UNSW SYDNEY

Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects

Author:
Wilkins, David; van Sebille, Erik; Rintoul, S; Lauro, Federico M; Cavicchioli, Ricardo

Publication details:
Nature Communications
v. 4
pp. 2457
2041-1723 (ISSN)

Publication Date:
2013

Publisher DOI:
http://dx.doi.org/10.1038/ncomms3457

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/53708 in https://unsworks.unsw.edu.au on 2024-04-19

Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects

David Wilkins[1], Erik van Sebille[2], Stephen R Rintoul[3], Federico M Lauro[1], Ricardo Cavicchioli[1*]

[1] School of Biotechnology and Biomolecular Sciences, University of New South Wales, New South Wales, 2052, Australia.

[2] Climate Change Research Centre & ARC Centre of Excellence for Climate System Science, University of New South Wales, New South Wales, 2052, Australia.

[3] CSIRO Marine and Atmospheric Research, and Centre for Australian Weather and Climate Research – A partnership of the Bureau of Meteorology and CSIRO, and CSIRO Wealth from Oceans National Research Flagship, and the Antarctic Climate and Ecosystems Cooperative Research Centre, University of Tasmania, Castray Esplanade, Hobart, Tasmania, 7001, Australia.

* To whom correspondence should be addressed: Ricardo Cavicchioli, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW, 2052, Tel. (+61 2) 9385 3516, Fax. (+61 2) 9385 2742, Email. r.cavicchioli@unsw.edu.au

Abstract

While environmental selection and spatial separation have been shown to shape the distribution and abundance of marine microorganisms, the effects of advection (physical transport) have not been directly tested. Here we examine 25 samples covering all major water masses of the Southern Ocean to determine the effects of advection on microbial biogeography. Even when environmental factors and spatial separation are controlled for, there is a positive correlation between advection distance and taxonomic dissimilarity, indicating that an "advection effect" plays a role in shaping marine microbial community composition. This effect is likely due to the advection of cells increasing the probability that upstream microorganisms will colonise downstream sites. Our study shows that in addition to distance and environmental selection, advection shapes the composition of marine microbial communities.

Introduction

The central goal of microbial biogeography is to understand how the distribution and abundance of microorganisms are shaped by their physical context. The Baas Becking hypothesis — that "*everything is everywhere*, but, *the environment selects*[1,2]" — posits that the rapid dispersal of microorganisms means microbial community structure is determined entirely by environmental selection. This stands in contrast to macroorganism biogeography, which has long been recognised as being under the control of historical (in addition to contemporary environmental) factors, particularly spatial influences such as barriers to dispersal. Microbial biogeography studies have begun to show that historical factors may also shape the distribution of microorganisms[3], e.g. a correlation between spatial and genetic distance (a "distance effect") in fluorescent *Pseudomonas* strains in soils[4]. This study, among others[5,6], also demonstrated the importance of taxonomic resolution in describing such biogeographic patterns. Other studies have found that dispersal potential varies between microbial species, leading to different or absent distance effects[7]. When combined with contemporary environmental selection ("environment effect"), distance effects explain some but not all variation between microbial communities, and the mechanism(s) by which a distance effect arises are not always clear[8].

In the ocean, several recent studies have found that microbial communities can be endemic to hydrographically distinct water masses. Surveys in the Arctic[9] and North Atlantic[10] oceans have found that bacterial assemblages within the same water mass can be similar across a range of thousands of kilometres, but assemblages can differ between water masses across a range of hundreds of meters. Water masses are defined by their distinct physicochemical properties, so such patterns do not directly imply the existence of factors beyond environmental selection. However, in some cases a water mass–community relationship has been shown to persist even when environment effects are statistically controlled for[11,12].

One explanation for these results is that microbial assemblages are influenced by the advection (physical transport) of cells by ocean currents. Higher dispersal rates cause the microbial community composition at a given site to increasingly resemble the dispersed colonisers, and less reflect local environmental selection and stochastic effects such as genetic drift[8]. Hence, it would be expected that

locations that are closely connected by advection would have more similar compositions than those that are not, even when the environment effect is accounted for. Indeed, advection is often invoked to explain observations of microbial diversity or abundance which do not seem attributable to environmental selection (e.g. refs. 13-16). The exchange of very small volumes of water between marine microbial mesocosms has been found to greatly reduce their β-diversity (in this case, compositional differences between communities from different mesocosms) even under consistent environmental conditions[17]. This suggests that advection of even small numbers of cells could have a large homogenizing effect independent of environmental selection. However, the existence of a relationship between advection and community composition that is independent of environment and distance effects has not been directly tested.

The Southern Ocean (SO) is composed of several water masses, which are physicochemically distinct but linked by circulation[18] (Fig. 1). Oxygen-depleted Upper Circumpolar Deep Water (UCDW), originating in the deep Indian and Pacific oceans[19], and high salinity Lower Circumpolar Deep Water (LCDW) originating in the Atlantic, spread poleward and shoal to reach the sea surface Antarctic Zone (AZ) between the Polar Front and the Antarctic continent. Some of the upwelled Circumpolar Deep Water (CDW) moves southwards, where it becomes colder and denser and sinks to form Antarctic Bottom Water (AABW), which spreads northwards to ventilate the densest layers of the global ocean. The remainder of the upwelled water is warmed and freshened as it moves northward before sinking to form Antarctic Intermediate Water (AAIW), which spreads northwards at about 1000 m depth. Overlying the AAIW is the warmer Subantarctic Mode Water (SAMW).

This study aims to determine whether advection shapes the community structure of bacteria and archaea, independent of environment and distance effects. By sampling each of the SO water masses (depths from surface to ≈6 km), we compare microbial communities over a large spatial distance (≈3000 km) and range of environments, and test whether advection plays a role in shaping their composition. When environment and distance effects are controlled for, a significant correlation between advection distance and community dissimilarity is observed. The findings indicate that advection shapes marine microbial community composition.

Results

Sequencing and taxonomic assignment After trimming, denoising and chimera removal, the 25 samples (each with three separately sequenced size fractions) yielded 1,008,963 16S rRNA gene pyrosequencing reads of length 251–561 bp (mean 426 bp). Individual fractions yielded 3,687–52,192 reads (mean 13,453). After pre-processing in QIIME, 2,473–45,046 (mean 11,599) reads per fraction were retained.

Operational Taxonomic Units (OTUs) were identified by clustering reads at a minimum sequence identity of 97%, broadly equivalent to the species level. This threshold has been widely used for the delineation of bacterial species, allowing our results to be compared to past and future studies that adopt this criterion. 15,868 unique OTUs were identified across all fractions of all samples, 13,011 of which were generated *de novo* i.e. did not cluster with sequences in the SILVA database. In many samples and size fractions, the total abundance of OTUs generated *de novo* was greater than that of OTUs identified by clustering with SILVA seed sequences (Fig. 2). However, the mean relative abundance of individual OTUs that clustered with SILVA was higher (2.1%) than that of OTUs generated *de novo* (0.43%). This suggests that while there exists a "long tail" of many rare OTUs that are not in SILVA, the most abundant OTUs in the sampled environments are well represented in SILVA. The Chao 1 statistic was calculated, and estimated OTUs were under-sequenced by 1.5–51% across all samples (mean 37%) (Supplementary Data1).

