Different processes shape prokaryotic and picoeukaryotic assemblages in the sunlit ocean microbiome

Ramiro Logares[1,2*], Ina M. Deutschmann[1], Caterina. R. Giner[1], Anders K. Krabberød[2], Thomas S. B. Schmidt[3], Laura Rubinat-Ripoll[4], Mireia Mestre[1], Guillem Salazar[5], Clara Ruiz-González[1], Marta Sebastián[1,6], Colomban de Vargas[4], Silvia G. Acinas[1], Carlos M. Duarte[7], Josep M. Gasol[1,8], Ramon Massana[1]

[1] Institute of Marine Sciences (ICM), CSIC, 08003, Barcelona, Catalonia, Spain

[2] Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, N-0316 Oslo, Norway

[3] European Molecular Biology Laboratory, Meyerhofstr. 1, D-69117, Heidelberg, Germany

[4] Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7144, Adaptation et Diversité en Milieu Marin, Equipe EPEP, Station Biologique de Roscoff, 29680 Roscoff, France

[5] ETH Zurich, Institute of Microbiology, 8093 Zurich, Switzerland

[6] Oceanography and Global Change Institute, IOCAG, University of Las Palmas de Gran Canaria, ULPGC, Gran Canaria 35214, Spain.

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

[8] Centre for Marine Ecosystems Research, School of Science, Edith Cowan University, Joondalup, WA, Australia

*** Corresponding author:**
Ramiro Logares
Institute of Marine Sciences (ICM), CSIC,
Passeig Marítim de la Barceloneta, 37-49,
08003, Barcelona, Catalonia, Spain
e-mail: ramiro.logares@gmail.com
Tel: +34 93 2309500; Fax: +34 93 2309555

Keywords: microbiome, ocean, structure, picoeukaryotes, prokaryotes

ABSTRACT

The smallest members of the sunlit-ocean microbiome (prokaryotes and picoeukaryotes) participate in a plethora of ecosystem functions with planetary-scale effects. Understanding the processes determining the spatial turnover of this assemblage can help us better comprehend the links between microbiome species composition and ecosystem function. Ecological theory predicts that *selection*, *dispersal* and *drift* are main drivers of species distributions, yet, the relative quantitative importance of these ecological processes in structuring the surface-ocean microbiome is barely known. Here we quantified the role of selection, dispersal and drift in structuring surface-ocean prokaryotic and picoeukaryotic assemblages by using community DNA-sequence data collected during the global Malaspina expedition. We found that dispersal limitation was the dominant process structuring picoeukaryotic communities, while a balanced combination of dispersal limitation, selection and drift shaped prokaryotic counterparts. Subsequently, we determined the agents exerting abiotic selection as well as the spatial patterns emerging from the action of different ecological processes. We found that selection exerted via temperature had a strong influence on the structure of prokaryotic communities, particularly on species co-occurrences, a pattern not observed among communities of picoeukaryotes. Other measured abiotic variables had limited selective effects on microbiome structure. Picoeukaryotes presented a higher differentiation between neighbouring communities and a higher distance-decay when compared to prokaryotes, agreeing with their higher dispersal limitation. Finally, drift seemed to have a limited role in structuring the sunlit-ocean microbiome. The different predominance of ecological processes acting on particular subsets of the ocean microbiome suggests uneven responses to environmental change.

SIGNIFICANCE STATEMENT

The global ocean contains one of the largest microbiomes on Earth and changes on its structure can impact the functioning of the biosphere. Yet, we are far from understanding the mechanisms that structure the global ocean microbiome, that is, the relative importance of environmental *selection*, *dispersal* and random events (*drift*). We evaluated the role of these processes at the global scale, based on data derived from a circumglobal expedition and found that these ecological processes act differently on prokaryotes and picoeukaryotes, two of the main components of the ocean microbiome. Our work represents a significant contribution to understand the assembly of marine microbial communities, providing also insights on the links between ecological mechanisms, microbiome structure and ecosystem function.

INTRODUCTION

The surface ocean microbiome is a pivotal underpinning of global biogeochemical cycles, thus being crucial for the functioning of the biosphere (1-4). The smallest ocean microbes, the picoplankton, have a key role in the global carbon cycle (4), being responsible for an important fraction of the total atmospheric carbon and nitrogen fixation in the ocean (5-7), which supports ~46% of the global primary productivity (8). Oceanic picoplankton plays a fundamental role in processing organic matter by recycling nutrients and carbon to support additional production as well as by channelling organic carbon to upper trophic levels through food webs (4, 5, 9).

The ocean picoplankton includes prokaryotes (both bacteria and archaea) and tiny unicellular eukaryotes (hereafter picoeukaryotes), which feature fundamental differences in terms of cellular structure, feeding habits, metabolic diversity, growth rates and behaviour (10, 11). Even though marine picoeukaryotes and prokaryotes are usually investigated separately, they are intimately connected through biogeochemical and food web networks (12-14). Overall, given the large effects picoeukaryotes can have on the populations of prokaryotes (and vice versa), it is highly relevant to determine whether or not their communities are structured by the action of similar ecological processes.

Ecological theory explains the structure of communities by a combination of four processes: *selection*, *dispersal*, *ecological drift* and *speciation* (15-17). Selection involves deterministic reproductive differences among individuals from different or the same species as a response to biotic or abiotic conditions. Selection can act in two opposite directions, it can constrain (homogeneous selection) or promote (heterogeneous selection) the divergence of communities (18). Dispersal, the movement of organisms across space, affects microbial assemblages by incorporating individuals

originating from the regional species pool. Dispersal rates can be high (homogenising dispersal), moderate, or low [dispersal limitation] (18). Dispersal limitation occurs when species are absent from suitable habitats because potential colonizers are too far away (19). The importance of dispersal limitation increases as geographic scale increases (20). Ecological drift (hereafter *drift*) in a local community refers to random changes in species' relative abundances derived from the stochastic processes of birth, death and offspring generation (15, 17). The action of drift in a *metacommunity*, that is, local communities that are connected via dispersal of multiple species (21), may lead to neutral dynamics (20), where random dispersal is the main mechanism of community assembly. In this neutral scenario, if dispersal is not limited, local communities will tend to resemble random subsamples of the metacommunity (20). Finally, speciation is the emergence of new species by evolution (15, 17), and it will not be considered hereafter as it is expected to have a small impact in the turnover of communities that are connected via dispersal (22).

The interplay of selection, dispersal and drift may generate different microbial assemblages that could feature diverse metabolisms and ecologies (16). The action of selection (in moderate to high strength) together with moderate rates of dispersal may generate a deterministic coupling between specific environmental conditions and combinations of species, a spatial pattern known as *species sorting* (23). In contrast, high or low levels of dispersal may produce the opposite effect, that is, a decoupling between abiotic environmental conditions (i.e. selection) and species assemblages. Particularly, high dispersal rates may maintain populations in habitats to which they are maladapted (21). Inversely, low dispersal rates may preclude species from reaching suitable habitats, leading to species assemblages that become more different as the geographic distance between them increases (distance decay). Still, both selection and

dispersal limitation can generate distance decay (24). Drift is expected to cause important random effects in local community composition in cases where selection is weak and populations are small (16, 25).

Whereas global-ocean connectivity patterns reveal the importance of dispersal in structuring communities in the upper ocean (26), our understanding of the relative role of selection, dispersal and drift in structuring the global-ocean microbiome is still poor (24, 27, 28). Multiple studies in diverse environments indicate that selection has a major role in structuring prokaryotic communities (23, 24), although there is also evidence pointing to drift as having a structuring role (29, 30). Here, we examine the mechanisms shaping the sunlit global-ocean microbiome by addressing the following questions: What are the relative roles of selection, dispersal and drift in shaping assemblages of prokaryotes and picoeukaryotes? What environmental variables exert selection? What spatial patterns emerge from the action of these three processes? We hypothesize that the major organismal differences between picoeukaryotes and prokaryotes (11) should result in a different relative importance of selection, dispersal and drift in structuring their communities. Specifically, we hypothesize that the lower capacity for dormancy of picoeukaryotes (11, 31) should result in a larger dispersal limitation when compared to prokaryotes, given that dormancy may allow prokaryotes transiting through large geographic areas that are unsuitable for growth. Furthermore, given that some prokaryotes may engage into metabolic cooperation (32), an uncommon behaviour in picoeukaryotes, we expect selection to generate more co-occurrences among prokaryotes than among picoeukaryotes.

RESULTS

Processes shaping the surface global-ocean microbiome

We analysed the prokaryotes and picoeukaryotes composing this microbiome in 120 tropical and subtropical stations sampled during the Malaspina-2010 expedition (33) **[Fig. S1]** by *Illumina* High-Throughput sequencing of 16S and 18S rRNA-genes. Based on these genes, we delineated Operational Taxonomic Units (OTUs) as species proxies (see Methods). We applied an innovative methodology based on phylogenetic and species turnover (22) that allowed us to quantify the relative importance of selection, dispersal and drift (See Methods).

We found that selection, dispersal and drift played a similar role in structuring prokaryotic communities, while dispersal limitation was the dominant force structuring picoeukaryotic communities **(Fig. 1)**. Selection explained ~34% of the turnover of prokaryotic communities, and ~17% of that in picoeukaryotes **(Fig. 1)**. Heterogeneous selection had a relatively higher importance in structuring picoeukaryotes as compared to prokaryotes (~16% vs. ~9%, respectively), while homogeneous selection was more important in structuring prokaryotic (~24%) than picoeukaryotic (~1%) communities **(Fig. 1)**. Dispersal limitation was by far the most important process structuring picoeukaryotic communities (~76%), while this process had a lower importance in prokaryotes (~35%) **[Fig. 1]**. Drift explained a larger fraction of community turnover in prokaryotes (31%) than in picoeukaryotes (~6%) **[Fig. 1]**. Homogenizing dispersal had a very limited role in the structuring of the global ocean microbiome (<1% for both picoeukaryotes and prokaryotes).

