Animals, protists and bacteria share marine biogeographic patterns

Holman, Luke E.; De Bruyn, Mark; Creer, Simon; Carvalho, Gary; Robidart, Julie; Rius, Marc

Nature Ecology and Evolution

Animals, protists and bacteria share marine biogeographic patterns

Luke E. Holman[1,*] (ORCID: 0000-0002-8139-3760)

Mark de Bruyn[2,3] (ORCID:0000-0003-1528-9604)

Simon Creer[3]

Gary Carvalho[3]

Julie Robidart[4] (ORCID: 0000-0001-9805-3570)

Marc Rius[1,5] (ORCID: 0000-0002-2195-6605)

[1] School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, United Kingdom

[2] The University of Sydney, School of Life and Environmental Sciences, Australia

[3] Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, United Kingdom

[4] Ocean Technology and Engineering Group, National Oceanography Centre Southampton, United Kingdom

[5] Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South Africa

* Corresponding author: lukeearlholman@gmail.com

Abstract

Over millennia, ecological and evolutionary mechanisms have shaped macroecological patterns across the tree of life. Research describing these patterns at both regional and global scales has traditionally focussed on the study of metazoan species. Consequently, there is a limited understanding of cross-phyla biogeographic structuring, and an escalating need to understand the macroecology of both microscopic and macroscopic organisms. Here we used environmental DNA (eDNA) metabarcoding to explore the biodiversity of marine metazoans, protists and bacteria along an extensive and highly heterogeneous coastline. Our results showed remarkably consistent biogeographic structure across the kingdoms of life despite billions of years of evolution. Analyses investigating the drivers of these patterns for each taxonomic kingdom found that environmental conditions, such as temperature, and to a lesser extent, anthropogenic stressors such as fishing pressure and pollution, explained some of the observed variation. Additionally, metazoans displayed biographic patterns that suggested regional biotic homogenisation. Against the backdrop of global pervasive anthropogenic environmental change, our work highlights the importance of considering multiple domains of life to understand the maintenance and drivers of marine biodiversity patterns across broad taxonomic, ecological and geographical scales.

Introduction

Researchers have long recognised the importance of grouping biota into distinct, geographically separated regions. Delineating these biogeographic areas is important to understand the factors shaping the range limits of species[1], to designate key geographic areas for biodiversity conservation[2] and to predict biological responses to environmental change[3,4]. One of the first efforts to define geographic regions of terrestrial biota were Alfred Russel Wallace's so-called 'Zoological Regions'[5], which included six major regions (hereafter realms) that are still recognised today[6]. The drivers responsible for these geographic classifications are predominantly environmental conditions or physical barriers. Biogeographic studies have shown that deep divergence in the geographic arrangement of terrestrial biota arose as a result of plate tectonics, while shallow divergence has been most frequently attributed to climatic conditions[7]. In aquatic ecosystems, the relative importance of biogeographic drivers is less understood, although both climatic (e.g. temperature)[8] and tectonic forces[9] have been identified as key determinants of marine biogeographic patterns. Recent studies have partitioned the oceans into distinct ecoregions (i.e. a geographically defined area, smaller than a realm, that contains characteristic species assemblages)[1,10], but the description of marine ecoregions has mostly considered conspicuous or well-described species. Similarly, most marine biogeographic research has focussed on readily identifiable eukaryotic species, principally metazoans[11], although considerable progress has been made in understanding global patterns of marine microbes[12]. In line with recent studies demonstrating strong cross-phyla interdependence[13], there is an increasing need to include prokaryotic species in our assessment of biogeographic patterns. The language of macroecology and microbial ecology is similar, both examining the incidence of species across different spatial scales, but these fields have long progressed independently. As a result, relatively few studies have explored biogeographic patterns simultaneously for both microscopic and macroscopic life, with examples of consistent and inconsistent patterns across different taxa[13-15]. Work is thus needed to explore the consistency of biogeographic breaks across different kingdoms of life.

Human driven habitat destruction, pollution and the introduction of non-native species are key drivers of recent global biodiversity change[3] and therefore have the potential to alter geographic

patterns of biota across multiple spatial scales. Cumulatively, anthropogenic stressors not only threaten vulnerable native species but also whole-community structure and function[3,16,17]. The magnitude and direction of human impacts are complex, with evidence for both gains and losses in local species richness across biomes[18-20]. However, a consistent global pattern is emerging, with a recent and rapid increase in species turnover[19] and an associated increase in community similarity (β diversity) between two or more geographically separated sites[20]. Incidences of increased community similarity are known as biotic homogenisation[21] and are driven by human activities that promote extinctions of native species and introductions of non-native species. In light of growing evidence that taxonomic, phylogenetic and functional diversity are strongly correlated[22], the homogenisation of biological communities has the potential to negatively affect ecosystem function. It is widely accepted that even uncommon species within an ecological community can contribute significantly to ecosystem function[23], demonstrating the importance of studying inconspicuous species to preserve ecosystem health. Studies have shown evidence for biotic homogenisation around the globe, with examples from plants[18], vertebrates[24] and invertebrates[25] demonstrating alteration of terrestrial biogeographic patterns. However, many studies are of limited taxonomic scope, focussing on highly conspicuous species for which reliable data can be easily produced[24,25]. Thus, most work overlooks inconspicuous species (e.g. microbes and microscopic eukaryotes), which show vastly different reproductive, demographic and dispersal patterns compared to metazoans[13], but are known to be key actors shaping the assembly of ecological communities and ultimately underpin ecosystem functioning[26]. Taken together, a more comprehensive characterisation of ecological communities is clearly needed when testing the role of anthropogenic activities on biogeographic patterns.

The advent of high-throughput sequencing has revolutionised our understanding of microbial life, with studies examining global patterns of prokaryotic life now increasingly common[12]. Moreover, the recent and rapid development of methods to infer the incidence of larger organisms using genetic material isolated from environmental samples (known as environmental DNA or eDNA) has provided an unparalleled ability to identify species across the entire tree of life[14,15,27]. Together, these methods can rapidly generate standardised biodiversity data for entire communities at unprecedented resolution, thereby minimising regional and taxonomic biases. In addition, these datasets can be analysed without complete taxonomic assignment and DNA

4

samples can be repurposed to test novel hypotheses. A common technique is to amplify DNA barcodes from eDNA and use high-throughput sequencing to produce high-resolution biodiversity data. This method (eDNA metabarcoding) has been shown to reliably detect organisms across many different ecosystems[27], but has infrequently been applied to understand spatial patterns of biodiversity across different kingdoms of life[12,15].

A unique geographic setting for testing biogeographic hypotheses is the South African coastline, where two large water masses (the Atlantic and Indian Oceans) meet, and a wide variety of abiotic and biotic conditions are found in a single region. This coastline has three well-defined coastal ecoregions bounded by the cold western boundary Benguela Current and the warm oligotrophic eastern boundary Agulhas Current. These ecoregions have been established on the basis of studies over several decades involving a number of conspicuous metazoan taxa[2,28]. Additionally, there is evidence for human exploitation of marine resources in the region spanning thousands of years[29,30] and some areas of the coastline have been subject to heavy maritime activity for centuries[31]. Other human activities also prevail such as the establishment of aquaculture facilities or the construction of harbours and breakwaters[29,30]. Thus, the South African coastline is an ideal study system to explore the mechanisms shaping biogeographic patterns.

Here we compared the biogeography of multiple marine kingdoms of life along the diverse coastline of South Africa. We first investigated the consistency of biogeographic boundaries across metazoans, protists and bacteria using eDNA metabarcoding. We then tested to what extent these patterns could be explained by anthropogenic and natural environmental factors. We finally evaluated if there was evidence for homogenisation of ecological communities along a coastline that has been affected by human activities for centuries.

Results

DNA sequencing

A total of 66.25 million sequences were produced across the three sequencing runs targeting sections of the standard DNA barcoding region of cytochrome c oxidase subunit I (hereafter COI), the V4 region of the eukaryotic nuclear small subunit ribosomal DNA (hereafter 18S) and the V3-V4 hypervariable region of prokaryotic small subunit ribosomal DNA (hereafter 16S). The number of unfiltered raw reads per experimental sample ranged from 61,958 to 859,580, with an average per sample across all three markers of $347,536 \pm 109,665$ (*s.d.*) (see Supplementary Table 1 for further details). Negative control samples exhibited very low levels of cross-contamination (Supplementary Note 1).