Non-metric Multidimensional Scaling (nMDS) ordination showed that the sampled water masses could be distinguished on the basis of taxonomic distance (Fig. 3c). This was supported by Analysis of Similarity (ANOSIM) analysis ($R = 0.83$, $p = 0.001$). While each water mass had a distinct taxonomic profile, some broad differences between surface and deep masses were observed (Fig. 2). Surface waters (AZ, PFZ and SAMW) had high abundances of OTUs clustering with SILVA sequences from the Alphaproteobacteria, Bacteroidetes and Gammaproteobacteria. The higher abundance of Bacteroidetes at the surface reflects their association with phytoplankton, as many species in this lineage specialise in the degradation of high molecular weight products of primary production[20]. Alphaproteobacteria were represented primarily by the SAR11 clade, abundant in ocean surface communities[21] including the SO[22], and Roseobacter clades, which have also been associated

with degradation of phytoplankton products[15,20] (Supplementary Data 2). The dominant Gammaproteobacterial orders were the Alteromonadales and Oceanospirillales, typical of SO surface waters[23]. Archaeal OTUs were not abundant at the surface, consistent with their well-described decline in abundance during summer[24,25]. The deep water masses (CDW, AAIW and AABW) were dominated by Crenarchaeota, Euryarchaeota and Gammaproteobacteria, again consistent with previous findings[26].

Environment and distance effects nMDS ordination showed that the sampled water masses clustered well on the basis of environmental distance (Fig. 3b). This was supported by ANOSIM ($R = 0.84$, $p = 0.001$). A partial Mantel test, comparing the taxonomic to environmental matrices with the spatial matrix held constant, found a correlation of $r = 0.54$ ($p = 0.001$), indicating a strong environment effect.

Distance-based Linear Models (distLM) analysis of the individual physicochemical variables found that considered separately, each of phosphate, silicate, nitrate, oxygen, salinity and pressure explained 12–25% of the taxonomic variance between samples ($p = 0.001–0.002$). Temperature had no significant effect on taxonomic composition when considered separately ($p > 0.05$). When all combinations of variables were considered (BEST modelling), the full set of variables was found to be a better solution (adjusted $R^2 = 0.33$) than any subset. This suggests that no redundant physicochemical parameters were measured.

The distance-based Redundancy Analysis (dbRDA) plot showed the physicochemical parameters structured the samples first along an axis separating surface and deep samples (dbRDA1), strongly related to dissolved oxygen ($r = 0.72$) (Fig. 4). The second axis was best correlated with temperature ($r = 0.91$). All variables had a moderate correlation with at least one of the first two axes (Supplementary Data 3). As with the nMDS ordinations (Fig. 3), the water masses were generally well separated by the first two dbRDA axes. Samples from the AABW and AAIW, which were not well separated by the first two axes, were clearly separated along the third (Supplementary Data 3),

which was best correlated with silicate ($r = -0.77$), reflecting the relative enrichment of silicate in the AABW.

A distance effect was detected by comparing the taxonomic and spatial matrices with the environmental matrix held constant (partial Mantel; $r = 0.41$, $p = 0.003$). This indicated that a process other than contemporary environmental selection was appreciably affecting variation in microbial community composition.

Testing of advection effect 244,000 encounters were recorded between particles and sample sites during the 100 y advection simulation. Encounter times spanned the full range of the simulation (5 d–100 y), with a median of 30 d and mean of 3018 d. 47 pairs of samples did not yield mutual encounters (i.e. at least one particle from one sample encountering the other). Of these, every pair included at least one AABW sample.

nMDS ordination showed that the sampled water masses could be broadly distinguished on the basis of their mutual advection distances (Fig. 3a). This was supported by ANOSIM ($R = 0.41$, $p = 0.002$).

A partial Mantel test, comparing the taxonomic and advection matrices with the spatial and environmental matrices held constant, showed that advection has a moderate ($r = 0.28$) and significant ($p = 0.01$) correlation with taxonomic composition independent of spatial and environmental factors. To ensure this result was not unduly influenced by the samples on which the 100 y ceiling was imposed (AABW) and those for which particle releases were not simulated (samples 11, 13, 17 and 22), the test was repeated with these samples removed. The correlation was stronger and remained significant ($r = 0.34$, $p = 0.03$). To ensure the result was robust to the choice of advection distance metric, correlations were recalculated for both the full set of samples and the subset (excluding AABW and samples 11, 13, 17 and 22) with pairwise advection distance redefined as the mean time for all particle encounters between samples. The observed correlation was higher using this metric for both the full set ($r = 0.41$, $p = 0.004$) and subset ($r = 0.50$, $p = 0.005$). SourceTracker analysis confirmed that the effect was moderately directional, with the proportion of OTUs contributed from a

169 given "source" sample to a "sink" sample correlated with the proportion of particle encounters it
170 generated (Spearman's ρ = 0.15, p = 0.001).

171 To explore the role of taxonomic resolution in the detection of an advection effect, partial Mantel
172 tests were repeated as above with OTUs clustered at different sequence similarity thresholds (0.80,
173 0.85, 0.90, 0.95, 0.99, 1.00). The advection effect (as measured by the magnitude of the partial Mantel
174 correlation) was strongest at a sequence similarity threshold of 0.90 (r = 0.31, p = 0.01), decreasing in
175 both directions from this maximum (Fig. 5).

176

177 **Explaining the advection effect** Two explanations can be proposed for the relationship between
178 advection and community composition. The first is that advection increases microbial dispersal by
179 increasing the probability that OTUs from one site will encounter and colonise another (the "dispersal
180 mechanism"). This assumes microbial dispersal is less than perfect; i.e. that "everything is *not*
181 everywhere". The second possibility is that advection transports a large numbers of cells at a rate
182 measurably outpacing environmental selection, regardless of whether or not those cells are able to
183 successfully colonise (i.e. grow and reproduce in) downstream sites: that "the environment selects, but
184 not fast enough" (the "bulk transport mechanism"). Considering the long advection times between
185 sites in this study, the relatively rapid growth of marine microorganisms, and the ability of small
186 numbers of cells to rapidly reduce β-diversity between isolated marine sites[17], the dispersal
187 mechanism seems the more plausible.

188 Two tests were performed to distinguish between these hypotheses. Firstly, the advection
189 distance matrix was reconstructed using the absolute pairwise number of particle encounters as the
190 distance metric between samples (distance metric for given sample pair = maximum number of
191 encounters across all sample pairs – encounters for given pair + 1). There was no significant
192 correlation between this matrix and the taxonomic distance matrix when the environmental and spatial
193 matrices were held constant (p > 0.05). This suggests that advection time, not the absolute number of
194 transported cells, is the relevant factor, supporting the dispersal mechanism.

Secondly, it was reasoned that under the dispersal mechanism, advection would have a large effect on the presence and diversity of taxa and a smaller effect on their abundances; while if the bulk transport hypothesis held, the effect would largely be on the relative abundances. To test this, the taxonomic profiles were transformed to a presence/absence measure (OTU present = 1, absent = 0), and a matrix of Sørensen dissimilarities generated. The advection effect was slightly stronger ($r = 0.29$, $p = 0.009$) than when abundances were considered, again supporting the dispersal hypothesis.

Discussion

The hypothesis that advection shapes microbial assemblages independent of an environment or distance effect is supported by our analyses of SO communities (Figures 1-5). Communities that are more closely connected by advection are more similar, and this effect exists even when environment and distance effects are controlled for. Our analyses also indicate that advection primarily shapes microbial community structure by increasing opportunities for colonisation, rather than transporting large numbers of cells to downstream sites.