Given that communities of prokaryotes and picoeukaryotes were predominantly structured by different processes, we expected both groups to present contrasting ß-diversity patterns. Accordingly, we found only moderate correlations between Bray-

Curtis and generalized UniFrac (gUniFrac) ß-diversity indices between picoeukaryotes and prokaryotes (Bray Curtis: ρ=0.58, gUniFrac: ρ=0.61, p=0.01, Mantel tests; **Fig. S2**). Rare species tend to occupy less sites than more abundant species (34), therefore communities featuring different proportions of abundant and rare species may display different spatial turnover. We found that picoeukaryotes had proportionally more regionally rare species (here defined as those with mean abundances below 0.001%) than prokaryotes (71% vs. 48% of the species respectively) **[Table S1, Fig. S3]**. This is consistent with the observation that picoeukaryotes had more restricted species distributions (i.e., occurring in <20% of the communities) than prokaryotes [95% vs. 88% of the species respectively] (**Table S2; Fig. S3**).

Selection acting on the microbiome

We investigated the abiotic agents exerting selection in the ocean microbiome by analysing the compositional differences between communities (ß-diversity) together with a set of environmental variables considered in the *Meta-119* dataset (See Supplementary Information). We used different ß-diversity indices (Bray-Curtis, $TINA_w$, $PINA_w$, gUniFrac; See Methods), as each captures distinct features of community differentiation. Water temperature was the most important driver of selection on prokaryotes **(Fig. 2)**. Furthermore, water temperature appeared to affect prokaryotic associations, given that the association-aware ß-diversity index $TINA_w$ (35) explained ~50% of community variance (PERMANOVA R^2) **[Fig. 2]**, while other ß-diversity indices tested that do not consider species associations explained considerably lower proportions **(Fig. 2)**. In contrast, temperature had limited effects on picoeukaryotic species associations **(Fig. 2)**. Our results were further confirmed by

independent data from the global surface-ocean collected during the TARA Oceans expedition (36), as both the Malaspina and TARA Oceans datasets displayed stronger positive correlations between $TINA_w$ and water-temperature differences in prokaryotes than in picoeukaryotes **[Fig. 3]**. This indicates that locations with similar temperatures feature co-occurring prokaryotic species, with this pattern disappearing as the temperature difference between stations increases.

Further analyses, exploring 17 environmental variables from 57 stations **(**Supplementary Information**, Fig. S4)**, showed that fluorescence (a proxy for Chlorophyll *a* concentration) explained 31% of $PINA_w$-based prokaryotic community variance (PERMANOVA R^2), while it was not significant for picoeukaryotes **(Fig. S5**; $PINA_w$ is a phylogeny-based ß-diversity index, See Methods**)**. The remaining combinations of environmental variables and ß-diversity metrics explained a minor fraction of community variance, suggesting that abiotic selection, at the whole microbiome level, operates via a very limited set of environmental variables, largely temperature.

The finding that selection via temperature influences species associations particularly in prokaryotes suggests that prokaryotes and picoeukaryotes may show different patterns of species co-occurrences and co-exclusions in association networks (37). We found that prokaryotes were more associated between themselves than picoeukaryotes in networks considering co-occurrences and co-exclusions as well as in networks including only co-occurrences **(Fig. S6)**. Specifically, in networks including both co-occurrences and co-exclusions, prokaryotes featured ~33% of connected species (i.e. prokaryotic species with at least one association to another prokaryotic species) and an average number of associations per species (i.e. average degree) of ~14, while picoeukaryotes displayed ~17% of connected species and an average degree of ~8

231 **(Table S3**; **Fig. S6)**. Networks including co-occurrences only displayed similar patterns
232 **(Table S3**; **Fig. S6)**. The prokaryotic network was more modular, containing a higher
233 number of highly-connected clusters of species than the picoeukaryotic network **(Table**
234 **S3)**.

235 The previous results were supported by analyses using the Maximal Information
236 Coefficient [MIC], which quantifies a wide array of functional and non-functional,
237 linear and non-linear, associations (38). MIC results indicated that prokaryotes had
238 more associations between themselves than picoeukaryotes **(Table S4)**, a pattern that
239 was also observed in other data from the upper global ocean collected during the TARA
240 Ocean expedition **(Table S5)**. Most associations detected by MIC were non-linear
241 [defining nonlinearity as MIC-ρ^2 >0.2; (38)] **[Table S4** & **S5]**, pointing to complex
242 associations that may be missed by classic correlation analyses, which evaluate linear
243 relationships.

244

245 ***Selection acting on species***

246 The potential effects of selection on single species was evaluated by determining
247 their individual correlations with multiple environmental variables using the MIC (38).
248 In these analyses, temperature was the variable with the highest number of associated
249 prokaryotic species (1.7%) when considering a MIC threshold ≥0.4, representing ~17%
250 of the total estimated species abundance, while picoeukaryotic species displayed a
251 considerable smaller proportion (~0.3% of the species representing ~5% of the
252 estimated species abundance) [**Fig. S7]**. Picoeukaryotic and prokaryotic species also
253 had associations with oxygen, conductivity and salinity **(Fig. S7)**, which co-vary with
254 temperature. The remaining environmental variables tested had limited associations
255 with individual species **(Fig. S7)**, thus agreeing with our previous results suggesting that

256 selection on the surface ocean microbiome operates via a limited set of environmental
257 variables, with a dominant role for temperature. Prokaryotes featured proportionally
258 more individual-species associations with environmental parameters than
259 picoeukaryotes **(Fig. S7**), thus pointing to a stronger abiotic selection pressure on
260 prokaryotes than on picoeukaryotes in the surface global ocean. Our results were further
261 validated by analyses using data from the global TARA Oceans cruise, which indicated
262 that prokaryotic species were associated predominantly with temperature and oxygen,
263 while small unicellular eukaryotes had limited associations to multiple variables
264 (Temperature, Salinity, Oxygen, Nitrate & Chlorophyll; **Table S6**).
265
266 ***Dispersal limitation***
267 Our quantifications indicated that dispersal limitation was almost twice as
268 important in structuring picoeukaryotic than prokaryotic communities **(Fig. 1)**.
269 Environmental conditions between pairs of adjacent stations over the trajectory of the
270 cruise, typically separated by 250-500 km, are generally comparable (i.e. selective
271 differences between stations tend to be low in the tropical and subtropical ocean).
272 Therefore, compositional differences between neighbouring communities could
273 manifest dispersal limitation. Following these premises, we analysed the change in
274 picoeukaryotic and prokaryotic community composition along the trajectory of the
275 cruise by comparing each community to the one sampled immediately before in a
276 sequential manner (i.e. sequential β-diversity) **[Fig. 4]**. Both picoeukaryotic and
277 prokaryotic communities displayed variable amounts of sequential β-diversity **(Fig. 4,**
278 **Panels A and B)**, although picoeukaryotes featured, on average, a higher sequential β-
279 diversity than prokaryotes **(Fig. 4**, **Panel C)**. This is concordant with the overall mean

β-diversity, which was significantly higher for picoeukaryotes than for prokaryotes **(Fig. S8)**.

Simultaneous action of selection and dispersal limitation

When geographic distance is correlated with environmental variation, spatial community variance may be the manifestation of both selection and dispersal limitation. We analysed community variance associated to different marine biogeographic provinces, as defined by Longhurst (39) based on nutrient concentration, structure of the water column, wind regimes, satellite-derived primary production and composition of abundant phytoplankton species. After removing the effects of the most important environmental variables that were correlated with these geographic regions and that likely exert selection (e.g. temperature), we found that differences among Longhurst provinces still accounted for ~20% of picoeukaryotic community variance when using Bray-Curtis and gUniFrac β-diversity indices (PERMANOVA R^2) **[Fig. 2]**. Likewise, Longhurst provinces explained ~20-25% of prokaryotic community variance with all tested β-diversity indices (PERMANOVA R^2) **[Fig. 2]**. The variability in community composition associated to these provinces most likely represent dispersal limitation, even though abiotic or biotic selection exerted by unmeasured variables cannot be ruled out. ß-diversity in picoeukaryotes and prokaryotes displayed positive correlations with distance (i.e. distance decay) predominantly within 1,000 km **(Fig. 5)**, although correlations were weaker in prokaryotes than in picoeukaryotes, being consistent with a higher dispersal limitation in picoeukaryotes than in prokaryotes **(Fig. 1)**.

Selection and dispersal limitation may operate more strongly in geographic areas considered ecological boundaries, for example, due to strong physicochemical change in the seawater, leading to abrupt changes in microbiome composition. We identified a

total of 14 communities where sequential β-diversity displayed abrupt changes, with 11 of them coinciding for both picoeukaryotes and prokaryotes **[Fig. 4, Panels A & B]**. The Local Contribution to Beta Diversity (LCBD) index (40) **(Fig. 4, Panel D)** indicated that ~22% of both picoeukaryotic and prokaryotic communities (26 stations each, totaling 36 stations) contributed the most to the β-diversity, with 16 communities coinciding for both prokaryotes and picoeukaryotes ($p<0.05$; **Fig. 4, Panel D, Table S7**). In addition, 8 of the 36 stations featuring a significant LCBD were also identified as zones of abrupt community change in sequential Bray Curtis analyses **(Table S7)**. These zones featuring abrupt community change in both prokaryotes and picoeukaryotes point to selection or dispersal acting simultaneously and strongly upon both life's domains.

