legendre & Legendre, 2012), based on the relative abundance of operational taxonomic units, or OTUs:

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j} \quad (1)$$

where *i* and *j* are the sites, C_{ij} is the sum of the lesser values for only those OTU counts in common between both sites, and S_i and S_j are the total number of individuals counted at both sites. The Bray-Curtis index is a well-accepted and commonly used metric to analyze spatial patterns of biodiversity, or beta diversity (Legendre & Legendre, 2012). It is sensitive to differences in abundance between species, and abundant species are weighted more than rare species. A dissimilarity value of 1 indicates that the two sites have no OTU in common, while a value 0 indicates the two sites have identical communities. In the text and Figures 2 and 4, we refer to "similarity", which is $1 - BC_{ij}$. In the case of similarity, a value of 1 indicates the communities are identical, and a value of 0 indicates they have no OTU in common. When we compare environmental and ocean distances to community structure, we used dissimilarity, as recommended by Legendre et al. (2012).

The oceanic distance matrix was calculated as the shortest oceanic path between two sampling sites (km) avoiding land. To do so, we used the *marmap* (Pante & Simon-Bouhet, 2013) package in R. We recognize that the oceanic distance is not the same as the advection distance i.e. physical transport of cells by ocean currents (Wilkins, Van

Sebille, Rintoul, Lauro, & Cavicchioli, 2013). However, the oceanic distance can be considered a proxy of the organism dispersal.

Temperature, salinity, dissolved oxygen, primary productivity, PAR, water mass age, a254, SiO_4, PO_4, and NO_3 were used to build the environmental difference matrix (Chust et al., 2013; Legendre & Legendre, 2012; Martiny et al., 2006). Each variable was scaled individually to have a mean of 0 and a variance of 1. Then, we calculated the environmental differences ($ED_{i,j}$) in n environmental parameters (p) between two sites (i and j) using the Euclidean distance:

$$ED_{i,j} = \sqrt{\sum_{p=1}^{n} (X_{i,p} - X_{j,p})^2} \quad (2)$$

High environmental distance between two sites means that the oceanic environment is more different than if the environmental distance is low. However, it is possible to have equivalent environmental distance because of changes across different variables. Moreover, not all the variables are necessarily independent of one another. These environmental variables were chosen because they have previously been shown to be important for determining microbial community beta diversity patterns (Villarino et al., 2018). Though simple, $ED_{i,j}$ is a relatively useful metric that allows us to quantify environmental distance incorporating many variables across strongly contrasting sites (Chust et al., 2013; Legendre & Legendre, 2012; Martiny et al., 2006).

We used a nonparametric Wilcoxon signed-rank test (Bauer, 1972) to determine whether there is a statistical difference in community similarities between two paired groups of samples. We compared community similarities of microbes across depth and across groups. We tested the null hypothesis (H_0) that the dissimilarity of the two populations is distributed equally *versus* the alternative hypothesis (H_1) where the distribution of dissimilarities in one of the populations is shifted to the left or right of the other.

Sample selection schemes

To consider possible biases due to different sampling effort across the datasets, we undertook three sets of analysis: a) "all sites", including data on all available sites for prokaryotes and picoeukaryotes, b) "group subset", with data sampled similarly in space, done by selecting the same or the closer sites across the groups in surface and deep ocean (118 *vs.* 118 sites in the surface, 12 *vs*. 12 sites in the deep) and c) "layer subset", including only sites matching the surface and deep ocean for prokaryotes (27 sites) and picoeukaryotes (12 sites) (Supplementary Fig. 1 a-b). In the main text, we present results for "all-sites" and the "subsets" in Table 1a-b, so that community similarity comparison across the biological groups and ocean layers are coherent. In the Supplementary Material, we include complementary analysis on Mantel correlations driving beta diversity and distance-decay analysis done on these subsets of data. No meaningful differences between the subsets and all available data analysis were found, but we present these data to increase confidence that the results are not dependent upon the sampling scheme.

RESULTS

Horizontal community similarity in the surface and deep ocean

Community similarity across all sites was significantly higher for prokaryotes in the surface and deep ocean (38.9% and 51.4% respectively), compared to picoeukaryotes (25.8% and 12.1% respectively; Wilcoxon signed-rank test, $p < 0.001$, Fig. 2, Table 1a-b). The same pattern was observed in the "group subset" analysis including a subset of data sampled similarly in space (Wilcoxon signed-rank test, $p < 0.001$, Supplementary Fig. 2a and Table 1a-b). For picoeukaryotes, the horizontal community similarity across pairs of sites was significantly higher in the surface compared to the deep ocean, both including all sites (Wilcoxon signed-rank test, $p < 0.001$, Fig. 2, Table 1a-b) and including only sites matching the surface and deep ocean in the "layer subset" analysis (Wilcoxon signed-rank test, $p < 0.001$, Supplementary Fig. 2b, Table 1a-b). In contrast, for prokaryotes, community similarities were significantly higher in the deep ocean compared to surface, including all sites and the "layer subset" analysis (Wilcoxon signed-rank test, $p > 0.05$, Fig. 2, Supplementary Fig. 2b, and Table 1a-b).

We also asked whether the community similarity within ocean basins differed from the similarity between basins. To do so, we calculated the community similarity within the Pacific, Atlantic, and Indian Oceans, using samples from just those Oceans (Supplementary Fig. 3a-b-c, Supplementary Fig. 4a-b). We then compared the community similarity across ocean basins using an ANOSIM test (see Supplementary Methods). The ANOSIM test revealed significantly higher similarities within than between basins, in both prokaryotes and picoeukaryotes in the surface and deep oceans when we included all sites (Supplementary Table 2, Supplementary Fig. 3a-b-c). Differences in community composition were higher in the deep ocean compared to surface, both including all observations and in the "layer subset" analysis, however in the latter differences were not statistically significant (Supplementary Table 2, Supplementary Fig 3a-b-c).

Relative contribution of environment and oceanic distance to horizontal community structure

For prokaryotes, we found that community composition changes due to environmental differences across sites (indicated by Mantel r correlations in Fig. 3) were consistently greater in the surface compared to the deep ocean. The most important and significant individual variables structuring surface prokaryote communities were oxygen concentration, followed by temperature and nutrient concentration (PO_4, and NO_3; Fig. 3, Supplementary Fig. 5). The same set of environmental variables shaped the spatial distribution of surface picoeukaryotes, together with the absorption coefficient of dissolved organic matter at 254 nm. In the deep ocean, the relative importance of the environmental variables was lower compared to surface and only water mass age appeared to be an important factor limiting the spatial distribution of prokaryote communities, and oxygen concentration in picoeukaryotes (Fig. 3, Supplementary Fig. 5). Both dissolved oxygen and water mass age showed larger horizontal gradients and greater variability in the deep ocean compared to surface (Supplementary Fig. 6). POC flux to the ocean floor is often linked to surface productivity (Yasuhara et al., 2012). Therefore, we correlated surface primary productivity, estimated during the Malaspina expedition (Duarte, 2015), with microbial community changes in the deep, instead of using estimates of POC flux derived from a model (Lutz, Caldeira, Dunbar, &

Behrenfeld, 2007). None of the Mantel correlations between surface primary productivity and deep ocean microbial beta diversity appears to be important driving biogeography of deep ocean microbes (Supplementary Fig. 7b Prokaryotes Mantel r = 0.02, p = 0.39, Picoeukaryotes Mantel r = 0.05 p = 0.37). Overall, a combination of temperature, salinity, oxygen, PO_4, NO_3, SiO_4, and water mass age showed the highest average Mantel correlation within the models tested (Mantel r = 0.258; Table 2). We kept the same set of environmental variables to define the microbial environmental niche across the layers. The high correlations between the environmental variables in the deep ocean are not surprising as the oldest waters are those in the Pacific, which have lower salinity and oxygen and more nutrients (Supplementary Fig. 8).