Taxonomic assignments & alpha diversity

After performing taxonomic assignment of sequences to metazoans, protists and bacteria, taxonomically grouped datasets with largest number of observations within each marker were used in subsequent analyses (COI for metazoans, 18S for protists & 16S for bacteria). Analyses for the remaining subsets are shown in Supplementary Note 2 and were consistent with the results presented hereafter. After taxonomic assignment to phyla 1,054, 1,433 and 2,826 ASVs (amplicon sequence variants) were retained for metazoans, protist and bacteria datasets respectively. Across all taxonomically grouped datasets the majority of detected ASVs came from a small number (<5) of phyla or supergroups (Figure 1b). This pattern was consistent across sampling sites and no major changes in the identity of ASVs at phyla or supergroup level across the study region were observed (Figure 1b). Similar patterns were observed in the proportions of taxonomically assigned reads (Extended Data Fig. 1).
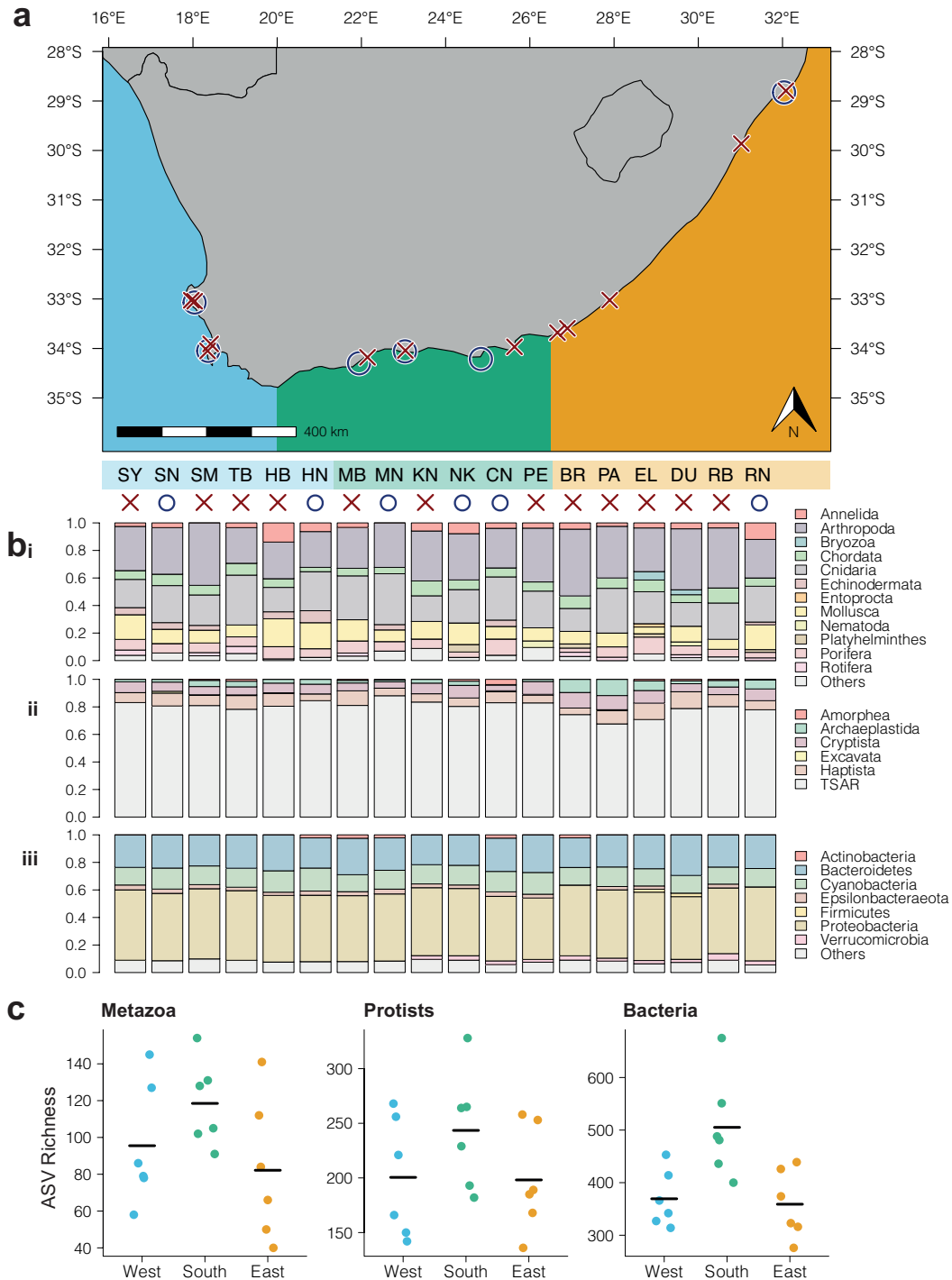
Figure 1. a Map of South Africa indicating the sampling sites and the site types (red crosses are artificial sites and blue circles natural sites), the east, south and west coastal regions are denoted by orange, green and blue respectively. Site codes as in Supplementary Table 9. Landmasses were plotted using the *map* function from the *maps* package in R (v3.6.1)[58]; **b** Proportion of ASVs per phyla across each site for **i** metazoans, **ii** protists and **iii** bacteria. Each bar represents a site indicated by the site code as in Supplementary Table 9; **c** Amplicon sequence variant (ASV) richness per site separated by coast (point colour matches section a) and taxonomic group, black line indicates mean ASV richness.

Across all markers the greatest mean ASV richness was found along the southern coast (Fig. 1). However, a one-way analysis of variance (ANOVA) showed a significant difference ($F_{2,15}=7.18$, p=0.007) between coastlines only in the bacterial dataset with no difference found in both the metazoan ($F_{2,15}=1.941$, p=0.178) and the protist datasets ($F_{2,15}=1.416$, p=0.273). A post-hoc Tukey test of the bacterial data (Supplementary Table 2) showed that the east and west coasts had significantly fewer ASVs compared to the south coast, but that they were not significantly different to one another in overall ASV richness.