The amount of variance in taxonomic composition explained by the advection effect can be estimated as the square of the correlation coefficient (r^2). This value, 8%, is likely to be a conservative estimate, as our study only captured advective pathways from one sample to another; it is possible that some samples with a large mutual advective distance shared a common advective source outside the study area. A recent review of studies partitioning variance in microbial community composition found that the mean reported variance explained by a distance effect was 10%, and by an environment effect 27%[8]. The estimated proportions of variance in our study, 17% ($p = 0.003$) for distance and 29% ($p = 0.001$) for environment, were close to these values, with the larger than average distance effect likely due to the large spatial scale of this study. While there are no comparative data for the advection effect in other systems, the value of at least 8% for the SO indicates that advection is important relative to both distance and environment effects.

Taxonomic resolution (i.e. the level of genetic difference at which OTUs are discriminated) often determines whether or not a pattern is detected in studies of microbial distance effects, with finer

222 resolutions generally leading to an increased likelihood of detecting a significant effect[8]. While this
223 held true in this study for OTU clustering thresholds of 0.80–0.90, OTU profiles generated with
224 thresholds finer than 0.90 had decreasing correlations to advection, although they remained
225 statistically significant (Fig. 5). There are several factors that may contribute to this pattern. At very
226 fine (0.99 and 1.00) clustering thresholds, sequencing error may increase random noise within the
227 OTU profiles. As the V6–V8 hypervariable regions were selected as the sequencing target, it may also
228 be due to intraspecific variation in 16S rRNA gene sequence within established populations[27], which
229 could increase the apparent divergence between sampled communities. While such variation is
230 unlikely to accumulate during the relatively short advection times observed in this study (median
231 encounter time 30 d), it is consistent with advection as a historical factor increasing the chance of
232 colonisers from "upstream" sites reaching and establishing long-term populations in "downstream"
233 sites. However, as it is unlikely that intraspecific variation would have an effect at thresholds < 0.97,
234 this pattern of decreasing effect at taxonomic resolutions finer than 0.90 should be further
235 investigated.

236 To obtain a better understanding of the advection effect, it would be useful to determine whether
237 specific taxonomic or physiological groups are more amenable to dispersal by advection; e.g. through
238 the formation of dormant spores[7,13], copiotroph resting states (e.g. *Photobacterium angustum*[28]) or
239 through the inherent stress resistance states of some oligotrophs (e.g. *Sphingopyxis alaskensis*[28]). As it
240 is possible that an unmeasured environmental variable (e.g. iron concentration; microenvironmental
241 factors due to particle attachment) also correlates with both advection and community composition,
242 future studies should address this. Mesocosm experiments[17] may also help to address the effects of
243 small-scale exchanges of cells on community structure.

244 While future work will help to clarify the nature of the advection effect, our data indicate that
245 advection should be considered alongside distance and environment effects in studies addressing
246 factors that shape marine microbial assemblages.

247

248 **Methods**

Sampling Sampling was conducted on board the RSV *Aurora Australis* during cruise V3 from January 20th–February 7th 2012. This cruise occupied a latitudinal transect from waters north of Cape Poinsett, Antarctica (65° S) to south of Cape Leeuwin, Australia (37° S) within a longitudinal range of 113–115° E. Sampling was performed as described in ref. 29, with sites and depths selected to provide coverage of all major SO water masses. At each surface station, ≈250–560 L of seawater was pumped from ≈1.5–2.5 m depth. At some surface stations, an additional sample was taken from the Deep Chlorophyll Maximum (DCM), as determined by chlorophyll fluorescence measurements taken from a Conductivity, Temperature and Depth probe (CTD) cast at each sampling station. Samples of mesopelagic and deeper waters (≈120–240 L) were also collected at some stations using Niskin bottles attached to the CTD. Sampling depths were selected based on temperature, salinity and dissolved oxygen profiles to capture water from the targeted water masses. Profiles were generated on the CTD descent, and samples collected on the ascent at the selected depths. Deep water masses were identified by the following criteria: CDW = oxygen minimum (Upper Circumpolar Deep) or salinity maximum (Lower Circumpolar Deep); AABW = deep potential temperature minimum; AAIW = salinity minimum[18]. The major fronts of the SO, which coincide with strong horizontal gradients in temperature and salinity[19,30], separate regions with similar surface water properties. The Antarctic Zone (AZ) lies south of the Polar Front (which was at 51° S during sampling), while the Polar Frontal Zone (PFZ) lies between the Polar Front and the Subantarctic Front. In total, 25 samples from the AZ, PFZ, SAMW, AAIW, CDW and AABW were collected for this study (Fig. 1, Supplementary Data 1). Seawater samples were prefiltered through a 20 µm plankton net, biomass captured on sequential 3.0 µm, 0.8 µm and 0.1 µm 293 mm polyethersulfone membrane filters, and filters immediately stored at −80 °C[31,32].

DNA extraction and sequencing DNA was extracted with a modified version of the phenol-chloroform method[31]. Tag pyrosequencing was performed by Research and Testing Laboratory (Lubbock, USA) on a GS FLX+ platform (Roche, Branford, USA), using a modification of the standard 926F/1392R primers targeting the V6–V8 hypervariable regions of bacterial and archaeal 16S rRNA genes (926wF: AAA-CTY-AAA-KGA-ATT-GRC-GG, 1392R: ACG-GGC-GGT-GTG-

277 TRC). Denoising, chimera removal and trimming of poor quality read ends were performed by the
278 sequencing facility.
279
280 **Generation of taxonomic dissimilarity matrix** Using Quantitative Insights Into Microbial Ecology
281 (QIIME) 1.6.0[33], tag pyrosequencing reads were clustered into OTUs with > 97% sequence similarity
282 using the uclust_ref algorithm[34], with rRNA gene sequences from the SILVA database (release 108,
283 eukaryote and chloroplast sequences removed)[35] used as seed sequences and *de novo* cluster
284 formation (i.e. formation of clusters with no seed sequence) allowed. Following the QIIME default
285 settings, reads with mean quality scores < 25, homopolymer runs > 6 nucleotides in length or errors in
286 the primer sequence were discarded during pre-processing. To generate a taxonomic profile for each
287 sample, the relative abundances of reads assigned to each OTU in each size fraction were encoded as
288 variables. To account for the reads discarded by QIIME, abundances in each size fraction were
289 standardised by the proportion of reads retained. The abundances were square root transformed and
290 Bray-Curtis dissimilarity indices between samples calculated in PRIMER 6 (PRIMER-E, Lutton,
291 UK). The Chao 1 statistic was calculated in QIIME to estimate the total number of OTUs in each size
292 fraction of each sample, and therefore the degree of under-sequencing. Although this statistic can be
293 sensitive to sequencing error due to its use of singleton OTU counts, this is unlikely to have a large
294 effect as OTUs were clustered at the 0.97 sequence identity threshold.
295
296 **Environmental measurements and generation of distance matrix** Environmental data were
297 collected from CTD casts at each sample site. Pressure, dissolved oxygen concentration and water
298 temperature measurements were collected with CTD instruments. Salinity and concentrations of
299 dissolved phosphate, nitrate and silicate were obtained from hydrochemical analysis of seawater
300 samples collected in Niskin bottles during CTD casts[36]. These samples were collected at discrete
301 depths, and the hydrochemical sample closest to the depth of the relevant biological sample was
302 selected. The exceptions were samples 32 and 33 (49.5° S, 115° E), for which nitrate concentrations
303 were not available, and sample 29 (53.2° S, 115° E) for which phosphate concentration was not
304 available. In these cases, a reading from the appropriate depth was substituted from the nearest

available cast (50.0°S, 115° E for samples 32 and 33; 53.8°S, 115° E for sample 29). Pressure values were $\log(x+1)$ transformed to reduce right-skew[37] and the combined instrument and hydrochemical data were used to create environmental profiles for each sample. The variables were normalised and a Euclidean distance matrix generated in PRIMER 6. Distance-based Linear Models (distLM) multivariate analysis[38] was performed to confirm the selection of physicochemical variables and explore their relationship with taxonomic composition. In PRIMER 6, all possible combinations of variables ("BEST selection") were explored by distLM, and the models (sets of variables) that best fit the taxonomic dissimilarity matrix (adjusted R^2 as the fitness measure) were selected. The relationship between the resulting model and the taxonomic dissimilarity between samples was visualised by distance-based Redundancy Analysis (dbRDA) ordination. To generate a spatial distance matrix, pairwise ellipsoidal distances between samples (including difference in depth) were calculated using INVERS3D (National Oceanic and Atmospheric Administration, Silver Springs, USA).