Microbial community composition changes were also correlated with the oceanic distance between pairs of sites, which we used to evaluate the influence of dispersal. In the surface, the oceanic distance explained a lower amount of variation in prokaryotes and picoeukaryotes community composition compared to the environmental spatial differences across sites (Table 3a, Fig. 3). In contrast, in the deep ocean, the oceanic distance was a slightly stronger driver of community composition relative to environmental conditions (Table 3a, Fig. 3). Partial Mantel tests were also used to determine the relative contribution of oceanic and environmental distance accounting for community similarity after controlling for the effect of each factor. Partial Mantel correlations between deep sample dissimilarities and environment were not significant for either of the biological groups when the influence of oceanic distance was considered (Table 3b), meaning a high shared covariation between environmental and ocean distance. In the surface, correlations between prokaryotic community composition and oceanic distance were not significant when the influence of the environment was considered (Table 3b). The portion of the variance explaining patterns of prokaryote and picoeukaryotes community composition slightly increased when accounting for the joint contribution of oceanic and environmental distance (Table 3c).

Community similarity decay over oceanic and environmental distance

Prokaryote and picoeukaryote horizontal community similarity decreased with increasing oceanic distance in both groups in the surface and deep ocean, showing

significant distance-decay patterns (Fig. 4a). Similarly, horizontal community similarity also decreased significantly with increasing environmental gradients in both prokaryotes and picoeukaryotes in the surface and deep ocean (Fig. 4b).

To address the difference in data density across the datasets, we complemented the analysis with (i) data sampled similarly in space "group subset', done by selecting the same or the closer sites across the groups in surface and deep ocean, and with (ii) data including only sites matching the surface and deep ocean "layer subset" (See "Sample selection schemes" section in methods). The results using the subsampled data are consistent with the results including all sites (Table 1a, Supplementary Table 3 and 4, Supplementary Fig. 2, 7a-b and 9a-b-c-d) as seen in the negative trends in distance-decay patterns in both analyses. The differences in the significance of the regressions are likely to be related to the decrease in sample size from "all data" to the "subsets".

DISCUSSION

Ocean currents are generally - though not always - stronger in the surface than in the deep ocean (Reid, 1981). This implies that microbial dispersal should be higher in the surface than in the ocean interior (Giner et al., 2019). On the other hand, the stronger environmental heterogeneity in the surface relative to the deep ocean can enhance community turnover due to niche effects. Our results showed that for picoeukaryotes, horizontal community similarity across pairs of sites is greater in the surface than at depth (Table 1a-b, Fig. 2), consistent with previous findings (Cordier et al., 2022). The effect of dispersal limitation on the spatial distribution of deep picoeukaryotes appears to be particularly strong, resulting in deep ocean communities that were more different to one another, over comparable distance to those on the surface (Fig. 3). Thus, we speculate that the greater degree of dispersal of picoeukaryotes in the surface compared to the deep ocean may underpin the observed depth differences in horizontal community similarity between the surface and deep ocean.

In contrast, we found that the prokaryote community similarity between pairs of sites was greater in the deep ocean compared to surface (Fig. 2). Despite the high dispersal capacity of prokaryotes, the strong environmental gradient observed in the surface seems to be driving stronger beta diversity compared to the deep ocean (Fig. 3), where

hydrodynamic activity is lower but environmental conditions are characteristically smooth (Reid, 1981). Hence, global dispersal potential for prokaryotes and subsequent environmental selection may represent a mechanism for driving patterns of microbial biogeography in the surface, in agreement with previous efforts at similar scales (Righetti et al., 2019; Sunagawa et al., 2015). Horizontal community similarity was generally higher in prokaryotes than in picoeukaryotes (Fig. 2). While the mechanisms that underly this contrast are uncertain, we suggest that the contrast could be driven by a suite of ecological mechanisms acting in concert, including differences in body and population sizes, dispersal capacity, and evolution strategies between prokaryotes and picoeukaryotes. Prokaryotes, due to their small body size and huge population sizes (De Bie et al., 2012; Peters, 1986), have high rates and distances of dispersal (Villarino et al., 2018) which reduces compositional differences between sites (Soininen et al., 2007), but also appear to be more sensitive to surface environmental changes (Fig. 2, Fig. 3). Picoeukaryotes are slightly larger, less abundant, and have typically more restricted ranges as they do not disperse as far or as quickly (De Bie et al., 2012; Villarino et al., 2018). In essence, our results show that prokaryotes, disperse widely both in the upper and deep ocean, such that horizontal similarity is high across depths (Fig. 3). Our results support the size-dispersal hypothesis which predicts that smaller organisms are more likely affected by species sorting than dispersal limitation, because smaller organisms can disperse almost everywhere (Cottenie, 2005) and thus their distributions strongly reflect environmental effects (Farjalla et al., 2012). In contrast, larger sized picoeukaryotes may have weaker dispersal, and depth contrasts in horizontal community similarity (Fig 2). The analysis on deep picoeukaryotes, however, was limited to relatively few sites and therefore we have to be cautious when interpreting the beta diversity patterns observed.

Other than size, the observed lower similarities of picoeukaryotes relative to prokaryotes could be associated with their greater diversity (Supplementary Table 1). Global-scale marine microbiome surveys such as *TARA* have shown that eukaryotic plankton are taxonomically and ecologically diverse, possibly even more so than prokaryotic plankton (De Vargas et al., 2015). Picoeukaryotes include many photosynthetic groups that, as primary producers, might be more variable in abundance due to light and nutrients conditions and grazing and sinking dynamics (Agusti et al., 2015; Boeuf et al., 2019). The increased beta diversity in picoeukaryotes compared to

prokaryotes could also be explained by the relatively more stable abundances of prokaryotes (Caporaso, Paszkiewicz, Field, Knight, & Gilbert, 2012; Yeh & Fuhrman, 2022), which include numerous heterotrophic lineages, for example the widely distributed SAR11 (Giovannoni, 2017). The observed differences in horizontal community similarity between picoeukaryotes and prokaryotes are intriguing and may result from the factors that we have discussed here, but further studies of the mechanisms underlying this divergence should be conducted. On the other hand, the depth contrast in picoeukaryote beta diversity found here is supported by Giner et al. (2019), where most photosynthetic groups (e.g., Pelagophyceae, green algae) were shown to decline in abundance with depth. In the same study, Giner et al., using the rRNA: rDNA ratio as a proxy of metabolic activity, found a lower fraction of metabolically active picoeukaryotes cells in bathypelagic waters compared to surface. This may imply that inactive taxa contribute more importantly to the beta diversity in deep waters as compared to surface, although it is assumed that dead cells should not accumulate in large abundance in environmental samples. The information presently available is still too scarce to draw conclusions. Future studies associating microbial identity with concurrent metabolic state would also help to understand the relationship between microbial community structure and their ecosystems functions.

