# Consistent marine biogeographic boundaries across the tree of life despite centuries of human impacts

Luke E. Holman[1,*] (ORCID: 0000-0002-8139-3760)
Mark de Bruyn[2,3] (ORCID:0000-0003-1528-9604)
Simon Creer[2]
Gary Carvalho[2]
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] Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, United Kingdom
[3] The University of Sydney, School of Life and Environmental Sciences, Australia
[4] Ocean Technology and Engineering Group, National Oceanography Centre Southampton, United Kingdom
[5] Department of Zoology, Centre for Ecological Genomics and Wildlife Conservation, University of Johannesburg, South Africa

* Corresponding author: Lukeearlholman@gmail.com

**Abstract**

Over millennia ecological and evolutionary mechanisms have shaped macroecological distributions across the tree of life. Research describing patterns of regional and global biogeography has traditionally focussed on the study of conspicuous species. Consequently, there is limited understanding of cross-phyla biogeographical structuring, and an escalating need to understand the macroecology of macrobial and microbial life. Here we used environmental DNA metabarcoding to explore the biodiversity of marine metazoans, micro-eukaryotes and prokaryotes along an extensive and highly heterogeneous coastline. We expected environmental forces to be responsible for creating similar biogeographical assemblages across kingdoms of life, while life-history strategies and pervasive anthropogenic impacts were predicted to disturb such patterns. We found that despite centuries of significant human activities, biogeographic patterns were consistent across kingdoms, underpinned primarily by abiotic factors. Our work highlights the importance of studying multiple domains of life to understand the maintenance of biodiversity and potential drivers of future change.

## Introduction

Researchers have long recognised the importance of grouping global 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 areas for biodiversity conservation[2] and in formulating predictive 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 in subsequent classifications today[6]. The drivers responsible for these geographical classifications are frequently environmental conditions or physical barriers. For example, broad tectonic dynamics can produce waterbodies or mountains, and dramatically different climatic conditions are known to define biogeographic boundaries and shape species distributions. 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 climate[7]. In aquatic ecosystems, the relative importance of macroecological drivers is less understood, although both climatic (e.g. temperature)[8] and tectonic forces[9] have been identified as key determinants of biogeographic patterns. Recent studies have successfully partitioned the oceans into distinct ecoregions (i.e. a geographically defined area, smaller than realm, that contains characteristic assemblages of species)[1,10], but the description of marine ecoregions has mostly considered conspicuous or well-described species. Similarly, most macroecological marine research has focussed on easily identifiable eukaryotic species, principally metazoans[11], although some progress has been made in understanding global patterns of marine microbes[12]. In line with recent studies demonstrating strong cross-phylum 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, only a few studies have simultaneously explored biogeographic patterns of both microscopic and macroscopic life to date[13,14], and to our knowledge no study has compared broad-scale biogeographic patterns across different kingdoms of life.

Human activities such as habitat destruction, pollution and the introduction of non-native species are the main drivers of more recent global biodiversity change[3] and therefore have the potential to alter geographic patterns of biota at multiple spatial scales. Cumulatively anthropogenic stressors not only threaten vulnerable native species but whole-community structure and function[3,15,16]. The magnitude and direction of human impacts is complex, with evidence for both gains and losses in local species richness across biomes[17-19]. However, a consistent global pattern is emerging, with a recent and rapid increase in species turnover[18] and an associated increase in community similarity (β diversity) between two or more geographically separated sites[19]. These incidences of increased community similarity are known as biotic homogenisation[20] and are driven by human activities that promote extinctions and the introductions of non-native species. In light of growing evidence that taxonomic, phylogenetic and functional diversity are strongly correlated[21], the homogenisation of biological communities has the potential to negatively affect ecosystem function. Furthermore, even uncommon species within an ecological community can contribute significantly to ecosystem function[22], demonstrating the importance of studying inconspicuous species to preserve ecosystems. Cumulatively, biotic homogenisation directly disrupts biogeographic patterns, threatens endemism at both local and regional scales, and has the potential to affect ecosystem function. Studies have shown evidence for biotic homogenisation around the globe, with examples from plants[17], vertebrates[23] and invertebrates[24] demonstrating alteration of terrestrial biogeographic patterns. While the overall magnitude and direction of this effect is becoming clear, many studies are of limited taxonomic scope, focussing on highly conspicuous species for which reliable data can be relatively easily produced[23,24]. Thus, most studies overlook inconspicuous species (e.g. microbes and microscopic metazoans), which show vastly different reproductive, demographic and dispersal patterns compared to eukaryotes[13], but are known to be key actors shaping the assembly of ecological communities and ultimately underpin ecosystem functioning[25]. Taken together, a more comprehensive characterisation of ecological communities is clearly needed when testing the role of biotic homogenisation 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[26,27]. Together, these methods can rapidly generate standardised biodiversity data for entire communities at an unprecedented resolution, thereby minimising regional and taxonomic biases. Such studies provide datasets that can be analysed without taxonomic assignment and DNA 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[26], but has infrequently been applied to understand spatial patterns of biodiversity across different kingdoms of life[12,28].