DISCUSSION

Our assessment of the tropical and sub-tropical sunlit fraction of the global-ocean microbiome during the Malaspina 2010 Circumnavigation Expedition indicated that dispersal limitation is the dominant process structuring picoeukaryotic communities, while selection, dispersal and drift have a balanced importance in structuring prokaryotic communities. These results summarise the general action of ecological processes at the microbiome level and cannot be extended to every single species or group; for example, some picoeukaryotes display cosmopolitan distributions (41). Our results also reflect the action of ecological processes in the tropical and subtropical sunlit ocean. Therefore, considering other zones, such as the polar oceans, could modify the relative importance of these processes. To determine the action of selection, we used the principle indicating that phylogenetically closely related taxa tend to be ecologically similar (and vice versa). This principle was supported by data from prokaryotes and unicellular eukaryotes (42-45). Yet, there can be exceptions (46), and failure to detect selection could inflate the estimates of dispersal limitation in our methodology. We consider that dispersal limitation in picoeukaryotes was not inflated in our analyses, as picoeukaryotes generally displayed more limited spatial distributions than prokaryotes and a higher sequential β-diversity. Drift was pragmatically analysed as the compositional variation between communities that does not differ from random community assembly, thus representing neutral metacommunity dynamics (20, 22, 44). Hence, our quantifications of drift do not reflect the impact of random demography in single communities. Lastly, our estimates of the importance of ecological mechanisms consider taxa with high or moderate abundances, which typically carry biogeographic information (47), thus, our results do not reflect the processes shaping the rare biosphere (48).

343 Selection, which is known to have an important role in structuring prokaryotic
344 communities (23, 24), explained a higher proportion of community turnover in
345 prokaryotes (~34% of the turnover) than in picoeukaryotes (~17% of the turnover). This
346 modest role of selection in structuring the tropical and subtropical sunlit-ocean
347 microbiome is consistent with the weak environmental gradients characterizing this
348 habitat. In other habitats featuring a high selective pressure, such as Antarctic
349 waterbodies that display a strong salinity gradient, the role of selection in structuring
350 bacteria has been reported to be much higher, accounting for up to ~70% of the
351 community turnover (49). The measured relative importance of selection is also a
352 consequence of the global scale of our survey. For example, in comparatively small
353 marine areas, where dispersal limitation is expected to be low (19), the relative
354 importance of selection could increase. In surface waters of the East China Sea it was
355 found that selection was ~40% more important than dispersal limitation in structuring
356 bacterial communities (50), while in our study, selection and dispersal limitation had a
357 similar importance in structuring prokaryotes. The same study (50) also found that
358 selection was ~40 times more important than dispersal limitation in structuring
359 communities of microbial eukaryotes. In contrast, our global assessment yields dispersal
360 limitation to be ~5 times larger than selection in structuring picoeukaryotic
361 communities.
362 Heterogeneous selection was more important in structuring picoeukaryotic
363 (~16%) than prokaryotic communities (~9%), while homogeneous selection was more
364 important in prokaryotic (~24%) than in picoeukaryotic communities (~1%).
365 Homogeneous environmental conditions should lead to homogeneous selection,
366 constraining community divergence, while heterogeneous environmental conditions
367 should promote community divergence (18). Our results suggest that prokaryotes and

368 picoeukaryotes in the same marine habitats respond differently to the same
369 environmental conditions. Thus, selection would be preventing community divergence
370 in prokaryotes while promoting it in picoeukaryotes. The fundamental cellular
371 differences between prokaryotes and picoeukaryotes (10, 11) may determine such
372 different responses to the same environmental heterogeneity. For example, comparable
373 environmental heterogeneity could select for a few species featuring wide
374 environmental tolerance or several species, which are adapted to narrow environmental
375 conditions.

376 Previous studies indicate that temperature is one of the most important variables
377 exerting selection on ocean prokaryotes (51-59). Earlier work (51) reported strong
378 correlations between prokaryotic ocean-microbiome composition and temperature, and
379 weak correlations with nutrients, consistent with our results. Less is known about the
380 effects of temperature on the community structuring of ocean picoeukaryotes, which
381 according to our results are minor. Yet, it is known that temperature affects the
382 distributions of MAST-4, a lineage of ubiquitous marine heterotrophic flagellates (41),
383 suggesting that the effects of temperature on small eukaryotes could be group specific.
384 Interestingly, our results suggest that selection, operating via temperature, affects
385 prokaryotic taxa co-occurrences, having limited effects on picoeukaryotic co-
386 occurences. In prokaryotes, the β-diversity associated to temperature explained by
387 $TINA_w$ (~50%) was substantially higher than Bray Curtis (~15%), reflecting the
388 importance of considering co-occurrences, as in $TINA_w$, to understand community
389 structure. These results suggest that temperature-driven selection determines the species
390 that can grow in different locations, yet in each site, species relative abundances and
391 presence-absence may vary due to local stochasticity (60, 61).

392 To what extent dispersal limitation affects the distribution of microbes is a
393 matter of debate (62-65). In surface open-ocean waters, prokaryotes typically display
394 abundances of 10^6 cells/mL, while picoeukaryotes normally have abundances of 10^3
395 cells/mL (66). Due to random dispersal alone, the more abundant prokaryotes are
396 expected to be distributed more thoroughly than the less abundant picoeukaryotes (34),
397 which is consistent with our findings. Furthermore, the absence of taxa from suitable
398 habitats that are separated by large distances is expected to be more pronounced in
399 picoeukaryotes than in prokaryotes. Still, our analyses compare actual species
400 distributions against those that would be expected by chance when considering species
401 abundances, suggesting that species-abundance is not the main reason of dispersal
402 limitation in picoeukaryotes.
403 Several studies in aquatic unicellular eukaryotes point to restricted dispersal and
404 endemism (64, 67-69), while others indicate the opposite (41, 70-72). This could reflect
405 different dispersal capabilities among small eukaryotic taxa (64, 73) and the generation
406 of dormant cysts in some species, such as in diatoms, dinoflagellates and
407 coccolitophorids (74, 75), which may increase dispersal rates. Cyst formation has not
408 been reported yet for picoeukaryotes (11) and this may partially explain their limited
409 dispersal. Regarding prokaryotes, other studies considering large geographic scales
410 indicated that dispersal limitation has a small influence in the structure of marine
411 communities (51, 52, 76), which is coherent with our results. Dormancy in prokaryotes
412 seems to be more common than in picoeukaryotes (11, 31), and this may allow the
413 former to disperse more thoroughly by reducing their metabolisms when moving
414 through unfavorable habitats (77). In sum, our quantifications of dispersal limitation
415 agree, in general terms, with known trends in both picoeukaryotes and prokaryotes.

The importance of drift or neutral dynamics (20) in structuring microbial communities is also a matter of debate (23, 78). Our results indicate that drift has a modest role in structuring the sunlit-ocean microbiome, being higher in prokaryotic (~31% of the turnover) than in picoeukaryotic communities (~6% of the turnover). Another study also found a higher importance of drift in determining the community structure of bacteria when compared with phytoplankton populating freshwater and brackish habitats (79). In contrast, drift was the prevalent community-structuring mechanism in unicellular eukaryotes populating lakes in a relatively small geographic area that features a strong salinity gradient, having a low importance for the structuring of prokaryotic communities (49). Likely, the relative importance of drift in structuring prokaryotes or unicellular eukaryotes is dependent on the selective strength of specific habitats, the occurrence of adaptive processes (49) or barriers to dispersal.

When geographic distance is correlated with environmental variation, a decrease in community similarity with distance (distance decay) can be the manifestation of both selection and/or dispersal limitation (24). Distance decay has been evidenced in diverse studies focusing on the surface and deep ocean microbiome (52, 80, 81). Yet, different to most previous studies, we have quantified the role of selection and dispersal limitation in structuring the surface ocean microbiome, and we can use this information to interpret the measured distance decay. As our quantifications indicated a strong dispersal limitation in picoeukaryotes, it is likely that this process explains the measured distance decay. In contrast, the distance decay observed in prokaryotes could be the outcome of both selection and dispersal limitation, as both presented comparable structuring roles. The amount of community variance in prokaryotes and picoeukaryotes associated to provincialism (i.e. Longhurst oceanic regions) likely reflects dispersal limitation, since the effects of important environmental variables were removed during

441 the analyses. Interestingly, another study investigating surface marine bacteria along
442 ~12,000 km in the Atlantic Ocean found that provincialism explained an amount of
443 community variance comparable to our results (52). Furthermore, a study of the
444 eukaryotic microbiome in the sunlit global-ocean indicated that provincialism
445 (considered in terms of ocean basins) was one of the most important variables
446 explaining community structure (67). In the light of our findings, we consider that
447 results from the previous study (67) manifest, to a large extent, dispersal limitation.
448 In the surface ocean, drastic changes in species composition across space may
449 point to strong changes in abiotic selection, as expected to occur across oceanographic
450 fronts (82, 83), or high immigration from the surface or deeper water layers. We
451 identified 14 stations featuring abrupt changes in prokaryotic or picoeukaryotic
452 community composition as well as 36 stations with a "unique" species composition
453 according to the *Local Contribution to Beta Diversity* analysis (40). Several of these
454 stations coincided for both picoeukaryotes and prokaryotes. Some of these areas
455 correspond to nutrient-rich coastal zones (the South African Atlantic coast and the
456 South Australia Bight) or potential upwelling zones, such as the Equatorial Pacific and
457 Atlantic as well as the Costa Rica Dome. This agrees with a scenario including strong
458 selective changes or immigration from deep water layers into the surface, which affects
459 both prokaryotes and picoeukaryotes.
460 Altogether, our results represent a significant contribution towards
461 understanding the structure of the sunlit-ocean microbiome by connecting patterns to
462 underlying ecological processes (24). Our findings indicate that comprehending the
463 idiosyncrasies of the main components of microbiomes is needed in order to attain a
464 holistic understanding of their structures and ecologies. In particular, our results suggest
465 that the structure of surface-ocean prokaryotic communities could be more susceptible

466 to global warming than that of picoeukaryotic communities. Prokaryotes represent an
467 important fraction of the total microbial biomass in the ocean (84, 85), and they have
468 fundamental roles for ecosystem function (1, 2). Therefore, understanding the specific
469 effects of temperature in their distributions and community metabolism (27) represents
470 an important challenge, which our results contribute to address.