For surface prokaryotes, and to a lesser extent deep picoeukaryotes, the similarity distributions (Fig. 2) are bimodal, while this is not the case for picoeukaryotes. When looking within ocean basins and hemispheres (Supplementary Fig. 4a), surface prokaryotes also exhibit a bimodal similarity distribution. We speculate that the bimodality may arise because prokaryotic communities are similar within large marine biomes (high similarity), but very different from communities in other biomes (low similarity). This could give rise to the bimodal distribution in the surface and explain why it does not appear in the deep. Distributions of picoeukaryotes, which disperse less readily, might not exhibit the bimodality, and instead have a smoother change in community structure over space (Supplementary Fig. 4a-b).

Overall, our findings revealed that patterns of microbial community similarity are sensitive to changes in environmental conditions, as expected from the ecological niche theory, which states that the environmental factors filter ecological communities (Leibold et al., 2004), but this sensitivity appears to differ between the surface and deep

ocean and between prokaryote and picoeukaryote communities (Fig. 3, Fig. 4a-b). Though the sampling conducted on the Malaspina cruise was restricted to tropical and subtropical regions, the environmental and ecological gradients observed are large enough to derive the spatial distribution patterns found (i.e., surface temperature = 15.7-29.3 °C, mean = 24.5 °C and standard deviation = 3.2 °C; depth temperature = 0.5-2.1 °C, mean = 1.4 °C and standard deviation = 0.5 °C, Supplementary Fig. 6). Temperature, oxygen, and phosphate were the strongest environmental factors structuring the sunlit microbiome, in agreement with previous global assessments (Righetti et al., 2019; Sunagawa et al., 2015). In the deep ocean, we found that dissolved oxygen was the most important driver of picoeukaryote beta diversity (Fig. 3). Contrasting oxygen conditions among ocean basins selecting for different marine communities have been described before for deep ocean microbes globally (Giner et al., 2019; Salazar et al., 2015). Water-mass age in the deep ocean explained a high portion of the prokaryote spatial patterns. The deep ocean holds a combination of water masses of contrasting origin, age (time elapsed since last in contact with the atmosphere) and ageing (time-integrated microbial respiration), which are related to global thermohaline circulation (Catalá et al. 2015). Several studies have pointed to differences in the biogeographical distribution of the major groups of planktonic Archaea (Teira, Lebaron, Van Aken, & Herndl, 2006) and bacteria (Giovannoni, 2017), which suggest that distinct water masses may harbour distinct prokaryotic communities. In contrast, surface waters show limited spatial variability in water mass age (Catalá et al. 2015).

The taxonomic diversity approach taken in this study should be complemented with other biodiversity facets, including functional, phylogenetic or the diversity of biotic interactions (Louca, Parfrey, & Doebeli, 2016). Here, we focused on spatial variation in the entire assemblage, but we did not consider the different responses of each individual microbe to environmental variations. In fact, selection exerted by an environmental variable drive species association networks (Lima-Mendez et al., 2015; Logares et al., 2020; Sommeria-Klein et al., 2021) increasing community similarity in regions with similar environments. However, the observed higher sensitivity of surface prokaryotes to the environment relative to picoeukaryotes found in our study are consistent with the biotic interaction aware analysis using *TARA* and *Malaspina* data by Logares et al. (2020).

To further constrain the importance of oceanic distance and dispersal limitation, we compared differences in community composition within and between the Atlantic, Pacific, and Indian oceans (Supplementary Table 2, Supplementary Fig. 3a-b-c). If dispersal limitation increases with geographic distance (Nekola & White, 1999), community similarities should progressively increase from regional to global datasets. It appears that the microbial communities in the deep ocean are, to some extent, more structured across basin compared to surface (Supplementary Table 2, Supplementary Fig. 3a-b-c), partially supporting our hypothesis that lower water mixing in the deep ocean is driving higher beta diversity in microbes. This finding also reinforces the key role of transport by oceanic currents shaping microbial biogeography, in particular by extending the distribution of taxa within basins and constraining long distance dispersal between basins (Sommeria-Klein et al., 2021; Sunagawa et al., 2015).

In this study, we showed that micro-organism exhibit biogeographical patterns as do larger organism (Martiny et al., 2006), and such patterns seem to be governed by similar underlying ecological determinants including present day environmental conditions as well as historical contingencies and dispersal limitation (Fig. 3, Fig. 4a-b). In the deep ocean, distance-decay patterns in benthic macrofauna have been found to be primarily determined by strong environmental gradients in temperature, POC and oxygen, however, dispersal limitation has also been found to be key in the community makeup (Jöst et al., 2019; McClain & Rex, 2015). The larger dispersal distances of microbes relative to bigger taxa may underpin the weaker associations with temperature and POC observed in deep microbial distribution (Fig. 3). In the upper ocean, in line with the observed microbial spatial patterns, environmental conditions, particularly temperature, are also strong correlates of coastal fish beta diversity in the Mediterranean Sea (Hattab et al., 2015).

In summary, we found distinct patterns beta diversity of marine microbes depending on the group and depth of the ocean. The higher horizontal community similarity of picoeukaryotes in the surface compared to the deep ocean may be driven by strong surface oceanic currents. In contrast, the similar horizontal community similarity of marine prokaryotes at the surface and in the deep could be caused by their extremely high dispersal. Our results show that in the surface ocean, the environment filters the

prokaryotic communities to a greater extent than picoeukaryote communities. On the other hand, dispersal limitation, alongside with niche specialization, drive the biogeography of deep ocean microbial communities. Overall, these findings expand current understanding of the ecological mechanisms underlying distributional patterns of marine microbial diversity in the global ocean.

DATA AVAILABILITY

DNA sequences for surface prokaryotes are publicly available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena; accession number PRJEB25224 [16S rRNA genes]), for deep prokaryotes at the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra) under accession ID SRP031469, and for surface and deep picoeukaryotes at the European Nucleotide Archive with Accession number PRJEB23771 (http://www.ebi.ac.uk/ena). Environmental data used in this study is available at https://github.com/ramalok/malaspina.surface.metabacoding, Giner et al. (2019) and Salazar et al. (2015).

REFERENCES

Agogue, H., Lamy, D., Neal, P. R., Sogin, M. L., & Herndl, G. J. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Molecular Ecology*, *20*, 258–274. https://doi.org/10.1111/j.1365-294X.2010.04932.x

Agusti, S., González-Gordillo, J. I., Vaqué, D., Estrada, M., Cerezo, M. I., Salazar, G., … Duarte, C. M. (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nature Communications 2015 6:1*, *6*(1), 1–8. https://doi.org/10.1038/ncomms8608

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology 2007 5:10*, *5*(10), 782–791. https://doi.org/10.1038/nrmicro1747

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, *67*(339), 687–690. https://doi.org/10.1080/01621459.1972.10481279

Boeuf, D., Edwards, B. R., Eppley, J. M., Hu, S. K., Poff, K. E., Romano, A. E., … DeLong, E. F. (2019). Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proceedings of the National Academy of Sciences*, *116*(24), 11824–11832.