The South African 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 delineated by the presence of a number of conspicuous metazoan taxa[2,29] established over decades of study. Additionally, there is evidence for human exploitation of marine resources in the region spanning thousands of years[30,31] and the coastline has been an area of high maritime activity for centuries[32]. Other human activities have also been prevalent such as the establishment of aquaculture facilities or the construction of harbours and breakwaters[30,31]. Thus, this coastline is an ideal study system to explore the mechanisms shaping biogeographic boundaries.

Here we used eDNA metabarcoding to examine the biogeography of multiple marine domains along the ecologically diverse coastline of South Africa. We hypothesised that environmental conditions would be responsible for the maintenance of clear and congruent cross-phyla biogeographic boundaries. In turn, life-history strategies were expected to reshape the geographic arrangement of different phyla. More specifically, we expected that microbial organisms would show a more homogeneous distribution along the studied coastline compared to eukaryotes, mainly due to the greater dispersal capacity of the former. Finally, in line with previous observations, we expected that long-term anthropogenic impacts would disrupt well-defined ecoregion boundaries and lead to an overall homogenisation of biogeographic patterns across the kingdoms of life.

## 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 1), representing the three major marine ecoregions. In order to assess the effects of anthropogenic impacts, we compared human altered 'artificial' sites (e.g. recreational marinas, harbours) and relatively unaltered rocky shore sites, hereafter 'natural' (see Supplementary Table 1). The artificial sites were previously sampled in Rius, et al. [33], and six adjacent natural rocky shore sites were selected, choosing the nearest site with matching aspect and exposure (Fig. 1) to an artificial site wherever possible. For each site, 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. [34]. 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[34,35]. 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. [36]. 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 spiked with eDNA compared to unspiked 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*

Three 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 section of the V4 region of the nuclear small subunit ribosomal DNA (hereafter 18S)[37] and a section of the standard DNA barcoding region of cytochrome c oxidase subunit I (hereafter COI)[38]. One gene region was used to target Prokaryotes, the V3-V4 hypervariable region of prokaryotic small subunit ribosomal DNA (hereafter 16S)[39]. Illumina double-indexed metabarcoding amplicon libraries were constructed with a two-step PCR protocol as detailed in Holman, et al. [34]. The first PCR 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 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 (v 2.0.0.9) module on the MiSeq control software (v 3.0.0.105). Cutadapt (v 2.3)[40] 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 (v 1.12)[41] in R (v 3.6.1)[42] 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 (amplicon sequence variant) table for analysis. The denoised ASVs were then curated using the default settings of the LULU algorithm[43] 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 runs as follows: an unconstrained (no limits on sequence