Beta diversity

Across all three taxonomic groups, non-metric multidimensional ordinations showed clustering of sites consistent with coastal ecoregions previously described in conspicuous metazoan species (Fig. 2). Furthermore, permutational multivariate analysis of variance (PERMANOVA) models showed a significant (p<0.001) effect of coastline in all cases (see Supplementary Table 3 for model output), with pairwise significant differences (p<0.01) between all pairs of coastlines in all taxa (Supplementary Table 4), and data subsets (Extended Data Fig. 2). There was evidence for heterogeneity of multivariate dispersion in the bacterial dataset (ANOVA on *betadispr* $F_{2,15}=4.09$, p=0.038, Supplementary Table 5). A Tukey test revealed a significant (p=0.031) pairwise difference between the east and west coast only, in line with the observations of Fig. 2., indicating that for bacteria, sites on the west coast were more variable in community composition than the more homogenous communities found at sites on the east coast. When the datasets were split by phyla, they demonstrated significant differences between ecoregions across all tested phyla (p<0.05 in all cases, see Supplementary Note 2 for full model output and visualisation). A power analysis (Supplementary Note 3) indicated that the number of ASVs allocated to each phylum was sufficient to detect a significant difference given the study design.
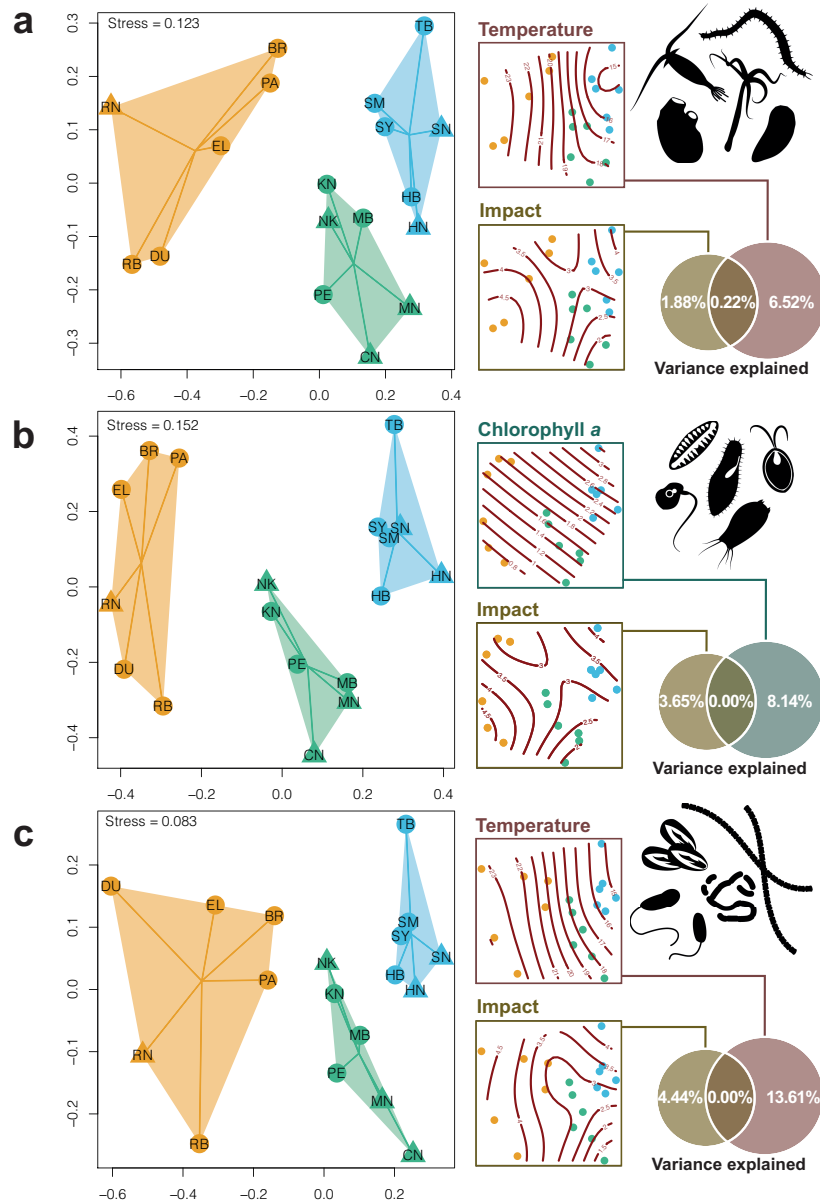
Figure 2. Observed patterns of β-diversity from environmental DNA metabarcoding of: **a** metazoans, **b** protists and **c** bacteria; based on Jaccard dissimilarities between amplicon sequence variants along the coast of South Africa. The first column of plots shows non-metric multidimensional scaling (nMDS) ordinations. Coloured hulls show the spread of the data and lines indicate the spread around the centroid grouped by coast with the east, south and west coasts denoted by orange, green and blue respectively. Site name abbreviations as in Supplementary Table 9, natural sites are denoted with triangles and artificial sites with filled circles. The second column of plots shows the same nMDS ordinations as the first column including the output of a generalised additive model with a 2D smoothed function for each of the significant environmental / impact variables overlaid; temperature – mean sea surface temperature (°C); Chlorophyll *a* – chlorophyll *a* concentration (mg m^{-3}); impact – human marine impact score (unitless measurement, see details in text) against the two nMDS axes. The Venn diagrams indicate the percentage total of the variance in the community dissimilarity explained by each significant variable, derived using variance partitioning of a distance-based redundancy analysis.

Corrected Mantel tests indicated that in the metazoan and bacterial datasets, sea surface temperature (SST) and human impact (measured by an index covering multiple anthropogenic stressors) were significantly correlated with the observed ASV dissimilarities after geographic distance between sites was accounted for (SST $p<0.05$ in all cases; human impact $p<0.01$ in all cases; full model outputs shown in Supplementary Table 6). In the protist dataset chlorophyll *a* concentration and the human impact index remained significant ($p<0.05$ in both cases). In contrast, across all datasets sea surface salinity (SSS) showed no correlation ($p>0.05$ in all cases) with observed ASV dissimilarities. These results indicated that both geographic and environmental distance have some effect on the observed community structure and also confirmed the appropriate variables to retain for analysis in each dataset. In all cases, partial Mantel tests gave similar R statistics and p values (see Supplementary Table 6).

A distanced-based redundancy analysis (dbRDA) showed a significant effect of both environmental variables ($p<0.001$ for SST and chlorophyll *a* in all cases) and human impact ($p<0.05$ in all cases) on the site similarity in metazoans, protists and bacteria (full model outputs are presented in Supplementary Table 7). Variance partitioning of the dbRDA models showed that human impact had a relatively smaller contribution to the observed dissimilarities compared to the chlorophyll *a* concentration or SST (as shown in Fig. 2). Across taxonomic groups there was negligible overlap in the variance explained by human impact and other variables. Generalised additive models with a 2D smoothed function showed significant terms ($p<0.001$, see individual full model outputs shown in Supplementary Table 8), indicating how each variable separately explained variation in the eDNA data in each of the taxonomic groups. SST and chlorophyll *a* concentration showed surfaces across nMDS plots for all markers (Fig. 2) that were simple, with gradients that were consistent across ecoregions. In contrast, human impact scores showed more complex surfaces with multiple peaks across ecoregions.

Distance-decay
Distance-decay slopes for all observations showed an exponential decrease in compositional similarity as the distance between sites increased (Fig. 3a). Regression models of log10 transformed compositional similarity indicating that this slope was statistically significant in all

cases (p<0.001 for all taxonomic groups, full model output in Supplementary Note 4). Comparisons between artificial sites (e.g. recreational marinas, harbours) and natural sites (e.g. relatively unaltered rocky shore and natural harbours) in the metazoan dataset showed a significant difference in the slope between artificial and natural site comparisons ($F_{3,76}$=47.73, p < 0.001). No statistically significant differences were found between site types in the protist or bacteria data, and the same pattern was observed within taxonomic groups for the metazoan and protist datasets across both the COI and 18S data (Supplementary Note 4).
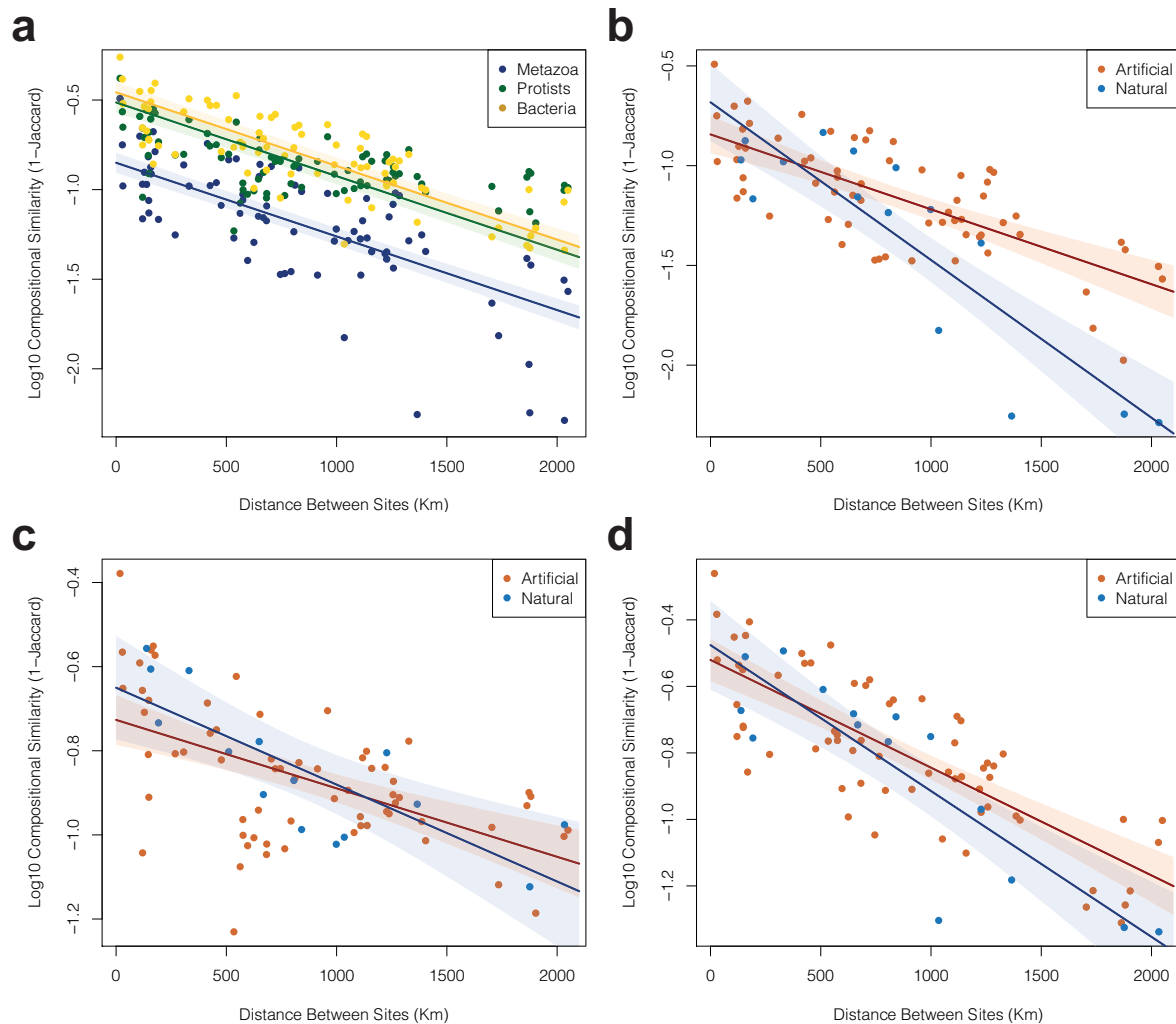
Figure 3. Plots showing distance between sites and community similarity measured using environmental DNA metabarcoding across South Africa. Logarithmically (base 10) transformed compositional similarity against distance is shown in **a**, which includes all datasets. Comparisons between artificial (coloured in red) and natural (coloured blue) sites are shown for **b** metazoans, **c** protists and **d** bacteria. 95% confidence intervals from the regression models are shown as light shaded areas around each regression slope.