Brown, M. V., Philip, G. K., Bunge, J. A., Smith, M. C., Bissett, A., Lauro, F. M., … Donachie, S. P. (2009). Microbial community structure in the North Pacific ocean. *ISME Journal*, *3*(12), 1374–1386. https://doi.org/10.1038/ismej.2009.86

Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R., & Gilbert, J. A. (2012). The Western English Channel contains a persistent microbial seed bank. *ISME Journal*, *6*(6), 1089–1093. https://doi.org/10.1038/ISMEJ.2011.162

Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.-E., Kotaki, Y., Rhodes, L., … Vyverman, W. (2010). Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proceedings of the National Academy of Sciences*, *107*, 12952–12957.

https://doi.org/10.1073/pnas.1001380107

Catalá, T. S., Reche, I., Álvarez, M., Khatiwala, S., Guallart, E. F., Benítez-Barrios, V. M., … Álvarez-Salgado, X. A. (2015). Water mass age and aging driving chromophoric dissolved organic matter in the dark global ocean. *Global Biogeochemical Cycles*, *29*(7), 917–934. https://doi.org/10.1002/2014GB005048

Cermeno, P., & Falkowski, P. G. (2009). Controls on diatom biogeography in the ocean. *Science*, *325*, 1539–1541. https://doi.org/10.1126/science.1174159

Chust, G., Irigoien, X., Chave, J., & Harris, R. P. (2013). Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography*, *22*, 531–543. https://doi.org/10.1111/geb.12016

Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., … Hubbell, S. P. (2002). Beta-Diversity in Tropical Forest Trees. *Science*, *295*, 666–669. https://doi.org/10.1126/science.1066854

Cordier, T., Angeles, I. B., Henry, N., Lejzerowicz, F., Berney, C., Morard, R., … Pawlowski, J. (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Science Advances*, *8*(5), 9309. Retrieved from https://www.science.org/doi/full/10.1126/sciadv.abj9309

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, *8*, 1175–1182. https://doi.org/10.1111/j.1461-0248.2005.00820.x

De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., … Declerck, S. A. J. (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters*, *15*, 740–747. https://doi.org/10.1111/j.1461-0248.2012.01794.x

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237). Retrieved from https://www.science.org/doi/abs/10.1126/science.1261605

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

Durham, W. M., Climent, E., Barry, M., De Lillo, F., Boffetta, G., Cencini, M., & Stocker, R. (2013). Turbulence drives microscale patches of motile phytoplankton. *Nature Communications*, *4*, 2148. https://doi.org/10.1038/ncomms3148

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, *281*, 200. Retrieved from http://science.sciencemag.org/content/281/5374/200.abstract

Farjalla, V. F., Srivastava, D. S., Marino, N. A. C., Azevedo, F. D., Dib, V., Lopes, P. M., … Esteves, F. A. (2012). Ecological determinism increases with organism size. *Ecology*, *93*, 1752–1759. https://doi.org/10.1890/11-1144.1

Giner, C. R., Pernice, M. C., Balagué, V., Duarte, C. M., Gasol, J. M., Logares, R., & Massana, R. (2019). Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *ISME Journal*. https://doi.org/10.1038/s41396-019-0506-9

Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science*, *9*(1), 231–255. Retrieved from https://pubmed.ncbi.nlm.nih.gov/27687974/

Grilli, J. (2020). Macroecological laws describe variation and diversity in microbial communities. *Nature Communications*, *11*(1), 4743. https://doi.org/10.1038/s41467-020-18529-y

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., … Gorsky, G. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*, 465–470. Retrieved from http://internal-pdf//Guidi et al 2016 Nature-0762905347/Guidi et al 2016 Nature.pdf

Hamdan, L. J., Coffin, R. B., Sikaroodi, M., Greinert, J., Treude, T., & Gillevet, P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *ISME Journal*, *7*(4), 685–696. https://doi.org/10.1038/ismej.2012.143

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. H. (2012, July). Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nature Reviews Microbiology*, Vol. 10, pp. 497–506. https://doi.org/10.1038/nrmicro2795

Hattab, T., Albouy, C., Ben Rais Lasram, F., Le Loc'h, F., Guilhaumon, F., & Leprieur,

F. (2015). A biogeographical regionalization of coastal Mediterranean fishes. *Journal of Biogeography*, *42*(7), 1336–1348. https://doi.org/10.1111/JBI.12505

Hellweger, F. L., Van Sebille, E., & Fredrick, N. D. (2014). Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science*, *345*(6202), 1346–1349. https://doi.org/10.1126/science.1254421

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Retrieved from http://www.jstor.org/stable/j.ctt7rj8w

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., … Zinger, L. (2019). Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*, *179*(5), 1084-1097.e21. https://doi.org/10.1016/j.cell.2019.10.008

Jonsson, B. F., & Watson, J. R. (2016). The timescales of global surface-ocean connectivity. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11239

Jöst, A. B., Yasuhara, M., Wei, C. L., Okahashi, H., Ostmann, A., Martínez Arbizu, P., … Brix, S. (2019). North Atlantic Gateway: Test bed of deep-sea macroecological patterns. *Journal of Biogeography*, *46*(9), 2056–2066. https://doi.org/10.1111/jbi.13632

Khatiwala, S., Primeau, F., & Holzer, M. (2012). Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age. *Earth and Planetary Science Letters*, *325–326*, 116–125. https://doi.org/10.1016/j.epsl.2012.01.038

Legendre, P., & Legendre, L. F. J. (2012). *Numerical Ecology*. Retrieved from https://books.google.es/books?id=6ZBOA-iDviQC

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., … Gonzalez, A. (2004, July 4). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, Vol. 7, pp. 601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., … Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*. Retrieved from http://science.sciencemag.org/content/348/6237/1262073.abstract

Logares, R., Deutschmann, I. M., Junger, P. C., Giner, C. R., Krabberød, A. K., Schmidt, T. S. B., … Massana, R. (2020). Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome*, *8*(1), 1–17. https://doi.org/10.1186/s40168-020-00827-8

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*(6305), 1272–1277.

Lutz, M. J., Caldeira, K., Dunbar, R. B., & Behrenfeld, M. J. (2007). Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. *Journal of Geophysical Research: Oceans*, *112*(C10), 10011. https://doi.org/10.1029/2006JC003706

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., … Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro*, *4*, 102–112. Retrieved from http://dx.doi.org/10.1038/nrmicro1341

McClain, C. R., & Rex, M. A. (2015). Toward a Conceptual Understanding of β-Diversity in the Deep-Sea Benthos. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 623–642. https://doi.org/10.1146/annurev-ecolsys-120213-091640

Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., & Sala, M. M. (2018). Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Sciences*, *115*(29), E6799–E6807. https://doi.org/10.1073/PNAS.1802470115

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*, 867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

Pante, E., & Simon-Bouhet, B. (2013). marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS ONE*, *8*, e73051. https://doi.org/10.1371/journal.pone.0073051

Pernice, M. C., Forn, I., Gomes, A., Lara, E., Alonso-Saez, L., Arrieta, J. M., … Massana, R. (2015). Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME Journal*, *9*, 782–792. https://doi.org/10.1038/ismej.2014.168

Pernice, M. C., Giner, C. R., Logares, R., Perera-Bel, J., Acinas, S. G., Duarte, C. M., … Massana, R. (2016). Large variability of bathypelagic microbial eukaryotic communities across the world/'s oceans. *ISME Journal*, *10*, 945–958. https://doi.org/10.1038/ismej.2015.170

Peters, R. H. (1986). *The Ecological Implications of Body Size*. Retrieved from https://books.google.es/books?id=OYVxiZgTXWsC

Reid, J. L. (1981). On the mid-depth circulation of the world ocean. *Evolution of Physical Oceanography*, *623*, 70–111.