Generation of advection distance matrix Advection distances between the sites were computed using three-dimensional velocity data from a hydrodynamic numerical ocean model in combination with a Lagrangian trajectory toolset. The ocean model used was the Southern Ocean State Estimate (SOSE)[39], a numerical model of the SO based on the Estimating the Circulation and Climate of the Ocean (ECCO) machinery[40] and constrained by a large set of *in situ* and remote-sensed observations. SOSE has been validated in the SO[41,42]. Here, the five-day averaged three-dimensional velocity fields for the period January 2005–December 2007 were used, on a 1/6° horizontal resolution and with 42 vertical levels. The Connectivity Modeling System (CMS)[43] v1.1 was used to integrate virtual Lagrangian particles within the SOSE velocity fields. For each site, 100 particles were released every 5 d (total of 22,000 per site). The particles were released at the latitude and depth of the site, evenly spaced in a 1° zonal line centred at the site longitudes. The particles were then advected for 100 y, looping through the 3 y of available velocity fields as described in ref. 44. Three-dimensional locations of the particles were saved every 5 d. The trajectory of each particle was analysed to detect encounters between particles and sample sites. An encounter was defined as the vector between any two consecutive 5 d particle locations intersecting a box bounded by ±0.2° of latitude, ±0.5° of

333 longitude and ±50 m of depth from a sample site. Only the first encounter between any particle and
334 sample was counted. Four pairs of samples (10/11; 12/13; 16/17; 21/22), where a DCM sample was
335 taken directly below a surface sample within the mixed layer, were too close to act as separate particle
336 release sites. For these samples, simulated particle releases were performed for only one of the pair,
337 and the generated encounters were attributed to both. For all samples pairwise, the mean time in
338 seconds between a particle being released from one sample and encountering the other was calculated.
339 Pairwise advective distance between samples was defined as the mean of these two directional mean
340 times between each sample in the pair. This metric was selected to ensure advective flows which may
341 be of high biological relevance, such as a small number of particles quickly transported between sites,
342 were appropriately weighted when paired with flows of lower biological relevance, such as a large
343 number of particles transported between sites over decades. To ensure the results were robust to this
344 choice of metric, subsequent statistical tests were repeated with pairwise advective distance redefined
345 as the mean time for all pairwise encounters. The pairwise distance between the surface/DCM
346 samples discussed above was set to zero. For pairs of samples that did not yield mutual encounters (47
347 pairs, all including at least one AABW sample; see Results), the distance was set to the maximum run
348 time of the simulation (100 y). To ensure these constraints were not unduly influencing the results,
349 subsequent statistical tests were rerun without the surface/DCM samples for which particle releases
350 were not simulated and AABW samples (see "Testing of advection effect", below).
351
352 **Ordination of distances and comparison to water masses** Ordinations of the taxonomic,
353 environmental and advection distance matrices were produced by non-metric Multidimensional
354 Scaling (nMDS) using custom R scripts. Analysis of Similarities (ANOSIM) was performed in
355 PRIMER 6 to test for statistically significant differences between water masses in each of these three
356 factors. Right-tailed p-values for each ANOSIM test were computed using 999 random label
357 permutations of one of the test matrices.
358
359 **Testing of advection effect** Mantel tests were performed in Pattern Analysis, Spatial Statistics and
360 Geographic Exegesis version 2 (PASSaGE 2)[45]. To test for and quantify distance and environment

361 effects, partial Mantel tests were performed comparing the taxonomic matrix to the spatial then
362 environmental matrices, with the remaining matrix held constant. To test the hypothesis that
363 advection shapes SO microbial assemblages independent of distance and environment effects, a
364 partial Mantel test was performed comparing the taxonomic and advection matrices, with both the
365 spatial and environmental matrices held constant. Right-tailed p-values for all tests were calculated
366 using 999 random label permutations of one the test matrices. To ensure that the result was not unduly
367 influenced by the samples to which the 100 y ceiling was applied (all AABW, see Results) and those
368 for which particle releases were not simulated (samples 11, 13, 17 and 22), the test was repeated with
369 these samples removed. To confirm that the advection effect was directional, i.e. that "upstream" sites
370 were acting as sources of diversity to "downstream" sites, SourceTracker[46] was used to identify
371 sources of OTUs in each sample. Each sample was sequentially designated a sink, with the remaining
372 samples as potential sources, and the most probable proportion of OTUs originating from each
373 potential source determined over 100 randomised trials per sample. Spearman's rank correlation was
374 then calculated between the SourceTracker predicted source proportions and particle encounter source
375 proportions for each sample pairwise, with right-tailed p-value determined by permutation. To explore
376 the role of taxonomic resolution in detecting a significant advection effect, OTU clustering was
377 repeated as above at sequence similarity thresholds of 0.80, 0.85, 0.90, 0.95, 0.99 and 1.00.
378 Generation of relative abundance profiles and partial Mantel tests were performed as above.
379

380 **References**

381 1. Baas Becking, L. G. M. *Geobiologie Of Inleiding Tot De Milieukunde*. (W.P. Van Stockum &
382 Zoon, 1934).
383 2. de Wit, R. & Bouvier, T. '*Everything is everywhere*, but, *the environment selects*'; what did
384 Baas Becking and Beijerinck really say? *Environ. Microbiol.* **8,** 755–758 (2006).
385 3. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nature*
386 *Rev. Microbiol.* **4,** 102–112 (2006).
387 4. Cho, J.-C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent
388 *Pseudomonas* strains in soil. *Appl. Env. Microbiol.* **66,** 5448–5456 (2000).

389 5. Ramette, A. & Tiedje, J. M. Multiscale responses of microbial life to spatial distance and
390 environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci. USA* **104,** 2761–
391 2766 (2007).
392 6. Storch, D. & Sizling, A. L. The concept of taxon invariance in ecology: do diversity patterns
393 vary with changes in taxonomic resolution? *Folia Geobot.* **43,** 329–344 (2008).
394 7. Bissett, A., Richardson, A. E., Baker, G., Wakelin, S. & Thrall, P. H. Life history determines
395 biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol. Ecol.*
396 **19,** 4315–4327 (2010).
397 8. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond
398 biogeographic patterns: processes shaping the microbial landscape. *Nature Rev. Microbiol.* **10,**
399 497–506 (2012).
400 9. Galand, P. E., Potvin, M., Casamayor, E. O. & Lovejoy, C. Hydrography shapes bacterial
401 biogeography of the deep Arctic Ocean. *Nature* **4,** 564–576 (2009).
402 10. Agogué, H., Lamy, D., Neal, P. R., Sogin, M. L. & Herndl, G. J. Water mass-specificity of
403 bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol.*
404 *Ecol.* **20,** 258–274 (2011).
405 11. Hamilton, A. K., Lovejoy, C., Galand, P. E. & Ingram, R. G. Water masses and biogeography
406 of picoeukaryote assemblages in a cold hydrographically complex system. *Limnol. Oceangr.*
407 922–935 (2008).
408 12. Hamdan, L. J. *et al.* Ocean currents shape the microbiome of Arctic marine sediments. *ISME J.*
409 **7,** 685–696 (2013).
410 13. Sul, W. J., Oliver, T. A., Ducklow, H. W., Amaral-Zettler, L. A. & Sogin, M. L. Marine
411 bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. USA* **110,** 2342–2347 (2013).
412 14. Ghiglione, J.-F. *et al.* Pole-to-pole biogeography of surface and deep marine bacterial
413 communities. *Proc. Natl. Acad. Sci. USA* **109,** 17633–17638 (2012).
414 15. Giebel, H.-A., Brinkhoff, T., Zwisler, W., Selje, N. & Simon, M. Distribution of *Roseobacter*
415 RCA and SAR11 lineages and distinct bacterial communities from the subtropics to the
416 Southern Ocean. *Environ. Microbiol.* **11,** 2164–2178 (2009).