Discussion

Here, we showed that metazoans, protists and bacteria have similar biogeographic patterns along an extensive and heterogenous coastline. We found that these remarkably consistent patterns could be partially explained by measured environmental conditions (chlorophyll *a* and temperature), and to a lesser extent, cumulative human impacts. Additionally, we found evidence for anthropogenically driven homogenisation of communities, but this effect was restricted to metazoans. Collectively, we provide evidence of congruent biogeographic boundaries across vastly different forms of life, and demonstrate that underlying processes, such as anthropogenic alterations, affect biogeographic patterns differentially across taxa.

Prokaryotes and eukaryotes diverged billions of years ago and have since evolved to inhabit a vast range of ecological niches. Previous studies have shown both similar[32,33] and dissimilar[34] patterns of β diversity between macro- and microscopic species across environmental and geographic gradients. Recent work has explored biogeographic regionalisation in marine plankton across kingdoms[15], showing that smaller planktonic organisms such as bacteria may have greater biogeographic structuring compared to larger metazoans or protists[15]. Together this evidence suggests that different ecological processes drive a number of taxon-specific responses to produce patterns that are not universal across ecosystems at different spatial scales[13]. Here we observed similar biogeographic patterns across life's kingdoms (Fig. 2), providing clear evidence of cross-phyla biogeographical congruence.

Our analyses suggest that environmental variables such as temperature or chlorophyll *a* concentration influenced the structure of marine communities across the study region (Fig. 2). Global studies of biogeographic patterns have shown a central role of temperature in the structuring of both microbial[12] and larger planktonic life[15,35] across the oceans. There is growing evidence that the range boundaries of marine organisms closely track their thermal limits[4]. Therefore, a general expectation was that species would remain within their thermal niche resulting in temperature-structured communities as observed here. In contrast to temperature, salinity had a minor role in structuring the studied communities (Fig. 2), an observation previously reported in a global marine analysis[12], with exceptions found in microbial[36] and meiofaunal[37] life in regions with unusually strong salinity gradients (e.g. Baltic Sea). The SSS

range across our study system was very narrow (35.0 - 35.4ppt) and so the negligible observed effect was unsurprising. In the case of protists, biogeographic patterns showed a stronger association with primary productivity (measured here as chlorophyll *a* concentration). Previous research has shown little or no role of productivity in driving coastal and oceanic scale biodiversity patterns[15,35]. However, these studies explored the global role of various environmental variables; the significance of more localised oceanographic systems such as upwelling (as along the western coast of southern Africa) might not be as apparent in global analyses.

Anthropogenic activities are known to alter both the physico-chemical properties of the marine environment and the trophic and ecological properties of ecosystems[38]. In our study system, human impact provided some explanatory power to understand the observed community structure, but to a much lesser extent compared to environmental variables (Fig. 2). The human impact index used here[39] covered a large number of different types of impact (e.g. pollution, shipping intensity) but even this aggregated approach adequately explained a small proportion of the total variation in ASVs observed among sampling sites. Previous work on marine metazoans has shown a strong effect of proximate urbanisation[40] and the ecological drivers produced through anthropogenic activities are well documented[38]. Interestingly, the pervasive and conspicuous urbanisation of the marine environment in the study area showed a much weaker effect on biogeographic patterns than other explanatory variables (Fig. 2). Anthropogenic pressures have become a major ecological driver only relatively recently in evolutionary time, with the most dramatic changes in biodiversity occurring within the 21st century[3]. It is clear that human activities are altering evolutionary trajectories[38], either through extinction, range expansions or contractions. However, our data suggests that centuries of human impacts in our study system have not yet demonstrably altered the main observed biogeographic boundaries across taxa.

Previous work on biotic homogenisation has shown a dramatic effect on whole communities at both regional[18,41] and global scales[24,25]. Here, we found support for biotic homogenisation along the South African coastline only in metazoan species, with a difference in the slope of a distance-decay relationship between artificial and natural sites. This pattern was consistent for metazoans

across the gene regions considered (Supplementary Note 4). Pervasive vessel activity in the region[31], along with evidence that artificial environments are hotspots for biological invasions[42], suggest that introduced metazoans are contributing to homogenisation of coastal communities. Further work should incorporate time series data to explore biotic homogenisation, given the significant but minor role of human impact in structuring ecological communities across the region.

Both environmental parameters and species interactions have a clear-cut effect on marine community structure across kingdoms of life[43], but the comparative role of deterministic (environmental filtering, niche processes, etc.) and stochastic processes (ecological drift, random dispersal, etc.) in explaining the observed patterns remains uncertain. The classical deterministic theory of microbial biogeography (Baas Becking hypothesis[44], often summarised as 'everything is everywhere but the environment selects') postulates that due to vast population sizes and dispersal, microbes are found in all environments and the variation in abiotic conditions selects for those that make up the vast majority of species in each region. This theory ignores neutral processes which have been shown to have a critical role in structuring microbial biogeography across biomes[45]. In line with previous efforts studying deterministic and stochastic processes across taxonomic kingdoms[46,47], we found that the majority of the observed variation could not be fully explained for both prokaryotic and eukaryotic species (Fig. 2). Indeed, recent biogeographic research in the oceans has provided both theoretical[48] and empirical[15] evidence of strong biogeographic patterns driven by both stochastic and deterministic forces, but much of the observed variation between communities remains unexplained. Understanding the comparative roles of different community structuring processes requires a more comprehensive examination of the observed variance between communities, species interactions, and the broader role of the environmental conditions where they live.

Several recent innovations will provide valuable data to help uncover the unexplained variation in community structure. For example, the extraction and analysis of sedimentary ancient DNA allows the reconstruction of high-resolution biodiversity change over time[49], providing evidence to evaluate the role of deterministic processes relative to temporal changes in environmental conditions. In addition, the analysis of co-occurrence networks from molecular data can provide

species interaction hypotheses (e.g. Djurhuus, et al. [14]) that could be used to explore how ecological interactions structure biogeographic patterns. Finally, we expect that very high-resolution multi-spectral remote sensing data (e.g. WorldView-3, $<100m^2$) will provide unparalleled insights into the role of environmental forces structuring the distribution of ecological communities[50].

Methods

Field sampling

We sampled a range of sites along 2,000 km of coastline (Fig. 1) between October and November of 2017 (see details in Supplementary Table 9), covering the three major marine coastal ecoregions of South Africa. In order to assess the effects of anthropogenic impacts, we compared human altered 'artificial' sites (e.g. recreational marinas, harbours) and 'natural' sites (relatively unaltered rocky shore and natural harbour sites) (see Supplementary Table 9). The artificial sites were previously surveyed for marine invertebrate biodiversity by Rius, et al. [51], and six adjacent natural sites were selected for this study. The natural sites were the nearest non-developed sites with matching aspect and exposure (Fig. 1) to each of the artificial sampling sites. Three 400 ml seawater samples were filtered with 0.22 μm polyethersulfone membrane Sterivex filters (Merck Millipore, MA, USA) following the sampling scheme of Holman, et al. [42] at each sampling site. Consequently, we sampled a total of 1,200 ml of seawater per site, a volume that has been shown to differentiate fine scale (<1 km^2) community structure in marine systems[40,42]. Filters were immediately preserved at ambient temperature with the addition of 1.5 ml of Longmire's Solution for preservation until DNA extraction. Field control filters and equipment cleaning blanks were taken, transported, stored and sequenced as the rest of the field samples.