METHODS

Sample collection

Surface waters (3 m depth) from a total of 120 globally-distributed stations located in the tropical and sub-tropical global ocean **(Fig. S1)** across ~50,000 km, were sampled from December 2010 to July 2011 as a part of the Malaspina-2010 expedition (33, 86). Water samples were obtained with a 20 L Niskin bottle deployed simultaneously to a CTD profiler that included sensors for conductivity, temperature, oxygen, fluorescence and turbidity. About 12 L of seawater were sequentially filtered through a 20 µm nylon mesh, followed by a 3 µm and 0.2 µm polycarbonate filters of 47 mm diameter (Isopore, Millipore). Only the smallest size-fraction [0.2 -3 µm, here called "picoplankton"; see (5)] was used in downstream analyses. Samples for inorganic nutrients (NO_3^-, NO_2^-, PO_4^{3-}, SiO_2) were collected from the Niskin bottles and measured spectrophotometrically using an Alliance Evolution II autoanalyzer (87). Chlorophyll measurements were obtained from Estrada et al. (86). In specific samples nutrient concentrations were estimated using the World Ocean Database (88-91) due to issues with measurements. Since not all environmental parameters were available for all stations, two contextual datasets were generated: *Meta-119*, including 119 stations **(Fig. S1)**, 5 environmental parameters and 5 spatial features and *Meta-57* **(Fig. S4)**, including 57 stations and 17 environmental parameters. See Supplementary Information for further details.

DNA extraction, 18S- & 16S-rRNA amplicon sequencing and bioinformatic analyses

DNA was extracted using a standard phenol-chloroform protocol (92). Both the 18S and 16S rRNA-genes were amplified from the same DNA extracts. The

hypervariable V4 region of the 18S (~380 bp) was amplified with the primers TAReukFWD1 and TAReukREV3 (93), while the hypervariable V4-V5 (~400bp) region of the 16S was amplified with the primers 515F-Y - 926R (94), which target both Bacteria and Archaea. Amplicon libraries were then paired-end sequenced on an *Illumina* MiSeq platform (2x250bp) at the Research and Testing Laboratory facility (Lubbock, TX, USA; http://www.researchandtesting.com/).

Reads were processed following an in-house pipeline (95). Briefly, raw reads were corrected using BayesHammer (96) following Schirmer *et al*. (97). Corrected reads were merged with PEAR (98) and sequences >200bp were quality-checked and de-replicated using USEARCH (99). Operational Taxonomic Units (OTUs) were delineated at 99% similarity using UPARSE V8.1.1756 (100). To obtain OTU abundances, reads were mapped back to OTUs at 99% similarity. Chimera check and removal was performed both *de novo* and using the SILVA reference database (101). After our stringent quality control (see Supplementary Information), a total of 42,505 picoeukaryotic and 10,158 prokaryotic OTUs were obtained. Taxonomic assignment of picoeukaryotic and prokaryotic OTUs was generated by BLASTing (102) representative sequences against different reference databases. Metazoan, Charophyta, nucleomorphs, Chloroplast and mitochondrial OTUs were removed from the OTU tables. See more details in Supplementary Information and **Table S8**. Sequences are publicly available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena; accession numbers PRJEB23913 [18S] & PRJEB25224 [16S]).

In specific analyses, we considered publicly-available data from the TARA Oceans expedition (36). We selected data from surface communities only, including 41 samples (40 stations) for pico-nano eukaryotes (0.22-3 μm [1 sample] and 0.8-5 μm [40 samples]; 18S-V9 rDNA amplicon data) (67) as well as 63 stations for prokaryotes

[picoplankton, 0.22-3 μm (45 samples) and 0.22-1.6 μm (18 samples); 16S rDNA, miTags] (51).

General analyses and phylogenetic inferences

Both picoeukaryotic and prokaryotic datasets were sub-sampled to 4,060 reads per sample using *rrarefy* in *Vegan* (103), resulting in sub-sampled tables containing 18,881 picoeukaryotic and 7,025 prokaryotic OTUs. All OTUs with mean relative abundances above 0.1% and below 0.001% were defined as regionally abundant or rare respectively (104). Phylogenetic trees were constructed for both the 16S and 18S datasets using OTU-representative sequences. Reads were aligned against an aligned SILVA template using *mothur* (105). Afterwards, poorly aligned regions or sequences were removed using *Trimal* (106). A phylogenetic tree was inferred using *FastTree* v2.1.9 (107). Most analyses were performed in the R statistical environment (108) using *adespatial* (109), *APE* (110), *ggplot2* (111), *gUniFrac* (112), *Maps* (113), *Mapplots* (114), *Picante* (115) and *Vegan* (103). See further details in Supplementary Information.

Quantification of selection, dispersal and drift

Selection, dispersal and drift were quantified using the approach proposed by Stegen et al. (22). This methodology consists of two main steps: the first uses phylogenetic turnover to infer the action of selection and the second uses OTU turnover to infer the action of dispersal and drift. Phylogenetic turnover was measured by calculating the abundance-weighted β-mean nearest taxon distance (βMNTD), which quantifies the mean phylogenetic distances between the evolutionary closest OTUs in two communities. βMNTD values can be larger, smaller or equal to the values expected

546 when selection is not affecting community turnover (that is, expected under a random
547 distribution [null model]). βMNTD values higher than expected by chance indicate that
548 communities are under heterogeneous selection (18). In contrast, βMNTD values which
549 are lower than expected by chance indicate that communities are experiencing
550 homogeneous selection. Null models were constructed using 999 randomizations as in
551 Stegen et al. (22). Differences between the observed βMNTD and the mean of the null
552 distribution are denoted as β-Nearest Taxon Index (βNTI), with | βNTI | > 2 being
553 considered as significant departures from random phylogenetic turnover, pointing to the
554 action of selection.

555 The second step of this method calculates whether the observed β-diversity,
556 based on OTU turnover, could be generated by drift (i.e. chance) or other processes. For
557 this, we calculated the Raup-Crick metric (116) using Bray-Curtis dissimilarities
558 [hereafter RC_{bray}], following Stegen et al. (22). RC_{bray} compares the measured β-
559 diversity against the β-diversity that would be obtained if drift was driving community
560 turnover (that is, under random community assembly). The randomization was run
561 9,999 times and only OTUs with >1,000 reads over the entire dataset were considered in
562 order to prevent any bias due to potential under sampling. RC_{bray} values between -0.95
563 and +0.95 point to a community assembly governed by drift. On the contrary, RC_{bray}
564 values > +0.95 or < -0.95 indicate that community turnover is driven by dispersal
565 limitation or homogenizing dispersal respectively (116). The previous framework was
566 applied as following: First, we determined the proportion of community pairwise
567 comparisons displaying a |βNTI| > 2, which points to the action of selection.
568 Subsequently, for the pairwise comparisons that did not indicate the action of selection,
569 we calculated the proportion of total comparisons that could be assigned to dispersal

570 limitation, homogenizing dispersal or drift according to their RC_{bray} values. See further
571 details in Supplementary Information.
572
573 ***Estimation of interaction-adjusted indices***
574 Taxa INteraction-Adjusted (TINA) and Phylogenetic INteraction Adjusted
575 (PINA) indices were estimated following Schmidt et al. (35) TINA is based on taxa co-
576 occurrences while PINA considers phylogenetic similarities (35). In particular, TINA
577 quantifies β-diversity as the average interaction strength between all taxa in different
578 samples. Thus, communities which are identical or include taxa which are perfectly
579 associated will give TINA values of 1. On the other hand, TINA values will approach
580 0.5 in communities sharing no taxa or having neutral associations, and approach 0 if
581 taxa display high avoidance. Dissimilarity matrices were generated as 1-TINA. Full
582 picoeukaryotic and prokaryotic subsampled OTU tables were used to calculate the
583 abundance-weighted $TINA_w$ and $PINA_w$.
584
585 ***Associations between taxa and environmental parameters***
586 We analysed whether OTUs had differential associations with environmental
587 parameters as well as between themselves using different algorithms. Firstly, we used
588 the Maximum Information Coefficient (MIC) which captures diverse relationships
589 (including non-linear and non-functional) between two pairs of variables (38). The
590 Malaspina dataset consisted of 119 stations and 17 environmental parameters (See
591 Supplementary Information for extra details). In the TARA Oceans dataset, prokaryotes
592 were analysed across 63 surface stations (including 8 environmental parameters), while
593 microbial eukaryotes were analysed across 40 surface stations (including 6
594 environmental parameters) [see Supplementary Information]. In both datasets, MIC