417 16. Lauro, F. M., Chastain, R. A., Blankenship, L. E., Yayanos, A. A. & Bartlett, D. H. The
418 unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl. Env.*
419 *Microbiol.* **73,** 838–845 (2007).
420 17. Declerck, S. A. J., Winter, C., Shurin, J. B., Suttle, C. A. & Matthews, B. Effects of patch
421 connectivity and heterogeneity on metacommunity structure of planktonic bacteria and viruses.
422 *ISME J.* **7,** 533–542 (2013).
423 18. Gordon, A. L. *Structure of Antarctic waters between 20°W and 170°W*. (Am. Geogr. soc.,
424 1967).
425 19. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the
426 Antarctic Circumpolar Current. *Deep-Sea Res. Pt. I* **42,** 641–673 (1995).
427 20. Williams, T. J. *et al.* The role of planktonic Flavobacteria in processing algal organic matter in
428 coastal East Antarctica revealed using metagenomics and metaproteomics. *Environ. Microbiol.*
429 **15,** 1302–1317 (2013).
430 21. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities.
431 *Nature* **420,** 806–810 (2002).
432 22. Brown, M. V. *et al.* Global biogeography of SAR11 marine bacteria. *Mol. Syst. Biol.* **8,** 595
433 (2012).
434 23. Wilkins, D. *et al.* Key microbial drivers in Antarctic aquatic environments. *FEMS Microbiol.*
435 *Rev.* **37,** 303–335 (2013).
436 24. Murray, A. E. *et al.* Seasonal and spatial variability of bacterial and archaeal assemblages in
437 the coastal waters near Anvers Island, Antarctica. *Appl. Env. Microbiol.* **64,** 2585–2595
438 (1998).
439 25. Grzymski, J. J. *et al.* A metagenomic assessment of winter and summer bacterioplankton from
440 Antarctica Peninsula coastal surface waters. *ISME J.* **6,** 1901–1915 (2012).
441 26. López-García, P., López-López, A., Moreira, D. & Rodríguez-Valera, F. Diversity of free-
442 living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol. Ecol.*
443 **36,** 193–202 (2001).
444 27. Youssef, N. *et al.* Comparison of species richness estimates obtained using nearly complete

445 fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based
446 environmental surveys. *Appl. Env. Microbiol.* **75,** 5227–5236 (2009).
447 28. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad.*
448 *Sci. USA* **106,** 15527–15533 (2009).
449 29. Wilkins, D. *et al.* Biogeographic partitioning of Southern Ocean microorganisms revealed by
450 metagenomics. *Environ. Microbiol.* **15,** 1318–1333 (2013).
451 30. Sokolov, S. & Rintoul, S. R. Structure of Southern Ocean fronts at 140ºE. *J. Marine Syst.* **37,**
452 151–184 (2002).
453 31. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic
454 through eastern tropical Pacific. *PLoS Biol.* **5,** e77–e77 (2007).
455 doi:10.1371/journal.pbio.0050077
456 32. Ng, C. *et al.* Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace
457 Lake, Antarctica. *ISME J.* **4,** 1002–1019 (2010).
458 33. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data.
459 *Nat. Meth.* **7,** 335–336 (2010).
460 34. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,**
461 2460–2461 (2010).
462 35. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing
463 and web-based tools. *Nucleic Acids Res.* **41,** D590–6 (2013).
464 36. Rosenberg, M. & Rintoul, S. R. *Aurora Australis marine science cruise AU1203 –*
465 *oceanographic field measurements and analysis*. (Antarctic Climate and Ecosystems
466 Cooperative Research Centre, 2012).
467 37. Clarke, K. R. & Gorley, R. N. *PRIMER v6: User Manual / Tutorial*. 1–194 (PRIMER-E,
468 2006).
469 38. Legendre, P. & Anderson, M. J. Distance-based redundancy analysis: testing multispecies
470 responses in multifactorial ecological experiments. *Ecol. Monogr.* **69,** 1–24 (1999).
471 39. Mazloff, M. R., Heimbach, P. & Wunsch, C. An eddy-permitting Southern Ocean state
472 estimate. *J. Phys. Oceanogr.* **40,** 880–899 (2010).

40. Wunsch, C. & Heimbach, P. Practical global oceanic state estimation. *Physica D* **230,** 197–208 (2007).

41. Cerovečki, I., Talley, L. D. & Mazloff, M. R. A comparison of Southern Ocean air-sea buoyancy flux from an ocean state estimate with five other products. *J. Climate* **24,** 6283–6306 (2011).

42. Firing, Y. L., Chereskin, T. K. & Mazloff, M. R. Vertical structure and transport of the Antarctic Circumpolar Current in Drake Passage from direct velocity observations. *J. Geophys. Res.* **116,** C08015 (2011).

43. Paris, C. B., Helgers, J., van Sebille, E. & Srinivasan, A. Connectivity Modeling System: a probabilistic modeling tool for the multi-scale tracking of biotic and abiotic variability in the ocean. *Environ. Model. Softw.* **42,** 47–54 (2013).

44. van Sebille, E., Johns, W. E. & Beal, L. M. Does the vorticity flux from Agulhas rings control the zonal pathway of NADW across the South Atlantic? *J. Geophys. Res.* **117,** C05037 (2012).

45. Rosenberg, M. S. & Anderson, C. D. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods Ecol. Evol.* **2,** 229–232 (2011).

46. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat. Meth.* **8,** 761–763 (2011).

Acknowledgements

This work was supported by the Australian Research Council and the Australian Antarctic Science program. DW was supported by an Australian Postgraduate Research Award. This work was supported in part by the Australian Government's Cooperative Research Centres Program, through the Antarctic Climate and Ecosystems Cooperative Research Centre (ACE CRC), and by the Department of Climate Change and Energy Efficiency through the Australian Climate Change Science Program. We thank Timothy Williams, Sheree Yau, Mark Rosenberg and the crew of the RV *Aurora Australis* for their assistance in collecting samples.

Author contributions

DW, FML and RC conceived the study; DW and EvS performed analyses; FML, SSR and RC provided guidance for the work performed by DW; DW, EvS, SRR and RC wrote the manuscript.

Competing financial interest

The authors declare no competing financial interests.

Accession codes

Pyrosequencing reads for this project have been deposited in the Sequence Read Archive (under project accession PRJNA194155, sample accessions SAMN01991086 to SAMN01991160.