Environmental DNA extraction

We used a PCR-free laboratory separated from the main molecular biology laboratory facilities. No post-PCR or high concentration DNA samples were permitted in the laboratory. All surfaces and lab equipment were cleaned thoroughly before use with 1.25% sodium hypochlorite solution (3:1 dilution of household bleach). DNA extraction followed the SX^{CAPSULE} method from Spens, et al. [52]. Briefly, filters were first externally cleaned with sterile water and Longmire's Solution was removed from the filter outlet using a sterile syringe, 720 μl Buffer ATL (Qiagen, Hilden, Germany) and 80 μl Proteinase K (20mg/ml) was added and filters were incubated overnight at 56°C. The lysate was then removed from the filter inlet and subjected to DNA extraction using the Qiagen DNeasy Blood and Tissue Kit under the manufacturers recommended protocol. DNA was eluted using 200 μl Qiagen Buffer AE and re-eluted once to increase DNA yield. All DNA

samples were checked for PCR inhibition using the Primer Design Internal Positive Control qPCR Kit (Primer Design, Southampton, UK) with 10 µl reactions under the manufacturer recommended protocol. Inhibition was detected by an increase of >1.0 Ct in reactions containing eDNA compared to reactions with extraction controls. As inhibition was detected in a minority of samples, all samples were treated using the Zymo OneStep PCR Inhibition Removal Kit (Zymo Research, California, USA) following the manufacturer recommended protocol. Inhibited samples showed no evidence for inhibition post cleaning.

High throughput eDNA amplicon sequencing

Different sets of primers were used to generate three separate eDNA metabarcoding libraries for all samples. Two gene regions were selected to target broad metazoan/eukaryotic diversity: a 313bp region of COI[53] and a variable length region of 18S[54]. A 16S gene region of variable length was used to target the prokaryotes[55]. Illumina unique double-indexed metabarcoding amplicon libraries were constructed with a two-step PCR protocol as detailed in Holman, et al. [42]. The first PCR setup was performed in a PCR-free laboratory. The three eDNA samples per site were pooled and three independent technical replicates were sequenced per pool. The process per sequenced pool was as follows. The first PCR reaction was conducted in triplicate in a total reaction volume of 20 µl. Each reaction contained 10 µl Amplitaq GOLD 360 2X Mastermix (Applied Biosystems, California, USA), 0.8 µl (5 nmol ml^{-1}) of each forward and reverse primers and 2 µl of undiluted environmental DNA template. The reaction conditions for PCR were an initial denaturation step at 95°C for 10 minutes followed by 20 cycles of 95°C for 30 seconds, variable annealing temp (46°C for COI, 50°C for 18S and 55°C for 16S) for 30 seconds, and extension at 72°C for 1 minute. A final extension at 72°C was performed for 10 minutes. The triplicate first PCR replicates were then pooled and cleaned using AMPure XP beads (Beckman Coulter, California, USA) at 0.8 beads:sample volume ratio following manufacturer's instructions. The second PCR reaction was conducted in a total volume of 20 µl containing 10 µl Amplitaq GOLD 360 2X Mastermix, 0.5 µl (10 nmol ml^{-1}) of both forward and reverse primers and 5 µl of undiluted cleaned PCR product from the first reaction. PCR conditions were an initial denaturation step at 95°C for 10 minutes followed by 15 cycles of 95°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 1 minute. A final extension at 72°C was performed for 10 minutes. PCR 2 products were cleaned using

AMPure XP beads as above. Negative control samples for the filters, extraction kit, PCR1 and 2 were included in library building and sequenced alongside experimental samples. Products were quantified following the manufacturer's instructions using the NEBNext Library Quant qPCR kit (New England Biolabs, Massachusetts, USA) and then normalised and pooled at an equimolar concentration for each marker. Each gene region was sequenced independently using a V3 paired-end 300bp reagent kit on the Illumina MiSeq Instrument with 5% PhiX genomic library added to increase sequence diversity.

Bioinformatics

Raw sequences were de-multiplexed using the GenerateFastQ (v2.0.0.9) module on the MiSeq control software (v3.0.0.105). Cutadapt (v2.3)[56] was used to filter sequences to include only those that contained both the forward and reverse primer sequence across both read pairs for each gene fragment, remaining sequences then had the primer region removed for each gene fragment using the default settings. Sequences were denoised using the DADA2 pipeline (v1.12)[57] in R (v3.6.1)[58] with the default parameters unless noted as follows. Sequences were filtered to retain only pairs of reads with an expected error of 1 or fewer per read. Read trimming was performed after manual examination of the read quality profile, the forward reads were trimmed to 250bp (COI), 240bp (18S) and 240 bp (16S) and the reverse reads were trimmed to 230bp (COI), 220bp (18S) and 220 bp (16S). As each marker was sequenced separately, the differences in read trimming length reflect typical variation in sequencing runs rather than any biological difference. The error rates per run were estimated and used to perform the denoising using the DADA2 algorithm. The denoised sequence pairs were then merged and resulting sequences were truncated if they were outside of the expected gene fragment range (303-323bp for COI, 400-450bp for 18S and 390-450bp for 16S). Chimeras were identified and removed before assembling a sample by ASV table for analysis. The denoised ASVs were then curated using the default settings of the LULU algorithm[59] which merges sequences based on sequence similarity and co-occurrence. Assigning taxonomy to a set of unknown sequences is a difficult task, particularly considering many marine species lack DNA barcodes, are undescribed, or have erroneous barcodes in online public databases. We therefore focused our analysis at a higher taxonomic level than species, assigning taxonomy to sequences from the COI and 18S data as follows. The RDP classifier (v2.13)[60] was used to assign taxonomy for COI using a previously

published COI database[61] (v4.0) and a modified version of the SILVA database[62] (v3.2 from https://github.com/terrimporter/18SClassifier). As species level assignments have been shown to be accurate for COI data[42] an unconstrained (no limits on sequence similarity or match length) BLAST search (v2.6.0+) was performed for each sequence against the entire National Centre for Biotechnology Information *nt* database (downloaded on 16th May 2019), 200 hits per sequence were retained (-*num_alignments*). These sequences were then parsed using an R script to exclude hits below 65% coverage, remaining assignments with percent identity above 97% for COI were used to collapse reads for ASVs assigned to the same species. Recent analyses have suggested that only exact (100% identity) matching of sequences to reference data is appropriate for species assignment for the prokaryotic 16S region[63]. The 16S sequences were matched to the SILVA database (release 132)[62] using the default settings of the *assignTaxonomy* function from the DADA2 package to assign taxonomy at genus level or above. The incidence of NUMTs (nuclear mitochondrial DNA) and chimeras in the final ASV list was evaluated following Supplementary Note 5.

The following quality control filters were applied to the ASV by sample table produced by DADA2. First, the minimum number of reads per observation was set at three. Any ASVs not represented in at least one other sample were discarded. ASVs were then filtered to retain only those found in all three technical replicates. For any ASV found in the negative control samples, the largest value among the read count across all negative control samples was used as the zero value for all other samples (i.e. any smaller values found in non-control samples were set to zero). The COI and 18S datasets were then subset by the RDP classifier taxonomic assignments to produce datasets for the protists and metazoans as follows. Phylum level assignments above a threshold of 30, a value well above that shown to accurately assign phylum level taxonomy [61], were parsed to include phyla that contained only metazoan or protist members for each group respectively, other assignments or unknown assignments were discarded. This resulted in a protist and metazoan dataset for each marker, which were subsequently used for separate analyses using these groupings. The 16S data was parsed to include only bacterial ASVs. Within each taxonomic dataset samples were then rarefied to the smallest number of reads (see Supplementary Table 1). Technical replicates were then collapsed to produce a dataset containing the mean value of rarefied reads per ASV. Finally, ASVs assigned using BLAST

(with no cases of multiple matches of equal quality) to the same species in the COI dataset were combined by summing reads per site. The taxonomic assignment method used for the 16S data assigns to genus, and species level assignments are not possible for all taxa using the selected 18S region, so no ASVs from these datasets were collapsed. In order to explore broad scale patterns of taxonomic diversity, the number of ASVs per phyla and number of rarefied reads per phyla were collapsed to produce per site assessments of taxonomic composition. As phylum level phylogeny is not resolved for all protist species, the protist dataset was grouped by supergroup designations according to Burki, et al. [64]. For plots, phyla represented by less than 2% of ASV counts were concatenated in an 'other' category.