Richter, D. J., Watteaux, R., Vannier, T., Leconte, J., Frémont, P., Reygondeau, G., … Jaillon, O. (2019). Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *BioRxiv*, *23*, 867739. https://doi.org/10.1101/867739

Righetti, D., Vogt, M., Gruber, N., Psomas, A., & Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, *5*(5), 6253–6268. Retrieved from https://www.science.org/doi/full/10.1126/sciadv.aau6253

Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M., Salazar, G., Agustí, S., … Gasol, J. M. (2020). Major imprint of surface plankton on deep ocean prokaryotic structure and activity. *Molecular Ecology*, *29*(10), 1820–1838. https://doi.org/10.1111/MEC.15454

Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., … Gasol, J. M. (2019). Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. *Molecular Ecology*, *28*(8), 1930–1945. https://doi.org/10.1111/mec.15026

Salazar, G., Cornejo-Castillo, F. M., Benítez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X. A., Duarte, C. M., … Acinas, S. G. (2015). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME Journal*, *10*, 596. Retrieved from http://dx.doi.org/10.1038/ismej.2015.137

Soininen, J., Lennon, J. J., & Hillebrand, H. (2007). A multivariate analysis of beta diversity across organism and environments. *Ecology*, *88*, 2830–2838. https://doi.org/10.1890/06-1730.1

Sommeria-Klein, G., Watteaux, R., Ibarbalz, F. M., Karlusich, J. J. P., Iudicone, D., Bowler, C., & Morlon, H. (2021). Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science*, *374*(6567), 594–599. Retrieved from https://www.science.org/doi/pdf/10.1126/science.abb3717

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237). https://doi.org/10.1126/science.1261359

Teira, E., Lebaron, P., Van Aken, H., & Herndl, G. J. (2006). Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. *Limnology and Oceanography*, *51*(5), 2131–2144. https://doi.org/10.4319/LO.2006.51.5.2131

Villareal, T. A., Altabet, M. A., & Culver-Rymsza, K. (1993). Nitrogen transport by vertically migrating diatom mats in the North Pacific Ocean. *Nature*, *363*(6431), 709–712. https://doi.org/10.1038/363709a0

Villarino, E., Watson, J. R., Jönsson, B., Gasol, J. M., Salazar, G., Acinas, S. G., … Chust, G. (2018). Large-scale ocean connectivity and planktonic body size. *Nature Communications*, *9*, 142. https://doi.org/10.1038/s41467-017-02535-8

Walworth, N. G., Zakem, E. J., Dunne, J. P., Collins, S., & Levine, N. M. (2020). Microbial evolutionary strategies in a dynamic ocean. *Proceedings of the National Academy of Sciences*, *117*(11), 5943–5948. Retrieved from https://www.pnas.org/content/117/11/5943

Ward, B. A., Cael, B. B., Collins, S., & Robert Young, C. (2021). Selective constraints on global plankton dispersal. *Proceedings of the National Academy of Sciences*, *118*(10). Retrieved from https://doi.org/10.1073/pnas.2007388118

Watson, J. R., Siegel, D. A., Kendall, B. E., Mitarai, S., Rassweiller, A., & Gaines, S. D. (2011). Identifying critical regions in small-world marine metapopulations. *Proceedings of the National Academy of Sciences*, *108*, E907. Retrieved from http://www.pnas.org/content/108/43/E907.abstract

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*, 279–338. https://doi.org/10.2307/1943563

Wilkins, D., Van Sebille, E., Rintoul, S. R., Lauro, F. M., & Cavicchioli, R. (2013).

Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3457

Yasuhara, M., Hunt, G., Van Dijken, G., Arrigo, K. R., Cronin, T. M., & Wollenburg, J. E. (2012). Patterns and controlling factors of species diversity in the Arctic Ocean. *Journal of Biogeography*, *39*, 2081–2088. https://doi.org/10.1111/j.1365-2699.2012.02758.x

Yeh, Y.-C., & Fuhrman, J. A. (2022). Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Communications 2022 2:1*, *2*(1), 1–12. https://doi.org/10.1038/s43705-022-00121-8

Zellweger, F., Roth, T., Bugmann, H., & Bollmann, K. (2017). Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure. *Global Ecology and Biogeography*, *26*(8), 898–906. https://doi.org/10.1111/geb.12598

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., … Ramette, A. (2011). Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. *PLoS ONE*, *6*, e24570. https://doi.org/10.1371/journal.pone.0024570

TABLES

Table 1. (a) Community similarity values (min, max, mean, median) across paired sites. (b) Non-parametric Wilcoxon sign rank test comparing community similarities of microbes across depth and across group. We tested the null hypothesis (Ho) that the similarity of the two populations are distributed equally versus the alternative hypothesis (H_1) where the distribution of similarities in one population is shifted to the left or right of the other. The The *p* value of the test is calculated at 95% confidence interval. Comparison across groups and layer was made using the subset data: "All sites" includes data on all available sites for prokaryotes and picoeukaryotes in the surface and deep ocean; "Group subset" includes data sampled similarly in space, done by selecting the same or the closer sites across the groups in surface and deep ocean; "Layer subset" includes only sites matching the surface and deep ocean for prokaryotes. (*) cases with a reduction in sample size (number of stations) due to the subsets.

a)

Group	Depth	Number of sites	Community similarity				Analysis
			Min.	Median	Mean	Max.	
Picoeukaryotes	Surface	118	0.003	0.264	0.258	0.646	All sites
	Deep	12	0.002	0.076	0.121	0.727	
Prokaryotes	Surface	118	0.035	0.357	0.389	0.809	
	Deep	30	0.187	0.534	0.514	0.794	
Picoeukaryotes	Surface	118	0.003	0.264	0.258	0.646	Group subset
	Deep	12	0.002	0.076	0.121	0.727	
Prokaryotes	Surface	118	0.035	0.357	0.389	0.809	
	Deep*	12	0.187	0.480	0.458	0.775	
Picoeukaryotes	Surface*	12	0.042	0.267	0.260	0.434	Layer subset
	Deep	12	0.002	0.076	0.121	0.727	
Prokaryotes	Surface*	27	0.075	0.440	0.438	0.773	
	Deep*	27	0.187	0.533	0.514	0.794	

b)

Community similarity		Number of sites	Number of dissimilarity pairs	V	p
Across depth	Deep vs Surface Prokaryotes	27	351	42052	<0.001
	Deep vs Surface Picoeukaryotes	12	66	247	<0.001
Across groups	Surface Prokaryotes vs Picoeukaryotes	118	6903	20977945	<0.001
	Deep Prokaryotes vs Picoeukaryotes	12	66	2182	<0.001

Table 2. Mantel correlations between environmental variables and community dissimilarity, as measured by the Bray-Curtis index, for prokaryotes and picoeukaryotes. Key environmental variables are selected using Mantel tests. We select the model with higher mantel *r* correlations between community dissimilarity and the Euclidean distances of scaled environmental variables (* indicates $p < 0.05$). The analysis includes all available sites. n: the number of sites, which varies according to the group and layer analysed (See Fig. 1).