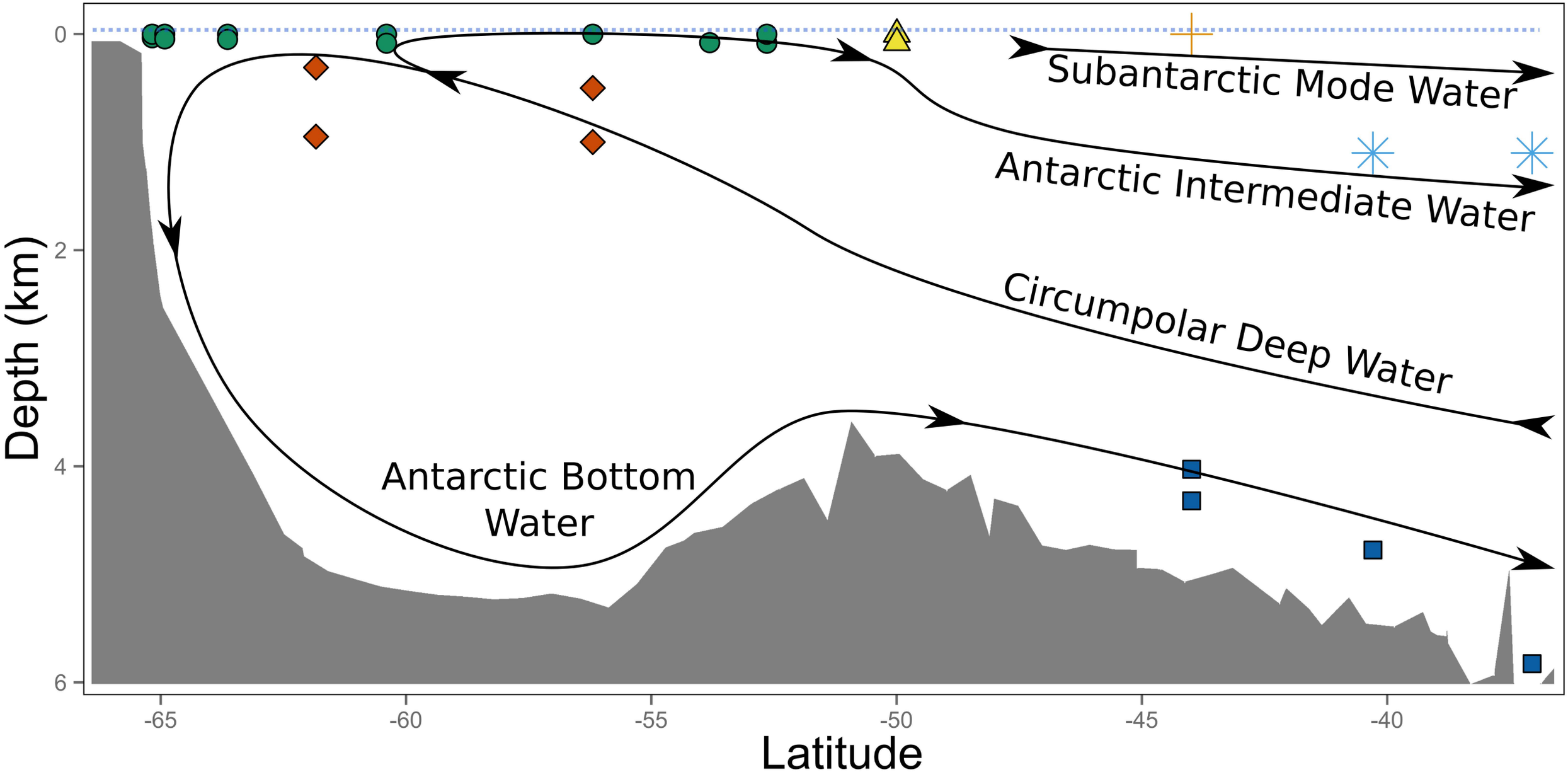
Figure Legends

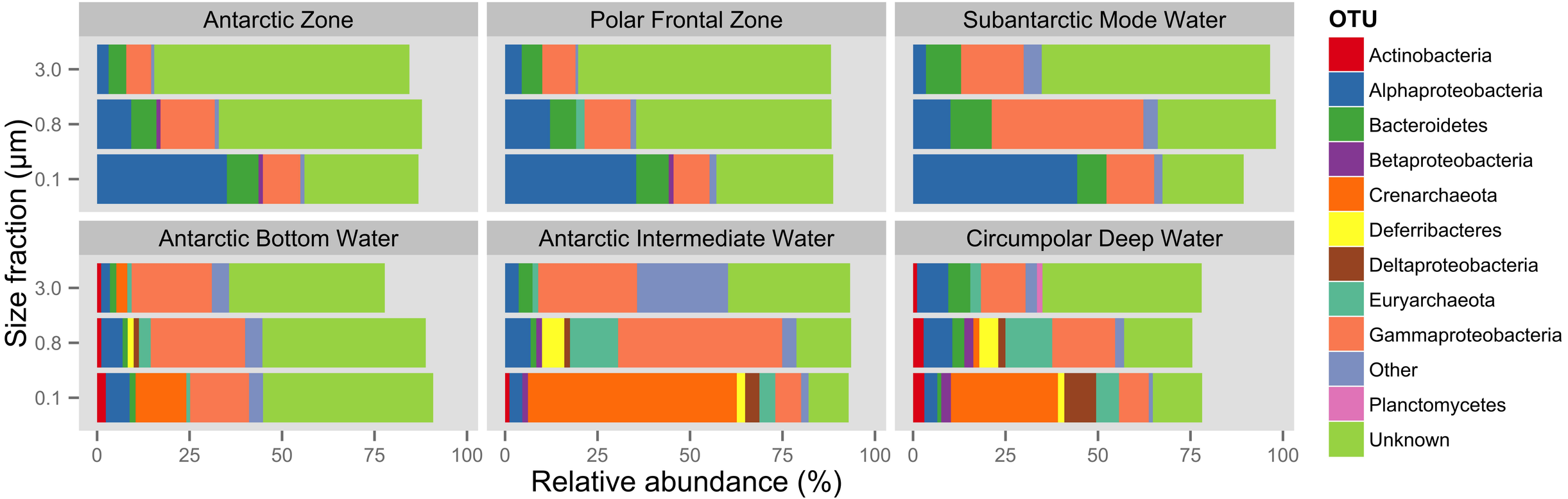
Figure 1 | Map of Southern Ocean sample locations. Antarctic Intermediate Waters (AAIW), light blue stars; Subantarctic Mode Water (SAMW), orange crosses; Antarctic Bottom Water (AABW), dark blue squares; Antarctic Zone (AZ), green circles; Polar Frontal Zone (PFZ), yellow triangles; Circumpolar Deep Water (CDW), red diamonds; sea surface, blue dashed horizontal line. Bathymetry is an approximate representation for 115° E, and is indicative only.