Environmental, human impact and geographic data
In situ environmental data reflects a snapshot of the total conditions experienced across the lifetime of the species that make up marine communities. Therefore, abiotic variables for the sites covering an ecologically relevant timescale were sourced as follows. High resolution (1 km^2) remote sensing average daily sea surface temperature data derived from multiple satellite deployments, combined with *in situ* data[65] was parsed in R to find the nearest datapoint to each site. For each point, a mean from two years of data from November 2017 was calculated. Interpolated average (2005-2017) sea surface salinity data (0.25° grid resolution) generated using gliders, oceanographic casts etc. from the 2018 World Ocean Atlas[66] was parsed to include only surface data for the sites. Monthly global ocean colour data (4 km^2) derived from multiple satellite deployments[67] was parsed to calculate an average value for chlorophyll *a* density per site across two years from November 2017. Finally, a previously described[39] 1 km^2 global resolution cumulative index for anthropogenic impact on marine ecosystems, comprising fishing pressure, climate change, shipping and land-based pollution, was parsed to produce a value for each site cumulatively across the entire period for which data were available (2003-2013). These global datasets have excellent temporal resolution, but are only appropriate for testing large-scale patterns as they have limited ability to discriminate highly localised observations.

Ecological statistics
Analyses were conducted in R (v3.6.1) unless otherwise stated. Differences in the mean number of ASVs per coastline were assessed using an ANOVA after testing for normally distributed

residuals using a Shapiro-Wilk test and equal variance between coasts using a Bartlett test. A Tukey's Honest Significant difference test was used to evaluate significant ANOVA results. Differences in community similarity were assessed using a PERMANOVA[68] implemented with the function *adonis* from the package *vegan* (v2.5-6)[69] to assess differences in multivariate centroids and dispersion between coastlines. The PERMANOVA was conducted on a matrix of Jaccard dissimilarities as this ecological index has been shown to be appropriate for biogeographical studies[70]. Significant pairwise differences were assessed using the function *adonis.pair* from the *EcolUtils* package (v0.1)[71]. To analyse if groups of samples have a difference in intra-group community variation, also known as heterogeneity of multivariate dispersion, the PERMDISP2 procedure[72] was used, implemented in the function *betadisper* from the *vegan* package. The pairwise group differences in heterogeneity of multivariate dispersion in the case of a global significant result from *betadisper* were analysed using a Tukey's Honest Significant difference test. Non-metric multidimensional scaling ordinations (nMDS) were calculated using Jaccard dissimilarities and the function *metaMDS* from the *vegan* package.

The influence of the abiotic and human impact data on the observed patterns of beta diversity were evaluated as follows. It has previously been common to use a partial Mantel test to evaluate the effect of a distance matrix (frequently environmental variables) on a second distance matrix (species composition) while 'cancelling out' the effect of a third matrix (geographic distance). However, this approach has been shown to be sensitive to spatial autocorrelation common in ecological datasets[73]. A recently developed method[73], which corrects spurious inflations of the parameter estimate for Mantel tests, was implemented. Across each taxonomic group Mantel tests were conducted comparing Jaccard dissimilarity against Euclidean distance for each environmental variable. For each test Moran spectral randomisation was performed including the geographic distance data with 10,000 permutations to assess statistical significance using the *msr* function from the *adespatial* package (v0.3-8). Partial Mantel tests were conducted with 10,000 permutations using the *mantel.partial* function from the *vegan* package for comparison.

Explanatory variables which had some correlation with the community dissimilarity after adjusting for geographic distance were then evaluated as follows. First, a distance-based redundancy analysis (dbRDA)[74], regressing site Jaccard dissimilarities against all remaining

variables, was performed using the function *dbrda* from the *vegan* package. The significance of terms was assessed with 10,000 permutations. The dbRDA ordination allows us to examine linear changes in the beta diversity in response to a number of predictor variables in tandem, and also to explore their relative impact. The function *varpart* from the *vegan* package was then used to partition the variance in the community dissimilarity by the environmental variables. We then used a generalised additive model to visualise the variation of each significant variable across the nMDS space via a restricted maximum likelihood 2D smoother, implemented in the function *ordisurf* from the *vegan* package.

Distance-decay relationships were explored by first measuring compositional similarity (1-Jaccard index) for each pair of sites, and then calculating distances between pairs of sites by drawing a continuous transect 1 km offshore parallel to the high-water mark using Google Earth Pro (v7.3.2.5776), taking the distance along the transect to measure distance between sites. We then used these data in least-square regression models using the function *lm* with an interaction function between distance and site type (artificial or natural) terms against compositional similarity as a response term. The compositional similarity values were log10 transformed to linearise the response, untransformed values of zero (no overlap of species) were omitted to avoid infinite response variable values.

Acknowledgments

Author Contributions Statement

L.E.H. and M.R. designed the study, L.E.H. collected samples, generated and analysed the data, prepared all figures and wrote the first draft of the paper. MdB, SC, GC, JR and MR supervised and advised the research. All authors substantially contributed to further manuscript drafts and provided final approval for publication.

Competing Interests Statement

The authors declare no competing interests.

Data Availability

Raw Illumina sequencing data are available from the European Nucleotide Archive under study accession number PRJEB38452, sample specific accessions are provided in Supplementary Data File 1. Associated metadata, R scripts and intermediate files are available at the following DOI: 10.5281/zenodo.4564075.

Code Availability

All code used in the current study can be found at the following DOI: 10.5281/zenodo.4564075.

References

1   Spalding, M. D. *et al.* Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience* **57**, 573-583, doi:10.1641/B570707 (2007).
2   Awad, A. A., Griffiths, C. L. & Turpie, J. K. Distribution of South African marine benthic invertebrates applied to the selection of priority conservation areas. *Diversity and Distributions* **8**, 129-145, doi:10.1046/j.1472-4642.2002.00132.x (2002).
3   Pecl, G. T. *et al.* Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, doi:10.1126/science.aai9214 (2017).
4   Sunday, J. M., Bates, A. E. & Dulvy, N. K. Thermal tolerance and the global redistribution of animals. *Nature Climate Change* **2**, 686-690, doi:10.1038/nclimate1539 (2012).
5   Wallace, A. R. *The geographical distribution of animals, with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface.* (Macmillan and Co., 1876).
6   Holt, B. G. *et al.* An update of Wallace's zoogeographic regions of the world. *Science* **339**, 74-78, doi:10.1126/science.1228282 (2013).
7   Ficetola, G. F., Mazel, F. & Thuiller, W. Global determinants of zoogeographical boundaries. *Nature Ecology and Evolution* **1**, 89, doi:10.1038/s41559-017-0089 (2017).
8   Kocsis, A. T., Reddin, C. J. & Kiessling, W. The stability of coastal benthic biogeography over the last 10 million years. *Global Ecology and Biogeography* **27**, 1106-1120, doi:10.1111/geb.12771 (2018).
9   Zaffosa, A., Finnegan, S. & Peters, S. E. Plate tectonic regulation of global marine animal diversity. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 5653-5658, doi:10.1073/pnas.1702297114 (2017).
10   Costello, M. J. *et al.* Marine biogeographic realms and species endemicity. *Nature Communications* **8**, doi:10.1038/s41467-017-01121-2 (2017).
11   Beck, J. *et al.* What's on the horizon for macroecology? *Ecography* **35**, 673-683, doi:10.1111/j.1600-0587.2012.07364.x (2012).
12   Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359, doi:10.1126/science.1261359 (2015).
13   Shade, A. *et al.* Macroecology to Unite All Life, Large and Small. *Trends in Ecology and Evolution* **33**, 731-744, doi:10.1016/j.tree.2018.08.005 (2018).
14   Djurhuus, A. *et al.* Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature communications* **11**, 254, doi:10.1038/s41467-019-14105-1 (2020).
15   Richter, D. J. *et al.* Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *bioRxiv*, 867739, doi:10.1101/867739 (2019).
16   Naeem, S., Duffy, J. E. & Zavaleta, E. The functions of biological diversity in an age of extinction. *Science* **336**, 1401-1406, doi:10.1126/science.1215855 (2012).
17   Tilman, D., Isbell, F. & Cowles, J. M. Biodiversity and Ecosystem Functioning. *Annual Review of Ecology, Evolution, and Systematics* **45**, 471-493, doi:10.1146/annurev-ecolsys-120213-091917 (2014).
18   Finderup Nielsen, T., Sand-Jensen, K., Dornelas, M. & Bruun, H. H. More is less: net gain in species richness, but biotic homogenization over 140 years. *Ecology Letters* **22**, 1650-1657, doi:10.1111/ele.13361 (2019).