595 analyses were run using CV=0.5, B=0.6, and statistically significant relationships with
596 MIC ≥0.4 (Malaspina) or MIC ≥0.5 (TARA) were considered; significance was
597 assessed using precomputed p-values (38). Non-linear associations were defined as
598 MIC-ρ^2 >0.2 (38). Secondly, we constructed association networks with the Malaspina
599 dataset considering OTUs with >100 reads using *SparCC* (117) as implemented in
600 *FastSpar* (118). To determine correlations, *FastSpar* was run with 1,000 iterations,
601 including 1,000 bootstraps to infer p-values. We used OTU associations with absolute
602 correlation scores >0.3 and p<0.01. Networks were visualized with *Cytoscape* (119) and
603 their properties determined using *igraph* (120).
604

ACKNOWLEDGEMENTS

We thank all scientists and crews from the Malaspina-2010 expedition. RL was supported by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). IMD was supported by an ITN-SINGEK fellowship (ESR2-EU-H2020-MSCA-ITN-2015, Grant Agreement 675752) and CRG by a Juan de la Cierva (IJCI-2015-23505, MINECO, Spain) fellowship. This work was supported by the projects Malaspina-2010 Expedition (CSD2008-00077, MINECO, Spain), INTERACTOMICS (CTM2015-69936-P, MINECO, Spain), REMEI (CTM2015-70340-P, MINECO, Spain) and MicroEcoSystems (240904, RCN, Norway). Bioinformatics analyses were performed at the MARBITS platform of the Institut de Ciències del Mar (ICM; http://marbits.icm.csic.es) as well as in MareNostrum (Barcelona Supercomputing Center) via grants obtained from the Spanish Network of Supercomputing (RES) to RL.

REFERENCES

1. Falkowski P (2012) The power of plankton. *Nature* 483(7387):S17-20.

2. Falkowski PG, Fenchel T, & Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320(5879):1034-1039.

3. Pedrós-Alió C, Acinas SG, Logares R, & Massana R (2018) Marine microbial diversity as seen by high throughput sequencing. *Microbial Ecology of the Oceans*, eds Gasol JM & Kirchman DL (Wiley-Blackwell, Hoboken, New Jersey), p 592.

4. Worden AZ*, et al.* (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347(6223):1257594.

5. Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annual review of microbiology* 65:91-110.

6. Jardillier L, Zubkov MV, Pearman J, & Scanlan DJ (2010) Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* 4(9):1180-1192.

7. Li WKW (1994) Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnology and Oceanography* 39(1):169-175.

8. Field CB, Behrenfeld MJ, Randerson JT, & Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281(5374):237-240.

9. del Giorgio PA & Duarte CM (2002) Respiration in the open ocean. *Nature* 420(6914):379-384.

10. Keeling PJ & Del Campo J (2017) Marine Protists Are Not Just Big Bacteria. *Curr Biol* 27(11):R541-R549.

11. Massana R & Logares R (2013) Eukaryotic versus prokaryotic marine picoplankton ecology. *Environ Microbiol* 15(5):1254-1261.

12. Massana R (2009) Picoeukaryotes. *Encyclopedia of Microbiology*, ed Schaechter M (Elsevier, Oxford), pp 674-688.

13. Seymour JR, Amin SA, Raina JB, & Stocker R (2017) Zooming in on the phycosphere: the ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* 2:17065.

14. Jürgens K & Massana R (2008) *Protistan grazing on marine bacterioplankton* (Wiley-Blackwell, Hoboken, New Jersey) 2nd Ed pp 383-441.

15. Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85(2):183-206.

16. Nemergut DR, *et al.* (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* 77(3):342-356.

17. Vellend M (2016) *The theory of ecological communities* (Princeton University Press, Princeton) pp ix, 229 pages.

18. Zhou J & Ning D (2017) Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiol Mol Biol Rev* 81(4):e00002-00017.

19. Heino J, *et al.* (2015) Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwater Biology* 60(5):845-869.

20. Hubbell SP (2001) *A unified neutral theory of biodiversity and biogeography* (Princeton University Press, Princeton, NJ).

21. Holyoak M, Leibold MA, & Holt RD (2005) *Metacommunities: Spatial Dynamics and Ecological Communities* (The University of Chicago Press, Chicago).

22. Stegen JC, *et al.* (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* 7(11):2069-2079.

23. Lindström ES & Langenheder S (2012) Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* 4:1-9.

24. Hanson CA, Fuhrman JA, Horner-Devine MC, & Martiny JB (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature reviews. Microbiology* 10(7):497-506.

25. Fisher CK & Mehta P (2014) The transition between the niche and neutral regimes in ecology. *Proc Natl Acad Sci U S A* 111(36):13111-13116.

26. Villarino E, *et al.* (2018) Large-scale ocean connectivity and planktonic body size. *Nature communications* 9(1).

27. Graham EB, *et al.* (2016) Coupling Spatiotemporal Community Assembly Processes to Changes in Microbial Metabolism. *Front Microbiol* 7:1949.

28. Graham EB, *et al.* (2016) Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? *Front Microbiol* 7:214.

29. Ofiteru ID, *et al.* (2010) Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci U S A* 107(35):15345-15350.

30. Sloan WT, *et al.* (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 8(4):732-740.

31. Jones SE & Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci U S A* 107(13):5881-5886.

692 32. Morris BE, Henneberger R, Huber H, & Moissl-Eichinger C (2013) Microbial
693 syntrophy: interaction for the common good. *FEMS Microbiol Rev* 37(3):384-
694 406.

695 33. Duarte CM (2015) Seafaring in the 21St Century: The Malaspina 2010
696 Circumnavigation Expedition. *Limnology and Oceanography Bulletin* 24(1):11-
697 14.

698 34. Gaston KJ*, et al.* (2000) Abundance–occupancy relationships. *Journal of*
699 *Applied Ecology* 37(s1):39-59.

700 35. Schmidt TS, Matias Rodrigues JF, & von Mering C (2017) A family of
701 interaction-adjusted indices of community similarity. *ISME J* 11(3):791-807.

702 36. Karsenti E*, et al.* (2011) A holistic approach to marine eco-systems biology.
703 *PLoS biology* 9(10):e1001177.

704 37. Faust K & Raes J (2012) Microbial interactions: from networks to models. *Nat*
705 *Rev Microbiol* 10(8):538-550.

706 38. Reshef DN*, et al.* (2011) Detecting novel associations in large data sets. *Science*
707 334(6062):1518-1524.

708 39. Longhurst AR (2007) *Ecological Geography of the Sea* (Academic Press).

709 40. Legendre P & De Caceres M (2013) Beta diversity as the variance of community
710 data: dissimilarity coefficients and partitioning. *Ecol Lett* 16(8):951-963.

711 41. Rodriguez-Martinez R, Rocap G, Salazar G, & Massana R (2013) Biogeography
712 of the uncultured marine picoeukaryote MAST-4: temperature-driven
713 distribution patterns. *ISME J* 7(8):1531-1543.

714 42. Pyron RA, Costa GC, Patten MA, & Burbrink FT (2015) Phylogenetic niche
715 conservatism and the evolutionary basis of ecological speciation. *Biol Rev Camb*
716 *Philos Soc* 90(4):1248-1262.

717 43. Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the
718 relationship between phylogenetic relatedness and ecological similarity among
719 species. *Ecol Lett* 11(10):995-1003.

720 44. Stegen JC, Lin X, Konopka AE, & Fredrickson JK (2012) Stochastic and
721 deterministic assembly processes in subsurface microbial communities. *ISME J*
722 6(9):1653-1664.

723 45. Violle C, Nemergut DR, Pu Z, & Jiang L (2011) Phylogenetic limiting similarity
724 and competitive exclusion. *Ecol Lett* 14(8):782-787.

725 46. Logares R*, et al.* (2007) Phenotypically different microalgal morphospecies with
726 identical ribosomal DNA: A case of rapid adaptive evolution? *Microbial*
727 *Ecology* 53(4):549-561.

728 47. Ser-Giacomi E, *et al.* (2018) Ubiquitous abundance distribution of non-
729 dominant plankton across the global ocean. *Nat Ecol Evol*.

730 48. Pedrós-Alió C (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* 4:449-466.

731 49. Logares R, *et al.* (2018) Contrasting prevalence of selection and drift in the
732 community structuring of bacteria and microbial eukaryotes. *Environ Microbiol*.

733 50. Wu W, *et al.* (2018) Contrasting the relative importance of species sorting and
734 dispersal limitation in shaping marine bacterial versus protist communities.
735 *ISME J* 12(2):485-494.

736 51. Sunagawa S, *et al.* (2015) Structure and function of the global ocean
737 microbiome. *Science* 348(6237):1261359.

738 52. Milici M, *et al.* (2016) Bacterioplankton Biogeography of the Atlantic Ocean: A
739 Case Study of the Distance-Decay Relationship. *Front Microbiol* 7:590.

740 53. Delmont TO, *et al.* (2017) The global biogeography of amino acid variants
741 within a single SAR11 population is governed by natural selection. *bioRxiv*.

742 54. Johnson ZI, *et al.* (2006) Niche partitioning among Prochlorococcus ecotypes
743 along ocean-scale environmental gradients. *Science* 311(5768):1737-1740.

744 55. Delmont TO & Eren AM (2018) Linking pangenomes and metagenomes: the
745 Prochlorococcus metapangenome. *PeerJ* 6:e4320.

746 56. Simon M, Glöckner FO, & Amann R (1999) Different community structure and
747 temperature optima of heterotrophic picoplankton in various regions of the
748 Southern Ocean. *Aquat Microb Ecol* 18(3):275-284.

749 57. Price PB & Sowers T (2004) Temperature dependence of metabolic rates for
750 microbial growth, maintenance, and survival. *Proc Natl Acad Sci U S A*
751 101(13):4631-4636.