Environmental variables	Picoeukaryotes		Prokaryotes		Mean Mantel *r* correlation
	Surface (n=118)	Deep (n=12)	Surface (n=118)	Deep (n=30)	
T+S+O+PO_4+SiO_4+NO_3+Water age	0.215*	0.257*	0.412*	0.148*	0.258
T+S+O+ PO_4+SiO_4+NO_3	0.233*	0.222	0.441*	0.104	0.250
T+S+O+PO_4	0.240*	0.206	0.454*	0.117	0.254
T+S+O+SiO_4	0.216*	0.208	0.386*	0.115	0.231
T+S+O+ NO_3	0.219*	0.203	0.445*	0.119	0.246
T+S+O	0.205*	0.188	0.395*	0.129	0.229

Table 3. (a) Mantel correlations, (b) Partial Mantel correlations and (c) Multiple Regression on distance Matrices (MRM) between microbial community dissimilarity, oceanic distance, and environmental distance. Mantel partial correlations are calculated after controlling for the effects of oceanic and environmental distance. The statistical significance of the Mantel is assessed using 9999 random permutations. MRM involves a multiple regression analysis of a response matrix (community dissimilarity) on any number of explanatory variables (in this case, environmental and oceanic distance), where each matrix contains distance or dissimilarities between all pair-wise combination of *n* objects. The test of statistical significance is performed by permutations (* indicates $p < 0.05$).

(a)

Biological group	Depth	Number of sites	Dissimilarity *vs* Oceanic distance		Dissimilarity *vs* Environmental distance	
			Mantel *r*	p	Mantel *r*	p
Picoeukaryotes	Surface	118	0.127*	<0.001	0.215*	<0.001
	Deep	12	0.341*	0.004	0.257*	0.028
Prokaryotes	Surface	118	0.082*	<0.001	0.412*	<0.001
	Deep	30	0.177*	<0.001	0.148*	0.031

(b)

Biological group	Depth	Number of sites	Dissimilarity *vs* Oceanic distance (control environment)		Dissimilarity *vs* Environmental distance (control oceanic distance)	
			Partial Mantel *r*	p	Partial Mantel *r*	p
Picoeukaryotes	Surface	118	0.077*	<0.001	0.191*	0.001
	Deep	12	0.280*	0.017	0.160	0.117
Prokaryotes	Surface	118	-0.024	0.889	0.406*	<0.001
	Deep	30	0.117*	0.028	0.063	0.246

(c)

Biological group	Depth	Number of sites	MRM (Oceanic distance + Environment)	
			MRM *r*	p
Picoeukaryotes	Surface	118	0.228*	0.001
	Deep	12	0.373*	0.011
Prokaryotes	Surface	118	0.413*	0.001
	Deep	30	0.188*	0.046

FIGURE CAPTIONS

Fig 1. Map showing the Malaspina sites for picoeukaryotes and prokaryotes in the surface and deep ocean. N indicates the number of sites.

Fig 2 Frequency of microbial horizontal community similarity values between pairs of sites in surface and deep ocean for picoeukaryotes and prokaryotes (bars). The y-axis indicates the relative proportion of the total for each bar. The x-axis indicates the community similarity values. A similarity value of 0 indicates that the two sites share no OTU in common, while a value 1 indicates the two sites have identical communities. The smooth line indicates the kernel density estimate of the distribution.

Fig 3. Mantel correlations of prokaryotes and picoeukaryote community dissimilarity with environmental factors. Red colours indicate positive Mantel correlation between community dissimilarity and environmental variables, blue colours indicate negative mantel correlations, and an *r* value of 0 (white) indicates no correlation. * indicates $p < 0.05$. PAR: Photosynthetic Active Radiation (µE cm^{-2} s^{-1}), a254: absorption coefficient of dissolved organic matter at 254 nm (m^{-1}), PP: Primary Productivity (mg C m^{-3} h^{-1}), SiO_4 (µmol L^{-1}), PO_4 (µmol L^{-1}), NO_3 (µmol L^{-1}), Water age (years), T: Temperature (ºC), S: Salinity, O: Oxygen (µmol kg^{-1}), Ocean dist: least cost oceanic distance avoiding land, Env dist: Environmental Euclidean distance.

Fig 4. Relationships between horizontal community similarity *vs* oceanic distance (a), and community similarity *vs* environmental distance between pairs of sites (b) in the surface and deep ocean for prokaryotes and picoeukaryotes. The points show the community similarity values for each unit of distance considering all pairs of sites. The relationship between community similarity and oceanic and environmental distance is fitted by a linear model (solid line), and the smooth shadows shows the 95% confidence level interval for predictions of a linear model. Regression lines are drawn on significant mantel correlations. A similarity value of 0 indicates that the two sites share no OTU in common, while a value 1 indicates the two sites have identical communities.