19   Blowes, S. A. *et al.* The geography of biodiversity change in marine and terrestrial assemblages. *Science* **366**, 339-345, doi:10.1126/science.aaw1620 (2019).

20   Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296-299, doi:10.1126/science.1248484 (2014).

21   Olden, J. D. & Rooney, T. P. On defining and quantifying biotic homogenization. *Global Ecology and Biogeography* **15**, 113-120, doi:10.1111/j.1466-822X.2006.00214.x (2006).

22   Stuart-Smith, R. D. *et al.* Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539-542, doi:10.1038/nature12529 (2013).

23   Mouillot, D. *et al.* Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biology* **11**, e1001569, doi:10.1371/journal.pbio.1001569 (2013).

24   Bernardo-Madrid, R. *et al.* Human activity is altering the world's zoogeographical regions. *Ecology Letters* **22**, 1297-1305, doi:10.1111/ele.13321 (2019).

25   Capinha, C., Essl, F., Seebens, H., Moser, D. & Pereira, H. M. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348**, 1248-1251, doi:10.1126/science.aaa8913 (2015).

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

27   Deiner, K. *et al.* Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* **26**, 5872-5895, doi:10.1111/mec.14350 (2017).

28   Emanuel, B. P., Bustamante, R. H., Branch, G. M., Eekhout, S. & Odendaal, F. J. A zoogeographic and functional approach to the selection of marine reserves on the west coast of South Africa. *South African Journal of Marine Science* **12**, 341-354, doi:10.2989/02577619209504710 (1992).

29   Griffiths, C. L., Robinson, T. B., Lange, L. & Mead, A. Marine biodiversity in South Africa: an evaluation of current states of knowledge. *PLoS One* **5**, e12008, doi:10.1371/journal.pone.0012008 (2010).

30   Griffiths, C. L. *et al.* Impacts of human activities on marine animal life in the Benguela: A historical overview. *Oceanography and Marine Biology: an Annual Review* **42**, 303-392, doi:10.1201/9780203507810.ch8 (2004).

31   Kaluza, P., Kolzsch, A., Gastner, M. T. & Blasius, B. The complex network of global cargo ship movements. *Journal of The Royal Society Interface* **7**, 1093-1103, doi:10.1098/rsif.2009.0495 (2010).

32   Rapacciuolo, G., Beman, J. M., Schiebelhut, L. M. & Dawson, M. N. Microbes and macro-invertebrates show parallel β-diversity but contrasting α-diversity patterns in a marine natural experiment. *Proceedings of the Royal Society B: Biological Sciences* **286**, 20190999, doi:doi:10.1098/rspb.2019.0999 (2019).

33   Astorga, A. *et al.* Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography* **21**, 365-375, doi:10.1111/j.1466-8238.2011.00681.x (2012).

34   Wang, J. *et al.* Patterns of elevational beta diversity in micro- and macroorganisms. *Global Ecology and Biogeography* **21**, 743-750, doi:10.1111/j.1466-8238.2011.00718.x (2012).

35   Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098-U1107, doi:10.1038/nature09329 (2010).

36   Herlemann, D. P. *et al.* Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal* **5**, 1571-1579, doi:10.1038/ismej.2011.41 (2011).

37   Broman, E. *et al.* Salinity drives meiofaunal community structure dynamics across the Baltic ecosystem. *Molecular Ecology* **28**, 3813-3829, doi:10.1111/mec.15179 (2019).

38   Shochat, E., Warren, P. S., Faeth, S. H., McIntyre, N. E. & Hope, D. From patterns to emerging processes in mechanistic urban ecology. *Trends in Ecology and Evolution* **21**, 186-191, doi:10.1016/j.tree.2005.11.019 (2006).

39   Halpern, B. S. *et al.* Recent pace of change in human impact on the world's ocean. *Scientific Reports* **9**, 11609, doi:10.1038/s41598-019-47201-9 (2019).

40   Kelly, R. P. *et al.* Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* **4**, e2444, doi:10.7717/peerj.2444 (2016).

41   Blouin, D., Pellerin, S. & Poulin, M. Increase in non-native species richness leads to biotic homogenization in vacant lots of a highly urbanized landscape. *Urban Ecosystems* **22**, 879-892, doi:10.1007/s11252-019-00863-9 (2019).

42   Holman, L. E. *et al.* Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports* **9**, 11559, doi:10.1038/s41598-019-47899-7 (2019).

43   Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science* **348**, 1262073, doi:10.1126/science.1262073 (2015).

44   Baas-Becking, L. G. M. *Geobiologie; of inleiding tot de milieukunde.* (WP Van Stockum & Zoon NV, 1934).

45   Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **10**, 497-506, doi:10.1038/nrmicro2795 (2012).

46   Farjalla, V. F. *et al.* Ecological determinism increases with organism size. *Ecology* **93**, 1752-1759, doi:10.1890/11-1144.1 (2012).

47   Wu, W. X. *et al.* Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME Journal* **12**, 485-494, doi:10.1038/ismej.2017.183 (2018).

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

49   Balint, M. *et al.* Environmental DNA time series in ecology. *Trends in Ecology and Evolution* **33**, 945-957, doi:10.1016/j.tree.2018.09.003 (2018).

50   He, K. S. *et al.* Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* **1**, 4-18, doi:10.1002/rse2.7 (2015).

51   Rius, M. *et al.* Range expansions across ecoregions: interactions of climate change, physiology and genetic diversity. *Global Ecology and Biogeography* **23**, 76-88, doi:10.1111/geb.12105 (2014).

52   Spens, J. *et al.* Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution* **8**, 635-645, doi:10.1111/2041-210x.12683 (2017).

53   Leray, M. *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for

characterizing coral reef fish gut contents. *Frontiers in Zoology* **10**, 34, doi:10.1186/1742-9994-10-34 (2013).

54   Zhan, A. *et al.* High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution* **4**, 558-565, doi:10.1111/2041-210x.12037 (2013).

55   Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS One* **9**, e105592, doi:10.1371/journal.pone.0105592 (2014).

56   Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**, 10-12 (2011).

57   Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).

58   R_Core_Team. R: a language and environment for statistical computing. *ISBN 3-900051-07-0* (2019).

59   Frøslev, T. G. *et al.* Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature communications* **8**, 1188, doi:10.1038/s41467-017-01312-x (2017).

60   Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* **73**, 5261-5267 (2007).

61   Porter, T. M. & Hajibabaei, M. Automated high throughput animal CO1 metabarcode classification. *Scientific Reports* **8**, 4226, doi:10.1038/s41598-018-22505-4 (2018).

62   Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).

63   Edgar, R. C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **34**, 2371-2375, doi:10.1093/bioinformatics/bty113 (2018).

64   Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. The new tree of eukaryotes. *Trends in Ecology & Evolution* **35**, 43-55 (2020).

65   JPL_OurOceanProject. GHRSST Level 4 G1SST Global Foundation Sea Surface Temperature Analysis. doi:10.5067/GHG1S-4FP01 (2010).

66   Zweng, M. M., J. R. Reagan, D. Seidov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, A. V. Mishonov, O. K. Baranova, K. Weathers, C. R. Paver, and I. Smolyar,. World Ocean Atlas 2018, Volume 2: Salinity. 50 (2019).

67   European_Space_Agency. Ocean Colour Climate Change Initiative Dataset, Version 4.2. (2020).

68   Anderson, M. J. in *Wiley Stats Ref: Statistics Reference Online* (eds N. Balakrishnan *et al.*) 1-15 (John Wiley & Sons, Ltd, 2014).

69   Oksanen, J. *et al.* Vegan: community ecology package. *R package, Version 2.5-6* (2011).

70   Kreft, H. & Jetz, W. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* **37**, 2029-2053, doi:10.1111/j.1365-2699.2010.02375.x (2010).

71   Salazar, G. EcolUtils: Utilities for community ecology analysis. *R package version 0.1* (2018).

72   Anderson, M. J. Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* **62**, 245-253, doi:10.1111/j.1541-0420.2005.00440.x (2006).

73   Crabot, J., Clappe, S., Dray, S. & Datry, T. Testing the Mantel statistic with a spatially-constrained permutation procedure. *Methods in Ecology and Evolution* **10**, 532-540, doi:10.1111/2041-210x.13141 (2019).