similarity or match length) BLAST search (v 2.6.0+) was performed for each sequence against the entire nt database (downloaded on 16th May 2019), 200 hits per sequence were retained (-*num_alignments*). These sequences were then parsed using an R script as follows. All hits below 65% coverage and 85% percentage identity were removed, remaining sequences with percent identity above 97% for COI and 99% for 18S were marked as 'high confidence'. Remaining sequences were marked 'low confidence'. For all remaining sequences with conflicting (non-identical) taxonomic assignments an R script selected the lowest common ancestor (broadest taxonomic lineage) that covered all assignments and assigned this taxa to the sequence. The NCBI taxonomy (downloaded 16th May 2019) was used for the lowest common ancestor and the *NCBItax2lin* utility from Mahmoudabadi and Phillips [44] was used to convert NCBI taxonomic IDs into lineages. 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[45]. The 16S sequences were matched to the SILVA database (release 132)[46] 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 Information 1.

*Abiotic, human impact and geographic data*

*In situ* temperature 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 (SST) data derived from multiple satellite missions combined with *in situ* data[47] 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 (SSS) data (0.25° grid resolution) generated using gliders, oceanographic casts etc. from the 2018 World Ocean Atlas[48] was parsed to include only surface data for the sites. Monthly global ocean colour data (4$km^2$) derived from multiple satellite missions[49] was parsed to calculate an average value for chlorophyll a density per site across two years from November 2017. Finally, a previously described[50] 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 a limited ability to discriminate highly localised observations. Distances between pairs of sites was estimated by drawing a continuous transect 1km offshore parallel to the high water mark using Google Earth Pro (v 7.3.2.5776), and taking the distance along the transect to measure distance between sites.

*Ecological statistics*

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 3. 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 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). Samples were then rarefied to the smallest number of reads found within each gene fragment (see Supplementary Table 2). Technical replicates were then collapsed to produce a dataset containing the mean value of rarefied reads per ASV. Finally, ASVs assigned with high confidence (and no cases of multiple matches of equal quality) to the same species in both the COI and the 18S datasets were combined. The taxonomic assignment method used for the 16S data assigns to genus, so no ASVs from the 16S data were collapsed. In order to explore broad scale patterns and avoid difficulties with taxonomic assignment, the number of ASVs per phyla and number of rarefied reads per phyla were collapsed to produce per site assessments of taxonomic composition. For plots phyla represented by less than 2% of ASV counts were concatenated in an ‘other’ category.

Differences in the mean number of ASVs per coastline were assessed using an analysis of variance (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 Permutational Multivariate Analysis of Variance (PERMANOVA)[51] implemented in R with the function *adonis* from the package *vegan* (v 2.5-6)[52] 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[53]. Significant pairwise differences were assessed using the R function *pairwise.adonis* from the *pairwiseAdonis* package (v.0.3)[54]. To analyse if groups of samples have a difference in intra-group community variation, also known as heterogeneity of multivariate dispersion, the PERMDISP2 procedure[55] was used, implemented in the R 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 R function *metaMDS* from the *vegan* package.

Two approaches were used to explore the influence of the abiotic and human impact data on the observed patterns of beta diversity. First, a distance-based redundancy analysis (dbRDA)[56], regressing site Jaccard dissimilarities against SST, SSS, chlorophyll a and human impact, was performed using the R function *dbrda* from the *vegan* package. The significance of terms was assessed with 10,000 permutations. Subsequently dbRDA models were constructed with every combination of independent variables for models with three, two or a single variable for a total of 15 models (including the full model). The adjusted $R^2$ values for each model were extracted to show the unique information provided by each variable. 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. In addition, we used a generalised additive model to assess the significance of each independent variable in explaining the variation in the two non-metric multidimensional axes from above using a restricted maximum likelihood 2D smoother, implemented in the function *ordisurf* in the R package *vegan*.

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[57]. A recently developed method[57], which corrects spurious inflations of the parameter estimate for Mantel tests, was implemented in R. Across each marker 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 in the *adespatial* package (v.0.3-8).

## Results

*Sequencing*

A total of 66.25 million sequences were produced across the three sequencing runs. 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 ± 109,665 (*s.d.*) (see Supplementary Table 2 for further details). Negative control samples exhibited very low levels of cross-contamination (Supplementary Information 2).