Fig. 1

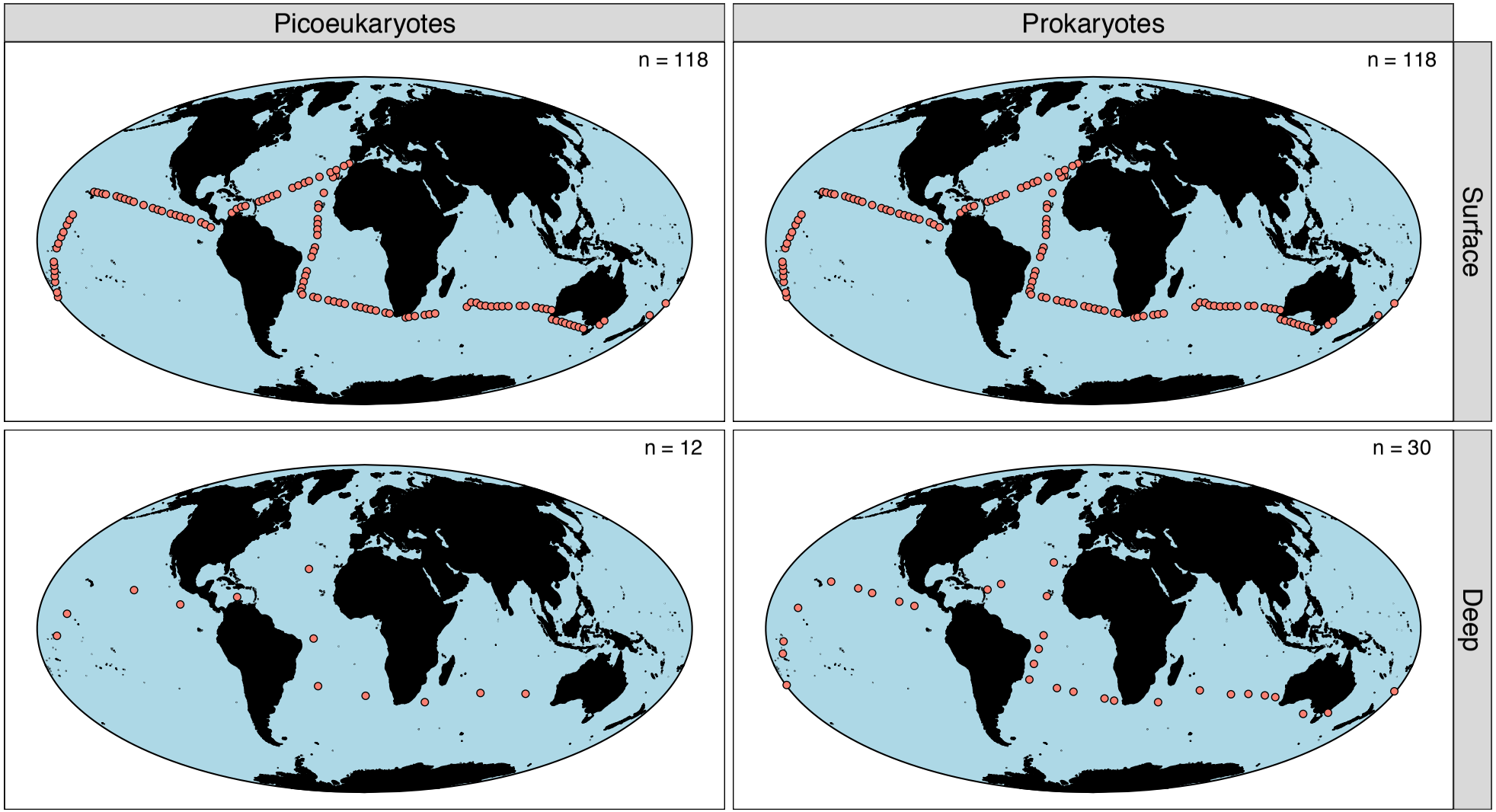

Picoeukaryotes
Prokaryotes
n = 118
n = 118
n = 12
n = 30
Surface
Deep

Fig. 2

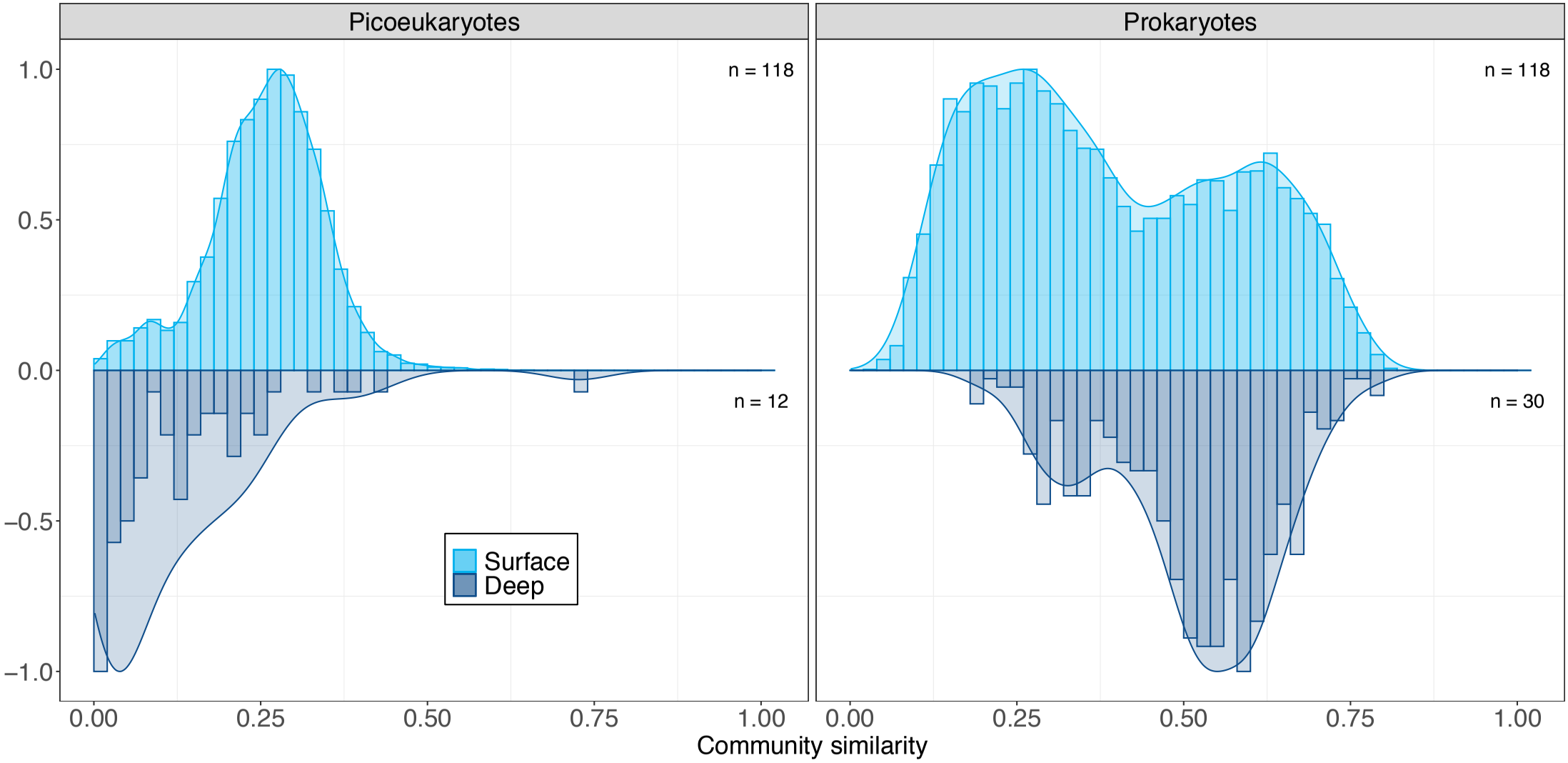

Picoeukaryotes
Prokaryotes
n = 118
n = 12
n = 118
n = 30
Surface
Deep
1.0
0.5
0.0
−0.5
−1.0
0.00
0.25
0.50
0.75
1.00
Frequency
Community similarity