74   McArdle, B. H. & Anderson, M. J. Fitting Multivariate Models To Community Data: A Comment On Distance-Based Redundancy Analysis. *Ecology* **82**, 290-297, doi:10.1890/0012-9658(2001)082[0290:Fmmtcd]2.0.Co;2 (2001).
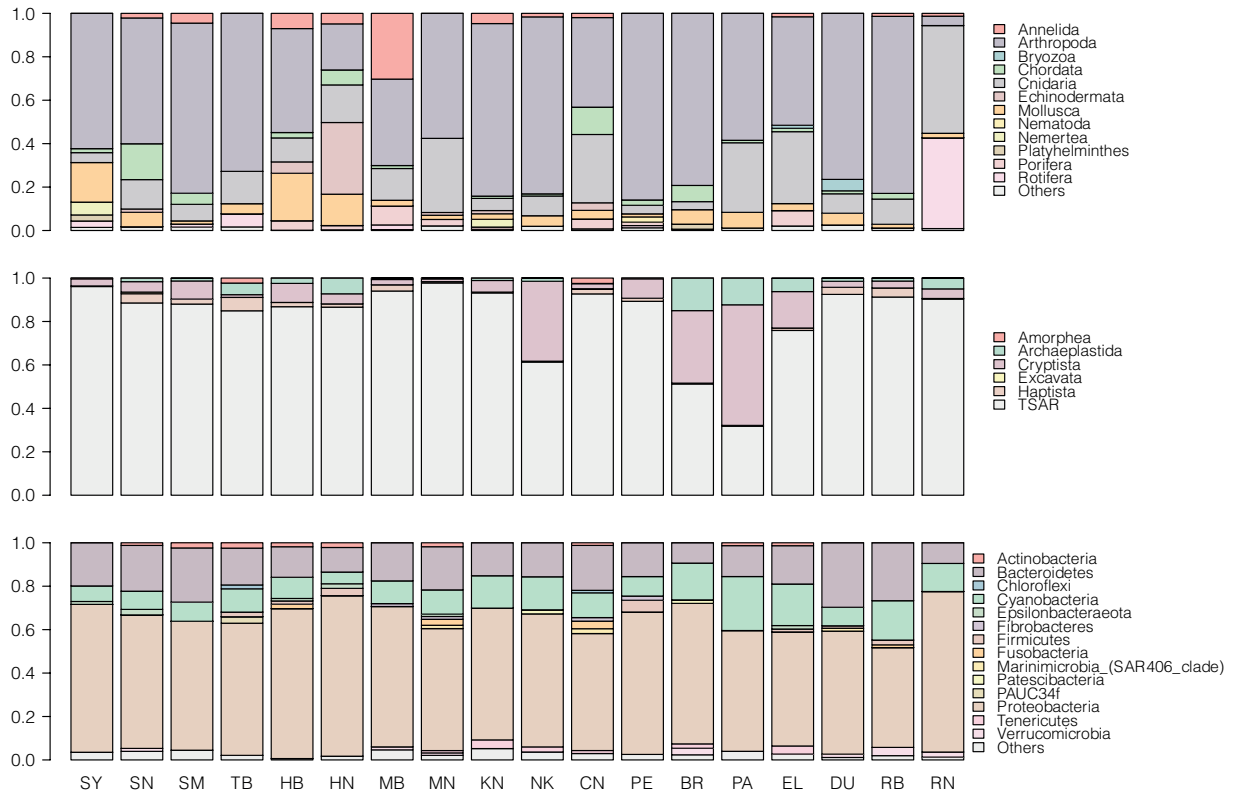
Extended Data Figure 1 Bar charts indicating the proportion of reads assigned per phyla (metazoans/bacteria) or supergroup (protists) from environmental DNA metabarcoding of seawater collected from sites across South Africa. The three rows correspond with data from metazoans (top), protists (middle) and bacteria (bottom). Site name abbreviations as in Supplementary Table 9.
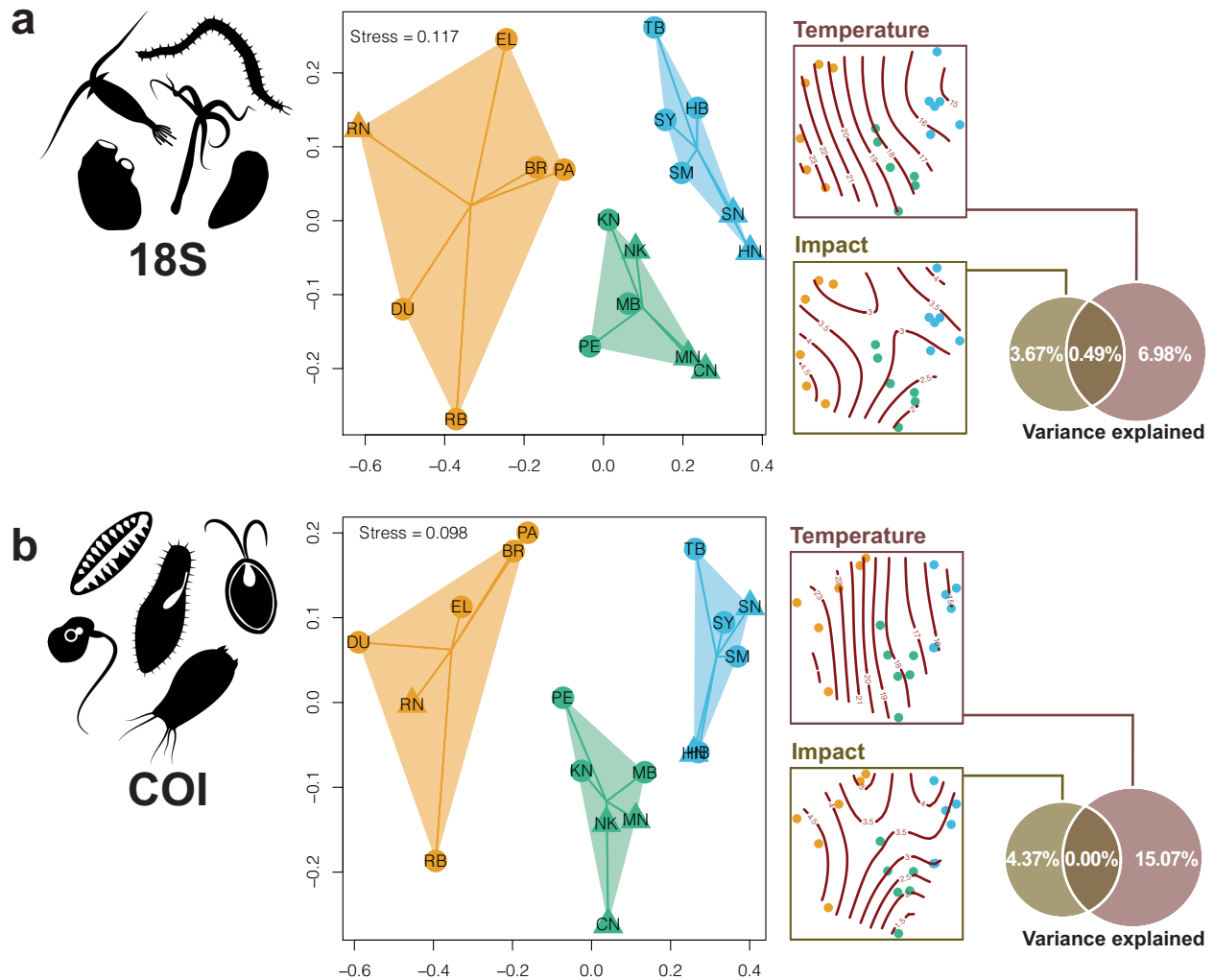
Extended Data figure 2 Observed patterns of β-diversity from environmental DNA metabarcoding of: **a** metazoans from the 18S dataset and **b** protists from the COI dataset; based on Jaccard dissimilarities between amplicon sequence variants along the coast of South Africa. The first column of plots shows non-metric multidimensional scaling (nMDS) ordinations. Coloured hulls show the spread of the data and lines indicate the spread around the centroid grouped by coast with the east, south and west coasts denoted by orange, green and blue respectively. Site name abbreviations as in Supplementary Table 9. Natural sites are denoted with triangles and artificial sites with filled circles. The second column of plots shows the same nMDS ordinations as the first column including the output of a generalised additive model with a 2D smoothed function for each of the significant environmental / impact variables overlaid; temperature – mean sea surface temperature (°C); impact – human marine impact score (unitless measurement, see details in main text) against the two nMDS axes. The Venn diagram charts indicate the percentage total of variance in community dissimilarity explained by each significant variable, derived using variance partitioning of a distance-based redundancy analysis.