*Alpha diversity and taxonomy*

Across all markers the greatest mean ASV richness was found along the southern coast (Fig. 1). However, a one-way ANOVA showed a significant difference ($F_{2,15}$=7.18, p=0.007) between coastlines only in the 16S data with no difference found in the COI ($F_{2,15}$=3.64, p=0.051) or 18S data ($F_{2,15}$=1.82, p=0.195). A post hoc Tukey test of the 16S data (Supplementary Table 3) showed that the east and west coasts had significantly less ASVs compared to the south coast, but that they were not significantly different to one another in terms of overall richness. The taxonomic composition, measured by ASV richness per phyla, was consistent across sites and markers (Fig. 1c). In contrast, the proportion of reads assigned to each phylum exhibited more variation across sites (Supplementary Fig. 1) in the COI and 18S data, although the 16S data were more even.

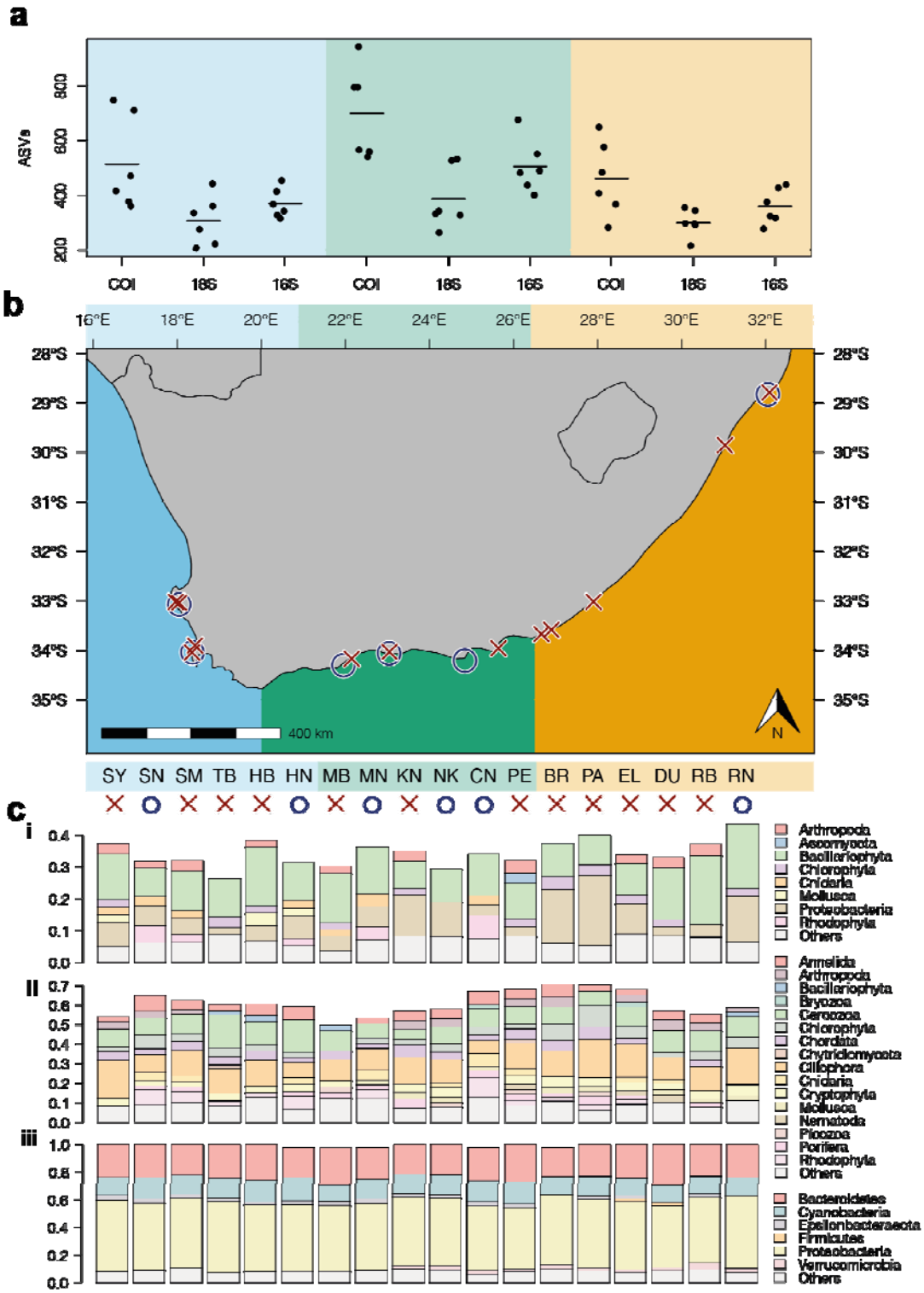


**Figure 1. a** Amplicon sequence variant (ASV) richness per site partitioned by coast (west = blue, south = green, east = orange) from environmental DNA metabarcoding of COI, 18S and 16S markers, black line indicates mean ASV richness; **b** Map of South Africa indicating the sampling sites and the site types (red crosses are artificial sites and blue circles natural sites). Site codes as in Supplementary Table 1; **c** Proportion of ASVs per phyla across each site for **i** COI, **ii** 18S and **iii** 16S. Each bar represents a site indicated by the site code as in Supplementary Table 1.

*Beta diversity*

Across all three markers, the non-metric multidimensional ordinations showed clustering of sites consistent with ecoregions previously described in conspicuous metazoan species (Fig. 2). Furthermore, PERMANOVA modelling showed a significant ($p<0.001$) effect of coastline in all cases (see Supplementary Table 4 for model output), with pairwise significant differences ($p<0.05$) between all pairs of coastlines in all markers (Supplementary Table 5). There was evidence (ANOVA on *betadispr* $F_{2,15}=3.94$, $p=0.042$) for heterogeneity of multivariate dispersion in the 16S dataset (Supplementary Table 6). A Tukey test revealed a significant ($p=0.036$) pairwise difference between the east and west coast only, in line with the observations of Figure 2 with a large observed difference in multivariate spread between these coasts in the 16S data.