Fig. 3

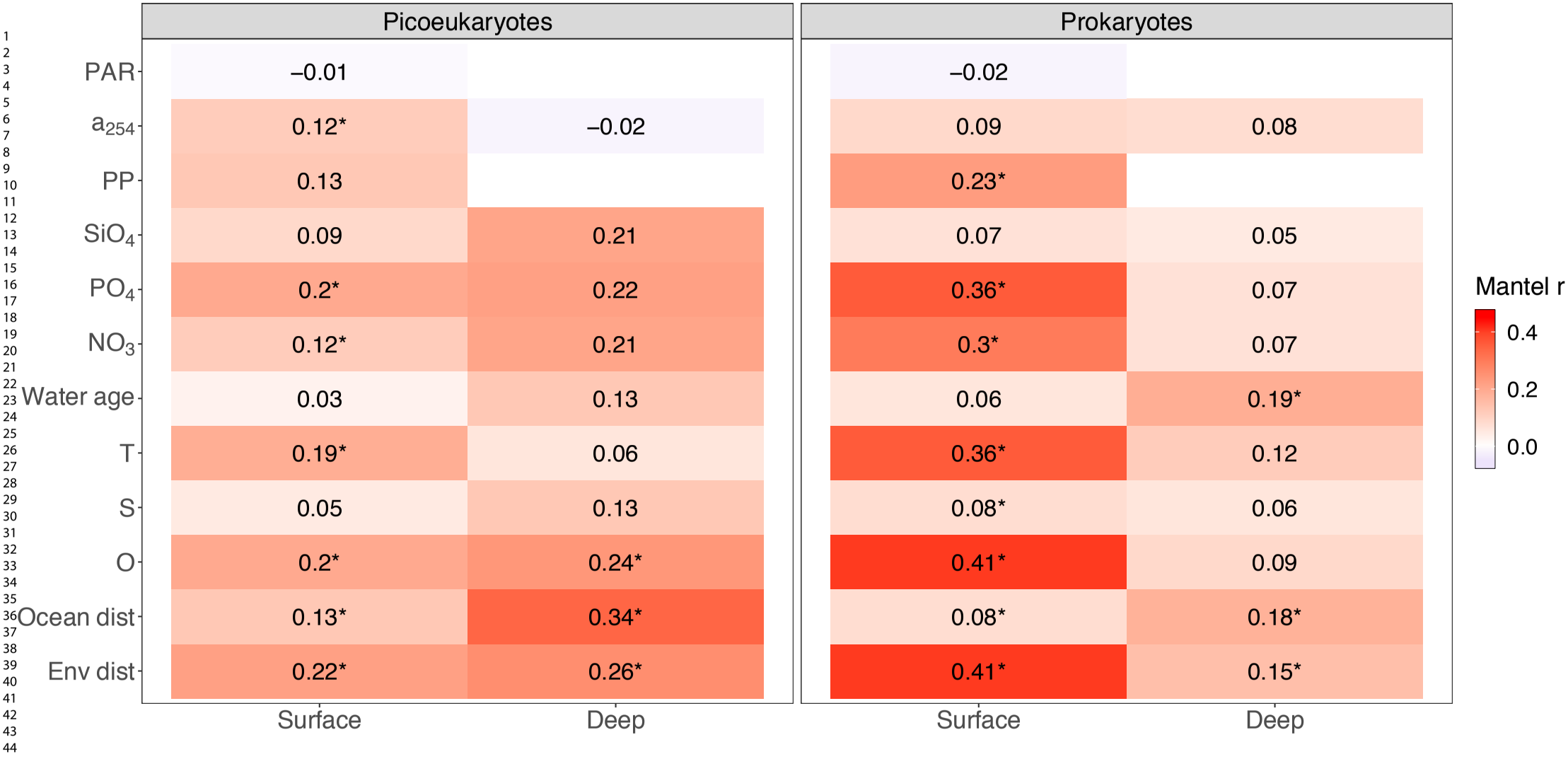
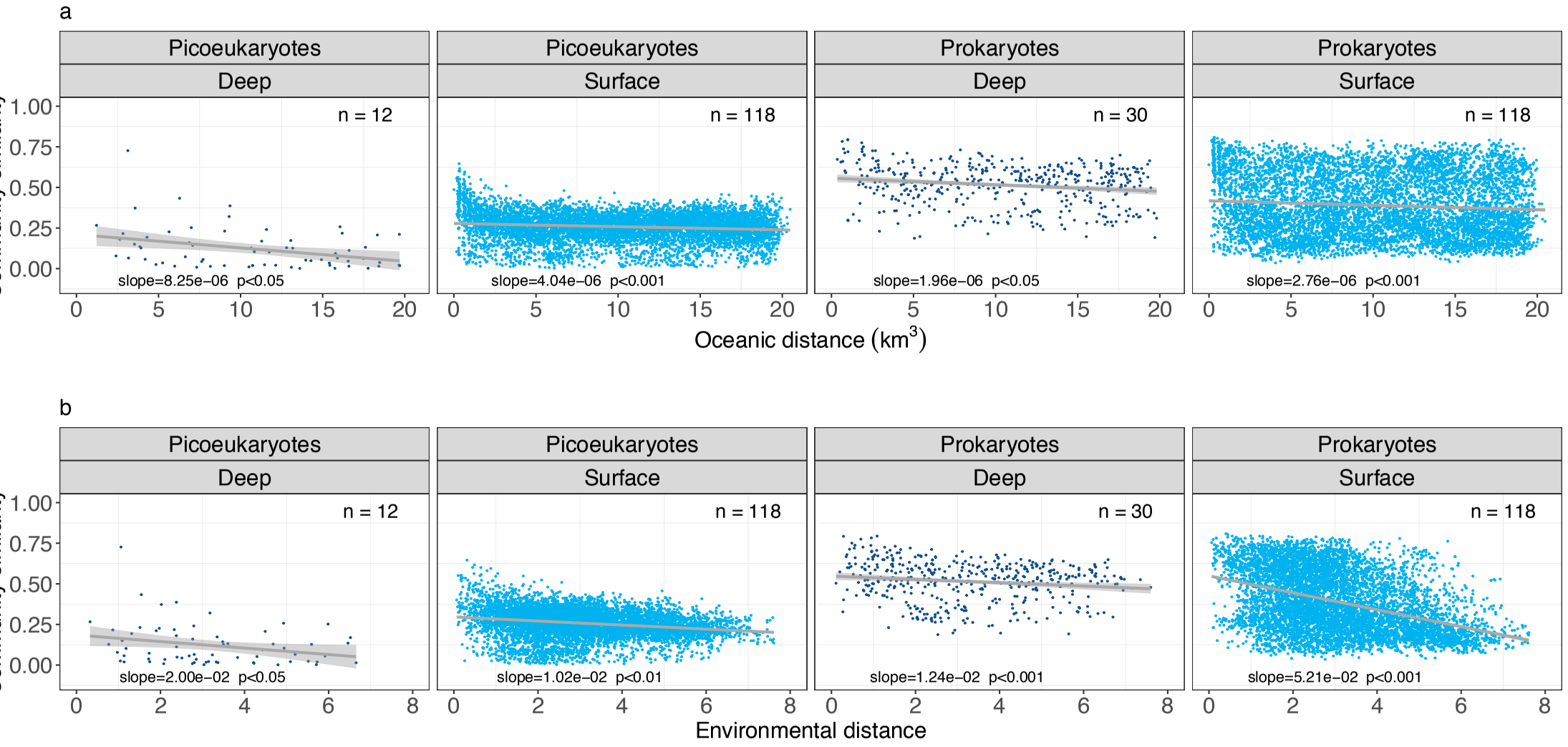

Picoeukaryotes
Prokaryotes
PAR
a_{254}
PP
SiO_4
PO_4
NO_3
Water age
T
S
O
Ocean dist
Env dist
–0.01
0.12*
–0.02
0.13
0.09
0.21
0.2*
0.22
0.12*
0.21
0.03
0.13
0.19*
0.06
0.05
0.13
0.2*
0.24*
0.13*
0.34*
0.22*
0.26*
–0.02
0.09
0.08
0.23*
0.07
0.05
0.36*
0.07
0.3*
0.07
0.06
0.19*
0.36*
0.12
0.08*
0.06
0.41*
0.09
0.08*
0.18*
0.41*
0.15*
Surface
Deep
Surface
Deep
Mantel r
0.4
0.2
0.0

Fig. 4


a
Picoeukaryotes
Deep
n = 12
slope=8.25e−06 p<0.05
Picoeukaryotes
Surface
n = 118
slope=4.04e−06 p<0.001
Prokaryotes
Deep
n = 30
slope=1.96e−06 p<0.05
Prokaryotes
Surface
n = 118
slope=2.76e−06 p<0.001
Community similarity
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Oceanic distance (km3)
b
Picoeukaryotes
Deep
n = 12
slope=2.00e−02 p<0.05
Picoeukaryotes
Surface
n = 118
slope=1.02e−02 p<0.01
Prokaryotes
Deep
n = 30
slope=1.24e−02 p<0.001
Prokaryotes
Surface
n = 118
slope=5.21e−02 p<0.001
Community similarity
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
Environmental distance