752 58. Follows MJ, Dutkiewicz S, Grant S, & Chisholm SW (2007) Emergent
753 biogeography of microbial communities in a model ocean. *Science*
754 315(5820):1843-1846.

755 59. Zinser ER, *et al.* (2007) Influence of light and temperature on *Prochlorococcus*
756 ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography*
757 52(5):2205-2220.

758 60. Chase JM (2003) Community assembly: when should history matter? *Oecologia*
759 136(4):489-498.

760 61. Gonze D, Lahti L, Raes J, & Faust K (2017) Multi-stability and the origin of
761 microbial community types. *ISME J* 11(10):2159-2166.

762 62. Martiny JB, *et al.* (2006) Microbial biogeography: putting microorganisms on
763 the map. *Nat Rev Microbiol* 4(2):102-112.

63. Logares R (2006) Does the global microbiota consist of a few cosmopolitan species? *Ecología Austral* 16:85-90.

64. Foissner W (2006) Biogeography and Dispersal of Micro-organisms: A Review Emphasizing Protists. *Acta Protozoologica* 45:111-136.

65. Papke RT & Ward DM (2004) The importance of physical isolation to microbial diversification. *FEMS Microbiol. Ecol.* 48(3):293-303.

66. Kirchman DL (2008) *Microbial Ecology of the Oceans* (John Wiley & Sons, Hoboken, New Jersey).

67. de Vargas C*, et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348(6237):1261605.

68. Darling KF, Kucera M, & Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proc Natl Acad Sci U S A* 104(12):5002-5007.

69. Casteleyn G*, et al.* (2010) Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc Natl Acad Sci U S A* 107(29):12952-12957.

70. Cermeño P, de Vargas C, Abrantes F, & Falkowski PG (Phytoplankton biogeography and community stability in the ocean. *PLoS One* 5(4):e10037.

71. Cermeno P & Falkowski PG (2009) Controls on diatom biogeography in the ocean. *Science* 325(5947):1539-1541.

72. Whittaker KA & Rynearson TA (2017) Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proc Natl Acad Sci U S A* 114(10):2651-2656.

73. Bass D, Richards TA, Matthai L, Marsh V, & Cavalier-Smith T (2007) DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evol Biol* 7(1):162.

74. Lewis J, Harris ASD, Jones KJ, & Edmonds RL (1999) Long-term survival of marine planktonic diatoms and dinoflagellates in stored sediment samples. *J Plankton Res* 21(2):343-354.

75. Billard C & Inouye I (2004) What is new in coccolithophore biology? *Coccolithophores: From Molecular Processes to Global Impact*, eds Thierstein HR & Young JR (Springer Berlin Heidelberg, Berlin, Heidelberg), pp 1-29.

76. Sintes E, De Corte D, Ouillon N, & Herndl GJ (2015) Macroecological patterns of archaeal ammonia oxidizers in the Atlantic Ocean. *Mol Ecol* 24(19):4931-4942.

77. Locey KJ (2010) Synthesizing traditional biogeography with microbial ecology: the importance of dormancy. *Journal of Biogeography* 37(10):1835-1841.

78. Louca S, *et al.* (2018) Function and functional redundancy in microbial systems. *Nature Ecology & Evolution* 2:936-943.

79. Östman Ö, *et al.* (2010) Regional invariance among microbial communities. *Ecology letters* 13(1):118-127.

80. Salazar G, *et al.* (2016) Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J* 10(3):596-608.

81. Zinger L, Boetius A, & Ramette A (2014) Bacterial taxa–area and distance–decay relationships in marine environments. *Mol Ecol* 23(4):954-964.

82. Díez B, Massana R, Estrada M, & Pedrós-Alió C (2004) Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean, studied by denaturing gradient gel electrophoresis. *Limnology and Oceanography* 49(4):1022-1034.

83. Flaviani F, *et al.* (2018) Distinct oceanic microbiomes (from viruses to protists) found either side of the Antarctic Polar Front. *Front Microbiol* 9.

84. Bar-On YM, Phillips R, & Milo R (2018) The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*.

85. Whitman WB, Coleman DC, & Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95(12):6578-6583.

86. Estrada M, *et al.* (2016) Phytoplankton across Tropical and Subtropical Regions of the Atlantic, Indian and Pacific Oceans. *PLoS One* 11(3):e0151699.

87. Grasshoff K, Ehrhardt M, & Kremling K (1983) *Methods on seawater analysis*.

88. Boyer TP, *et al.* (2013) World Ocean Database 2013. in *NOAA Atlas NESDIS 72*, eds Levitus S & Mishonov A (NOAA, Silver Spring, MD).

89. Catalá TS, *et al.* (2016) Drivers of fluorescent dissolved organic matter in the global epipelagic ocean. *Limnology and Oceanography* 61(3):1101-1119.

90. Catalá TS, *et al.* (2016) Chromophoric signatures of microbial by-products in the dark ocean. *Geophysical Research Letters* 43(14):7639-7648.

91. Fernandez-Castro B, *et al.* (2015) Importance of salt fingering for new nitrogen supply in the oligotrophic ocean. *Nature communications* 6:8002.

92. Massana R, Murray AE, Preston CM, & DeLong EF (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* 63(1):50-56.

93. Stoeck T, *et al.* (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 19 Suppl 1:21-31.

835 94. Parada AE, Needham DM, & Fuhrman JA (2016) Every base matters: assessing
836 small subunit rRNA primers for marine microbiomes with mock communities,
837 time series and global field samples. *Environ Microbiol* 18(5):1403-1414.

838 95. Logares R (2017) Workflow for Analysing MiSeq Amplicons based on Uparse
839 v1.5. (https://doi.org/10.5281/zenodo.259579).

840 96. Nikolenko SI, Korobeynikov AI, & Alekseyev MA (2013) BayesHammer:
841 Bayesian clustering for error correction in single-cell sequencing. *BMC*
842 *Genomics* 14 Suppl 1:S7.

843 97. Schirmer M, *et al.* (2015) Insight into biases and sequencing errors for amplicon
844 sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 43(6):e37.

845 98. Zhang J, Kobert K, Flouri T, & Stamatakis A (2014) PEAR: a fast and accurate
846 Illumina Paired-End reAd mergeR. *Bioinformatics* 30(5):614-620.

847 99. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST.
848 *Bioinformatics* 26(19):2460-2461.

849 100. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial
850 amplicon reads. *Nat Methods* 10(10):996-998.

851 101. Quast C, *et al.* (2013) The SILVA ribosomal RNA gene database project:
852 improved data processing and web-based tools. *Nucleic Acids Res* 41(Database
853 issue):D590-596.

854 102. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local
855 alignment search tool. *Journal of molecular biology* 215(3):403-410.

856 103. Oksanen J, *et al.* (2008) vegan: Community Ecology Package. R package
857 version 1.15-0.

858 104. Logares R, *et al.* (2014) Patterns of rare and abundant marine microbial
859 eukaryotes. *Curr Biol* 24(8):813-821.

860 105. Schloss PD, *et al.* (2009) Introducing mothur: open-source, platform-
861 independent, community-supported software for describing and comparing
862 microbial communities. *Appl Environ Microbiol* 75(23):7537-7541.

863 106. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for
864 automated alignment trimming in large-scale phylogenetic analyses.
865 *Bioinformatics* 25(15):1972-1973.

866 107. Price MN, Dehal PS, & Arkin AP (2009) FastTree: computing large minimum
867 evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*
868 26(7):1641-1650.

869 108. R-Development-Core-Team (2008) *R: A language and environment for*
870 *statistical computing.* (R Foundation for Statistical Computing, Vienna,
871 Austria).

872 109. Dray S*, et al.* (2017) adespatial: Multivariate Multiscale Spatial Analysis).

873 110. Paradis E, Claude J, & Strimmer K (2004) APE: Analyses of Phylogenetics and
874 Evolution in R language. *Bioinformatics* 20(2):289-290.

875 111. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer-
876 Verlag).

877 112. Chen J*, et al.* (2012) Associating microbiome composition with environmental
878 covariates using generalized UniFrac distances. *Bioinformatics* 28(16):2106-
879 2113.

880 113. Becker RA, Wilks AR, Brownrigg R, Minka TP, & Deckmyn A (2017) maps:
881 Draw Geographical Maps).

882 114. Gerritsen H (2014) mapplots: Data Visualisation on Maps).

883 115. Kembel SW*, et al.* (2010) Picante: R tools for integrating phylogenies and
884 ecology. *Bioinformatics* 26(11):1463-1464.

885 116. Chase JM, Kraft NJB, Smith KG, Vellend M, & Inouye BD (2011) Using null
886 models to disentangle variation in community dissimilarity from variation in α-
887 diversity. *Ecosphere* 2(2):1-11.

888 117. Friedman J & Alm EJ (2012) Inferring correlation networks from genomic
889 survey data. *PLoS Comput Biol* 8(9):e1002687.