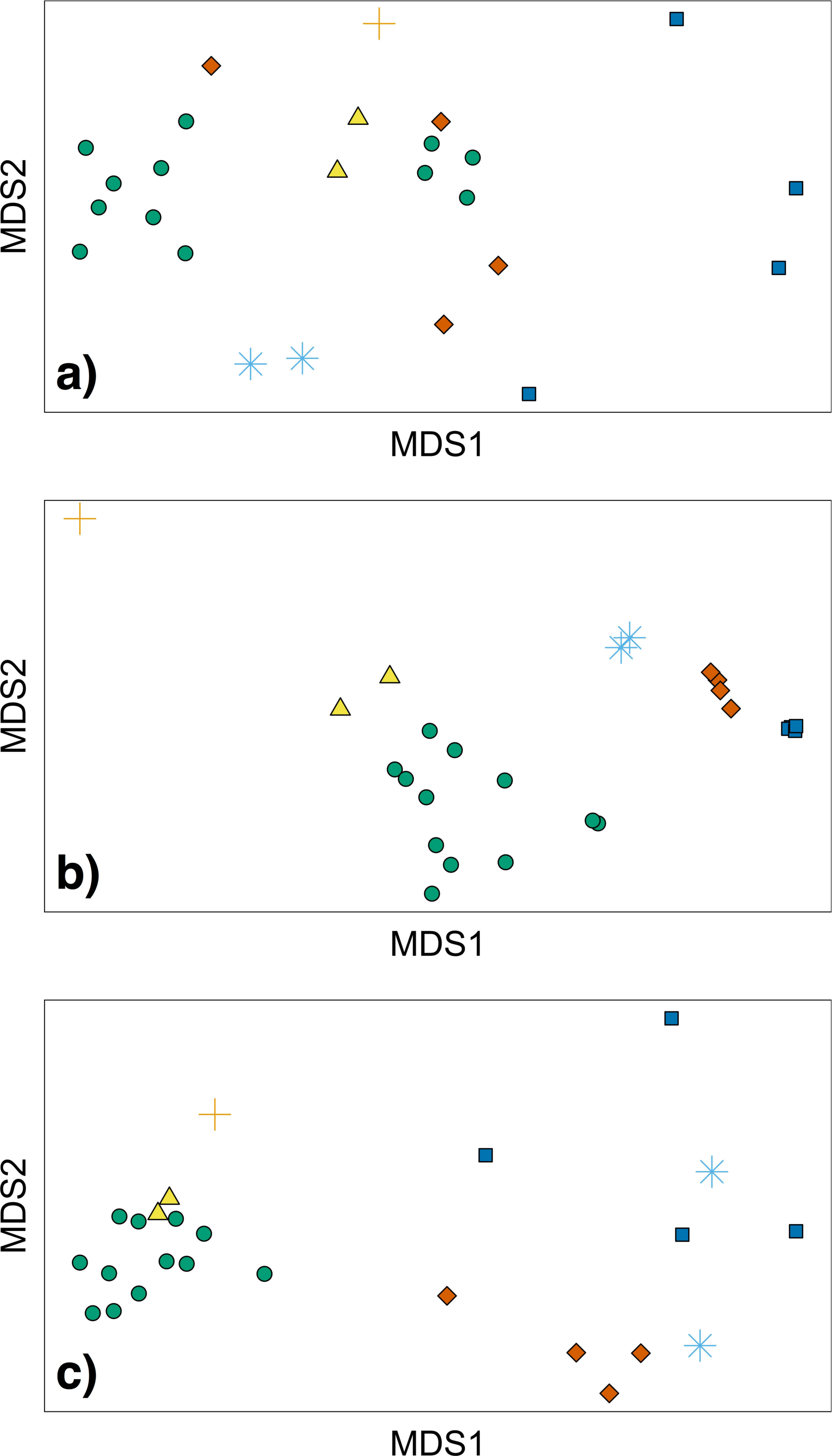
Figure 2 | Taxonomic assignments for each sampled water mass. Relative abundances of OTUs in each sampled water mass and size fraction. OTUs that formed by clustering against SILVA 16S rRNA gene seed sequences were aggregated to the phyla associated with those sequences, except for members of the Proteobacteria which were aggregated to class when known. OTUs generated by *de novo* clustering were aggregated as “Unknown”. For ease of comprehension, phyla or classes with a mean relative abundance across all samples and fractions < 1% were aggregated into a single group labeled "Other". Relative abundance was scaled to account for discarded reads. Surface water masses are in the top row, and deeper water masses are in the bottom row.

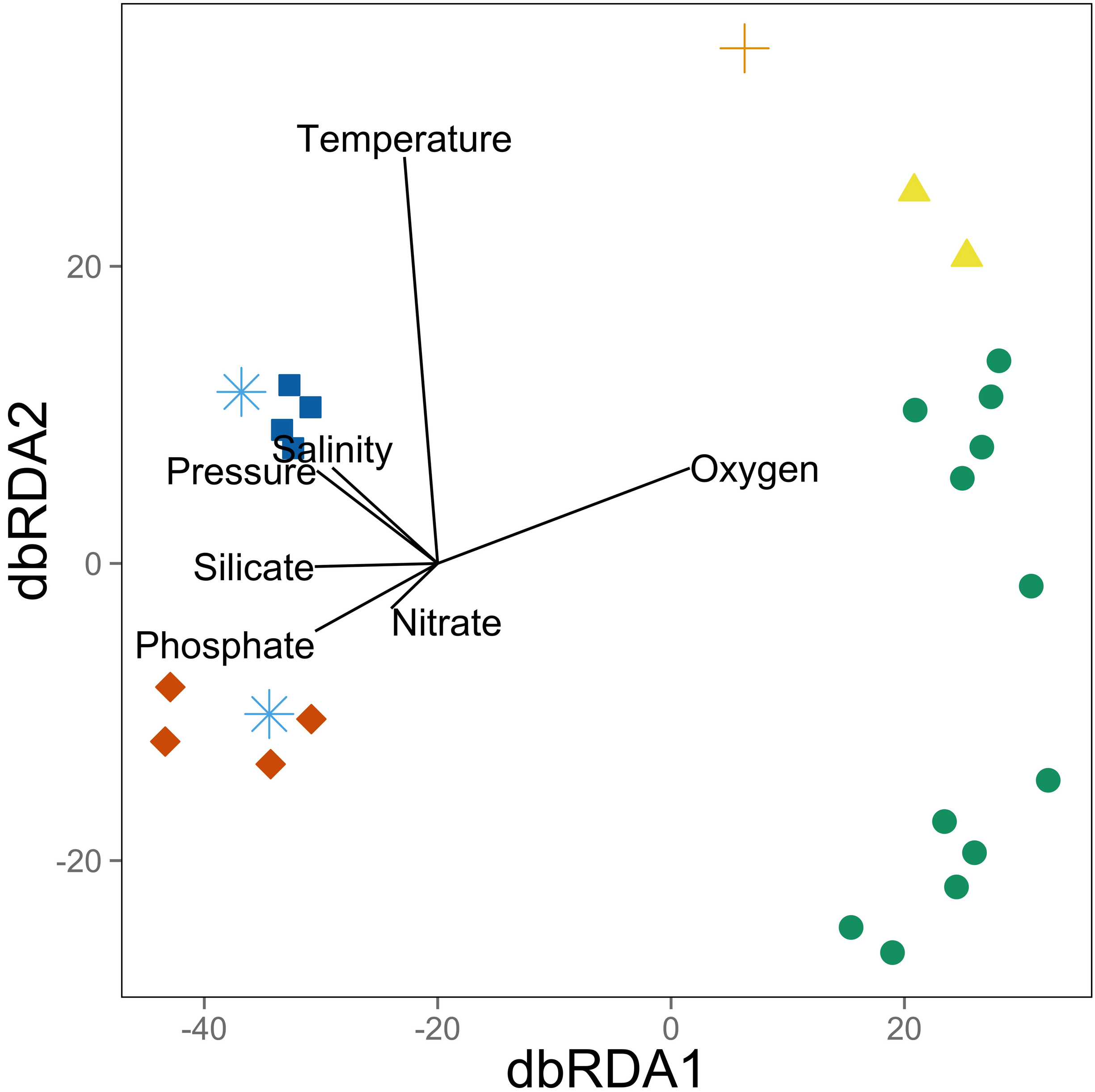
527 **Figure 3 | nMDS ordinations.** First two nMDS dimensions of the (**a**) advection (2D stress = 0.19),
528 (**b**) environmental (2D stress = 0.02) and (**c**) taxonomic (2D stress = 0.08) distance/dissimilarity
529 matrices. Antarctic Intermediate Waters (AAIW), light blue stars; Subantarctic Mode Water
530 (SAMW), orange crosses; Antarctic Bottom Water (AABW), dark blue squares; Antarctic Zone (AZ),
531 green circles; Polar Frontal Zone (PFZ), yellow triangles; Circumpolar Deep Water (CDW), red
532 diamonds.
533
534 **Figure 4 | Relationship between physicochemical variables and sample taxonomic composition.**
535 dbRDA ordination of the distLM model describing the relationship between physicochemical
536 variables and the taxonomic dissimilarity between samples. Vectors represent the effect of each
537 predictor variable on the two visualised axes. Vector length corresponds to the relative size of the
538 effect, while direction represents the correlations to the two displayed axes. The first axis (dbRDA1)
539 captures 51% of fitted and 26% of total variation between the samples' taxonomic profiles; the second
540 (dbRDA2) captures 16% of fitted and 8% of total variation. Antarctic Intermediate Waters (AAIW),
541 light blue stars; Subantarctic Mode Water (SAMW), orange crosses; Antarctic Bottom Water
542 (AABW), dark blue squares; Antarctic Zone (AZ), green circles; Polar Frontal Zone (PFZ), yellow
543 triangles; Circumpolar Deep Water (CDW), red diamonds.
544
545 **Figure 5 | Taxonomic resolution.** Effect of taxonomic resolution (OTU clustering sequence
546 similarity threshold) on partial Mantel correlation (r) between advection and taxonomic distances,
547 with environmental and spatial distances held constant. All correlations are statistically significant
548 (right-tailed p from 999 permutations < 0.05). The full set of 25 samples was used.
549
550 ## Supplementary Data File Legends
551 **Supplementary Data 1 | Full sample data.** Full location, summary of taxonomic assignments and
552 full physicochemical data for each of the 25 samples in this study. Units are given in column headers.
553