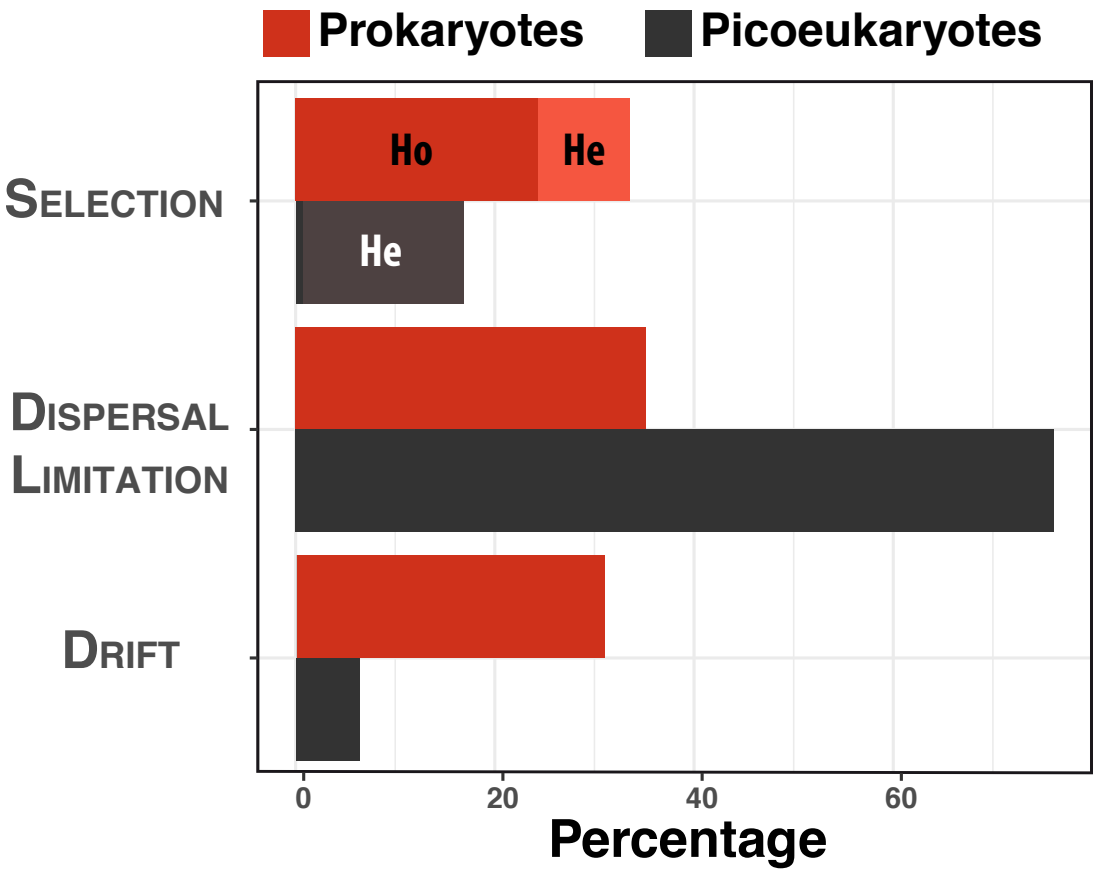
890 118. Watts SC, Ritchie SC, Inouye M, & Holt KE (2018) FastSpar: Rapid and
891 scalable correlation estimation for compositional data. *bioRxiv*.

892 119. Shannon P*, et al.* (2003) Cytoscape: a software environment for integrated
893 models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.

894 120. Csardi G & Nepusz T (2006) The igraph software package for complex network
895 research. *InterJournal* Complex Systems:1695.
896
897

898 **FIGURES**

899



900

901 **Figure 1. Relative importance of the processes shaping the tropical and subtropical**

902 **sunlit-ocean microbiome.** Percentage of the spatial turnover governed by different

903 processes in prokaryotes and picoeukaryotes. Ho = Homogeneous selection, He =

904 Heterogeneous selection (light red and light black).

905

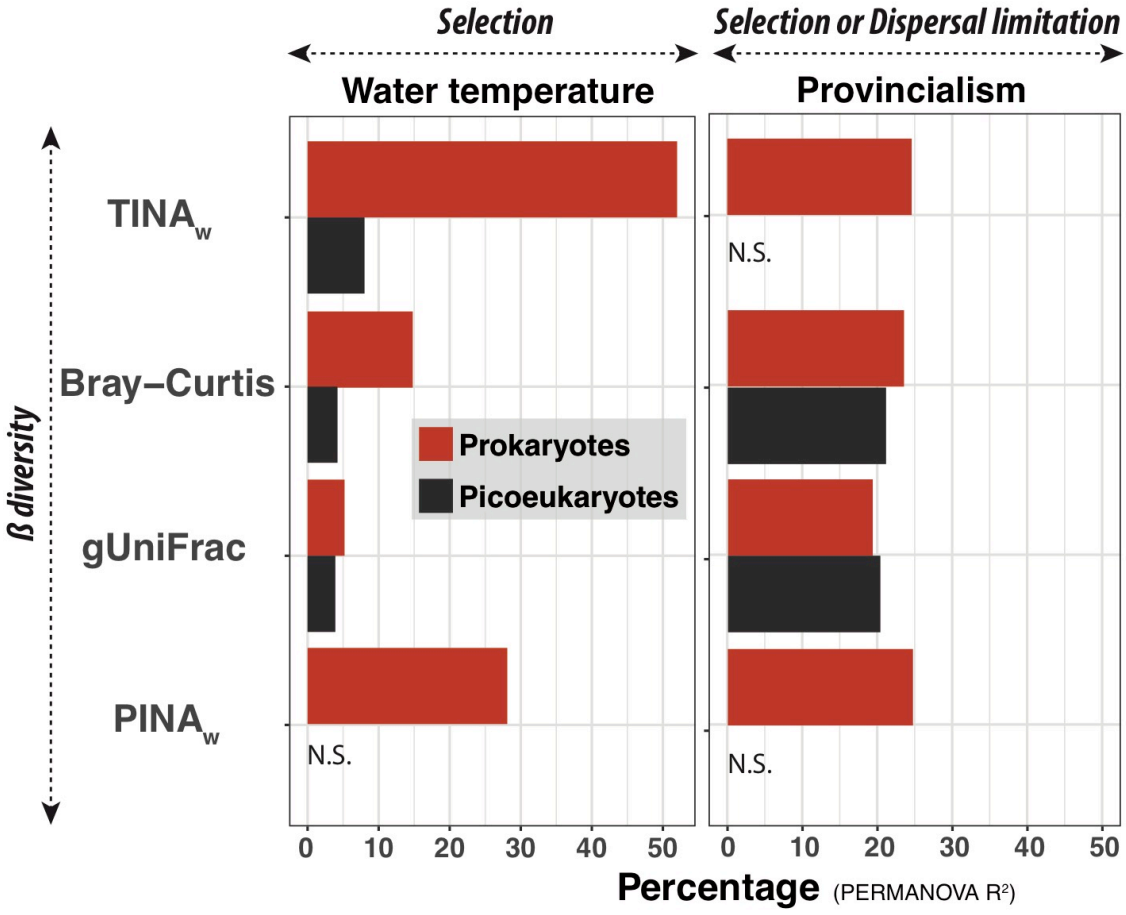


Figure 2. Water Temperature and Provincialism were the most important variables explaining the structure of the tropical and subtropical surface-ocean microbiome. Percentage of variance in picoeukaryotic and prokaryotic community composition (PERMANOVA R^2) explained by Water Temperature and Longhurst Provinces when using different β-diversity metrics. Figure based on the *Meta-119* dataset (see Methods). $TINA_w$: TINA weighted, gUniFrac: Generalized Unifrac, PINAw: PINA weighted. N.S. = Non-Significant. Note that $TINA_w$ captures a significantly higher proportion of community variance than Bray-Curtis in prokaryotes.

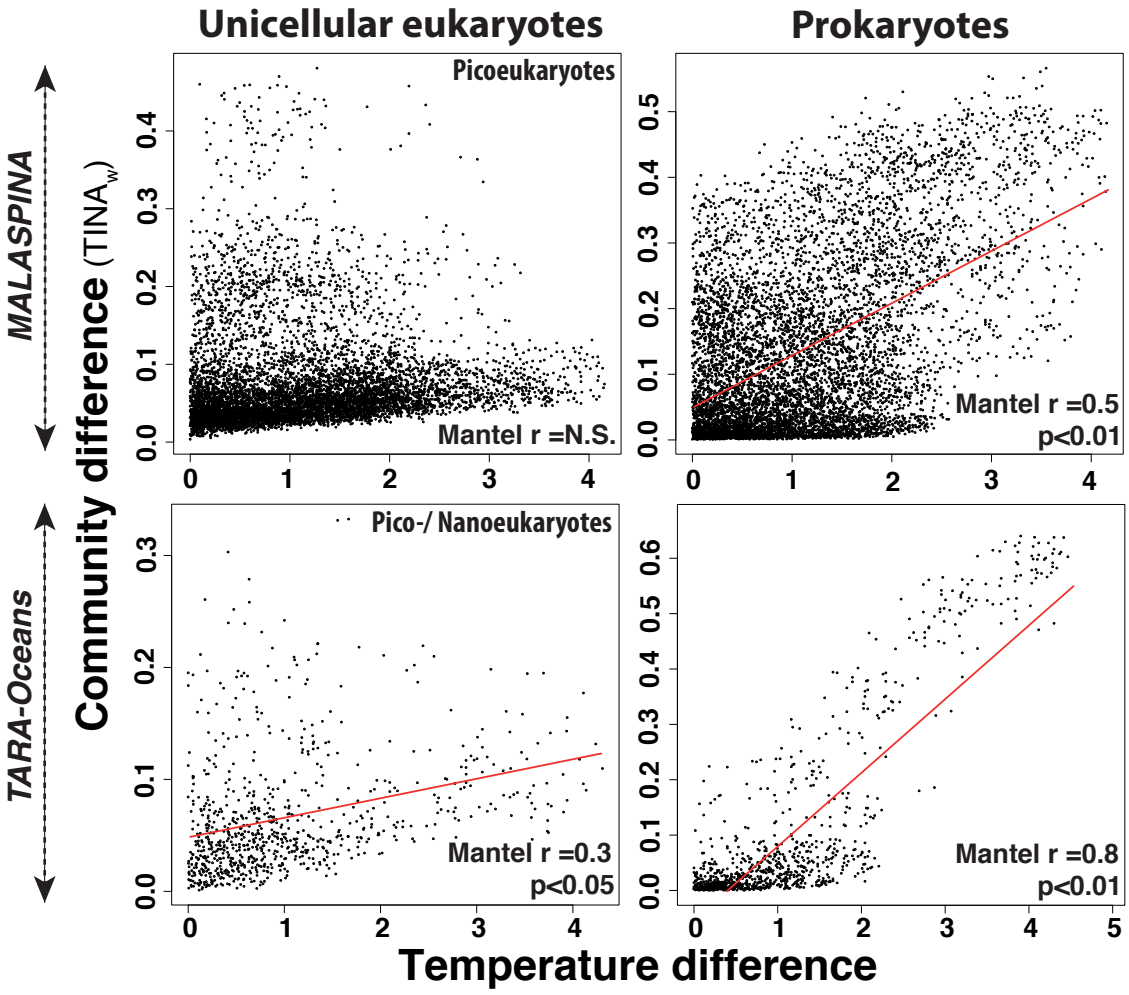


Figure 3. Temperature-driven selection affects species co-occurrences in prokaryotes but not in small unicellular eukaryotes. Community differences (Tina-weighted dissimilarities) vs. temperature differences (Euclidean distances) for both small unicellular eukaryotes and prokaryotes sampled during the Malaspina and TARA Oceans expeditions. NB: While only picoeukaryotes were contemplated in Malaspina (cell sizes <3 µm), TARA Oceans data included pico- and nanoeukaryotes (cell sizes <5 µm). Pico- and nanoeukaryotes from both expeditions (left panels) displayed low or no correlations between $TINA_w$ distances and temperature differences (Note Mantel test results in the figures). On the contrary, prokaryotes (right panels) had high to moderate

929 correlations between $TINA_w$ distances and temperature differences. The regression line
930 is shown in red (Malaspina microbial eukaryotes N.S., Malaspina Prokaryotes R^2=0.3,
931 TARA microbial eukaryotes R^2=0.1, TARA Prokaryotes R^2=0.7; $p<0.05$).
932

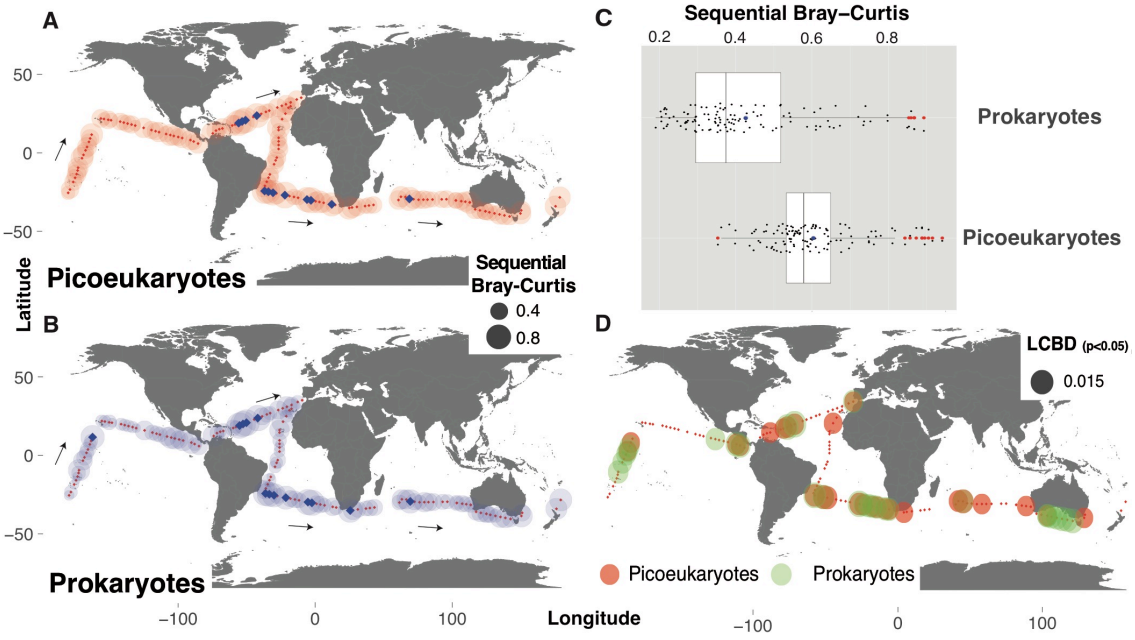


Figure 4. Spatial patterns likely manifesting differential dispersal or strong selective change. Panels A-C: Sequential change in community composition across space (sequential β-diversity). Communities were sampled along the R/V Hespérides trajectory (Panels A and B, black arrows), and the composition of each community was compared against its immediate predecessor. In Panels A and B, the size of each bubble represents the Bray-Curtis dissimilarity between a given community and the community sampled previously. Blue squares in Panels A and B represent the stations where pairwise β-diversity displayed abrupt changes (Bray Curtis values >0.8 for picoeukaryotes and >0.7 for prokaryotes). Abrupt changes coincided in a total of 11 out of 14 stations for both picoeukaryotes and prokaryotes, while one station presented marked changes for picoeukaryotes and two for prokaryotes. Panel C summarizes the sequential Bray-Curtis values for prokaryotes and picoeukaryotes [Means were

950 significantly different between domains (Wilcoxon text, $p<0.05$)]. Panel D shows the 36
951 stations featuring a comparatively large contribution to the overall β-diversity (Local
952 Contribution to Beta Diversity (40); $p<0.05$).
953
954

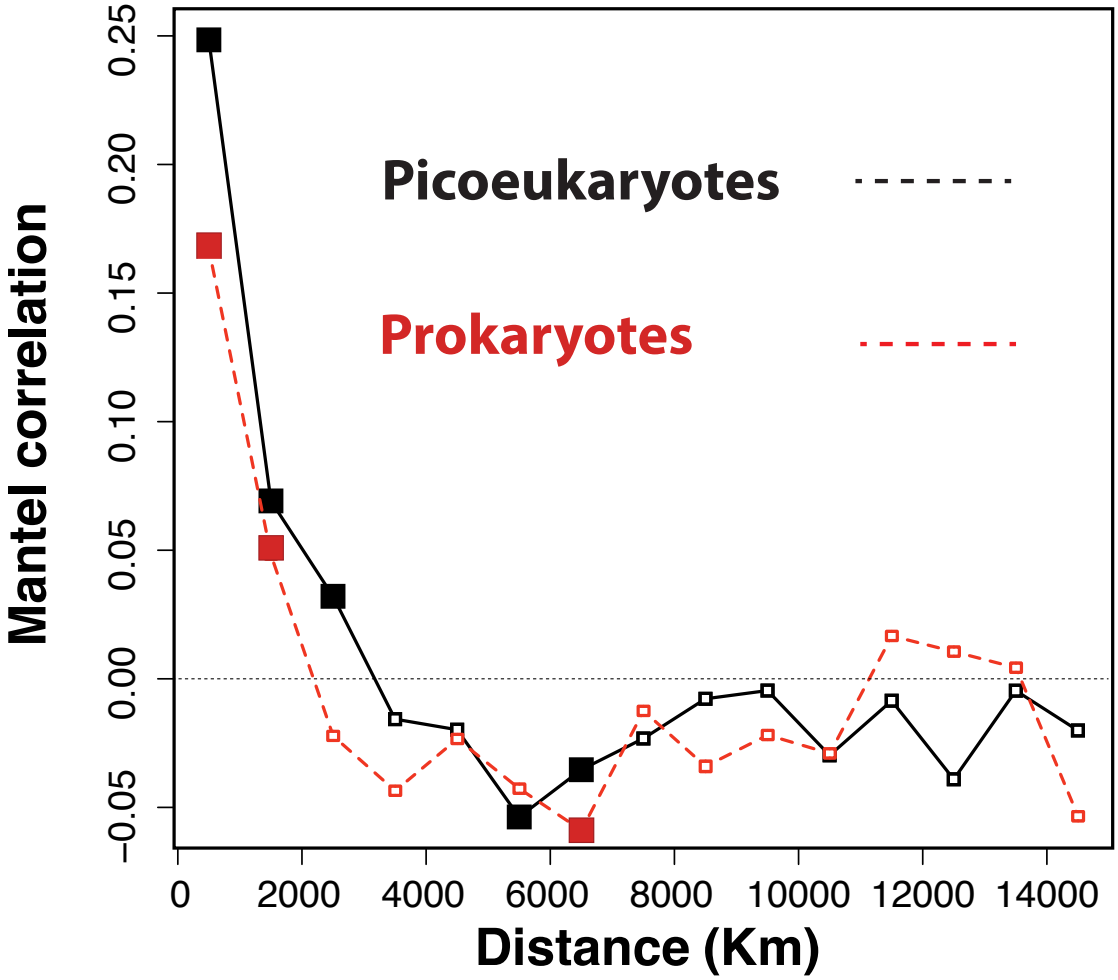


Figure 5. Decrease in community similarity with distance (distance decay). Mantel correlograms between geographic distance and β-diversity featuring distance classes of 1,000 km for both picoeukaryotes and prokaryotes. Coloured squares indicate statistically significant correlations ($p<0.05$). Note that β-diversity in picoeukaryotes displayed positive correlations with increasing distances up to ~3,000 km, while prokaryotes had positive correlations with distances up to ~2,000 km. Correlations tended to be smaller in prokaryotes than in picoeukaryotes, indicating smaller distance decay in the former compared to the latter.