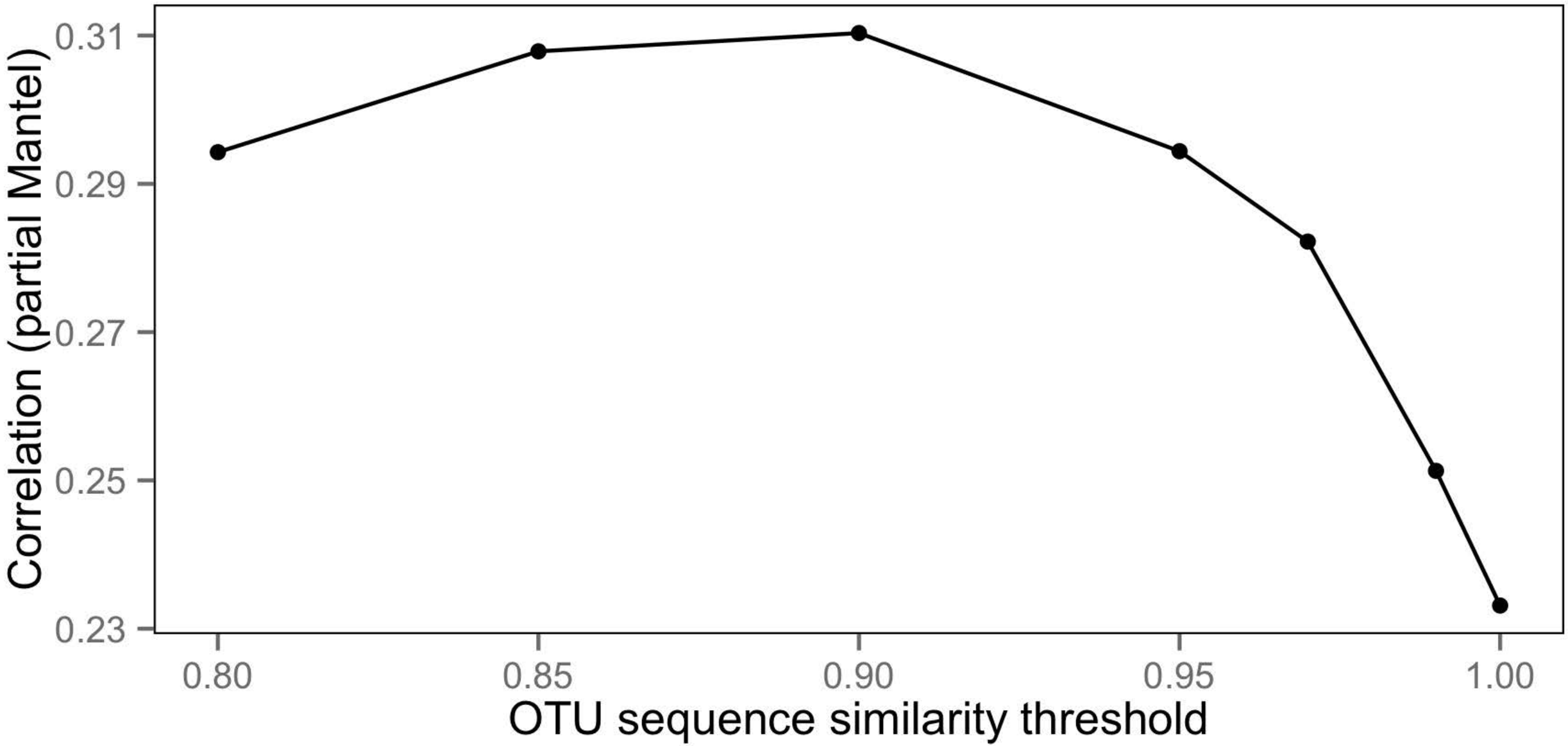
554 **Supplementary Data 2 | All OTU assignments.** Complete list of OTU assignments, with relative
555 abundances, for each sample in this study. OTU labels are either the complete SILVA lineage (for
556 OTUs clustered against a SILVA reference sequence) or an automatically assigned number prefixed
557 with "deNovo" (for OTUs clustered *de novo*). Relative abundances have been scaled to account for
558 excluded reads.
559
560 **Supplementary Data 3 | dbRDA correlations.** Coordinate scores for each sample and correlations
561 between each retained physicochemical variable and dbRDA axis.


Subantarctic Mode Water
Antarctic Intermediate Water
Circumpolar Deep Water
Antarctic Bottom Water
Depth (km)
0
2
4
6
-65
-60
-55
-50
-45
-40
Latitude


Antarctic Zone
Polar Frontal Zone
Subantarctic Mode Water
Antarctic Bottom Water
Antarctic Intermediate Water
Circumpolar Deep Water
Size fraction (µm)
3.0
0.8
0.1
0
25
50
75
100
Relative abundance (%)
OTU
Actinobacteria
Alphaproteobacteria
Bacteroidetes
Betaproteobacteria
Crenarchaeota
Deferribacteres
Deltaproteobacteria
Euryarchaeota
Gammaproteobacteria
Other
Planctomycetes
Unknown


MDS2
a)
MDS1
MDS2
b)
MDS1
MDS2
c)
MDS1


Temperature
Salinity
Pressure
Oxygen
Silicate
Nitrate
Phosphate
20
0
-20
-40
-20
0
20
dbRDA2
dbRDA1


Correlation (partial Mantel)
0.31
0.29
0.27
0.25
0.23
0.80
0.85
0.90
0.95
1.00
OTU sequence similarity threshold