The dbRDA showed a significant global effect of the abiotic and human impact variables on the site similarity in COI ($F_{4,16}=2.277$, $p<0.001$), 18S ($F_{4,16}=1.987$, $p<0.001$) and 16S ($F_{4,16}=2.456$, $p<0.001$). Testing individual variables for statistical significance by permutation revealed an ($p<0.05$) effect in all variables across all markers except for human impact in the 16S dataset ($p=0.052$), full model outputs are presented in Supplementary Table 7. Decreasing the number of variables in the model decreased the proportion of explained variance in the ASV dissimilarity as shown in Figure 2. Across all markers, SST explained the most variance in the models (as measured by $R^2$ values in single term models) followed by SSS then chlorophyll a concentration, with the human impact index providing the lowest amount of explanatory information in models. Across all three markers, the model with all variables explained more variation than any model with three variables, indicating that all variables provided unique information.

Generalised additive models with a 2D smoothed function showed significant terms ($p>0.001$) across all models (individual full model outputs shown in Supplementary Table 8), indicating that each variable separately explains variation in the eDNA data between communities. SST, SSS and chlorophyll a concentration showed surfaces across nMDS plots for all markers (Fig. 2) that were simple, with gradients across the plot consistent with ecoregions. In contrast, human impact scores showed more complex surfaces with multiple peaks across the ecoregions.

The corrected Mantel tests indicated that across all markers, SST and the human impact index 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 9). In contrast, SSS and chlorophyll a concentration showed no correlation ($p>0.05$ in all cases) with observed ASV dissimilarities in any marker. These results indicate that both geographic and environmental distance have some effect on the observed community structure, but that out of the tested variables only temperature and the human impact index have a substantial effect after geographic distance between sites has been accounted for.

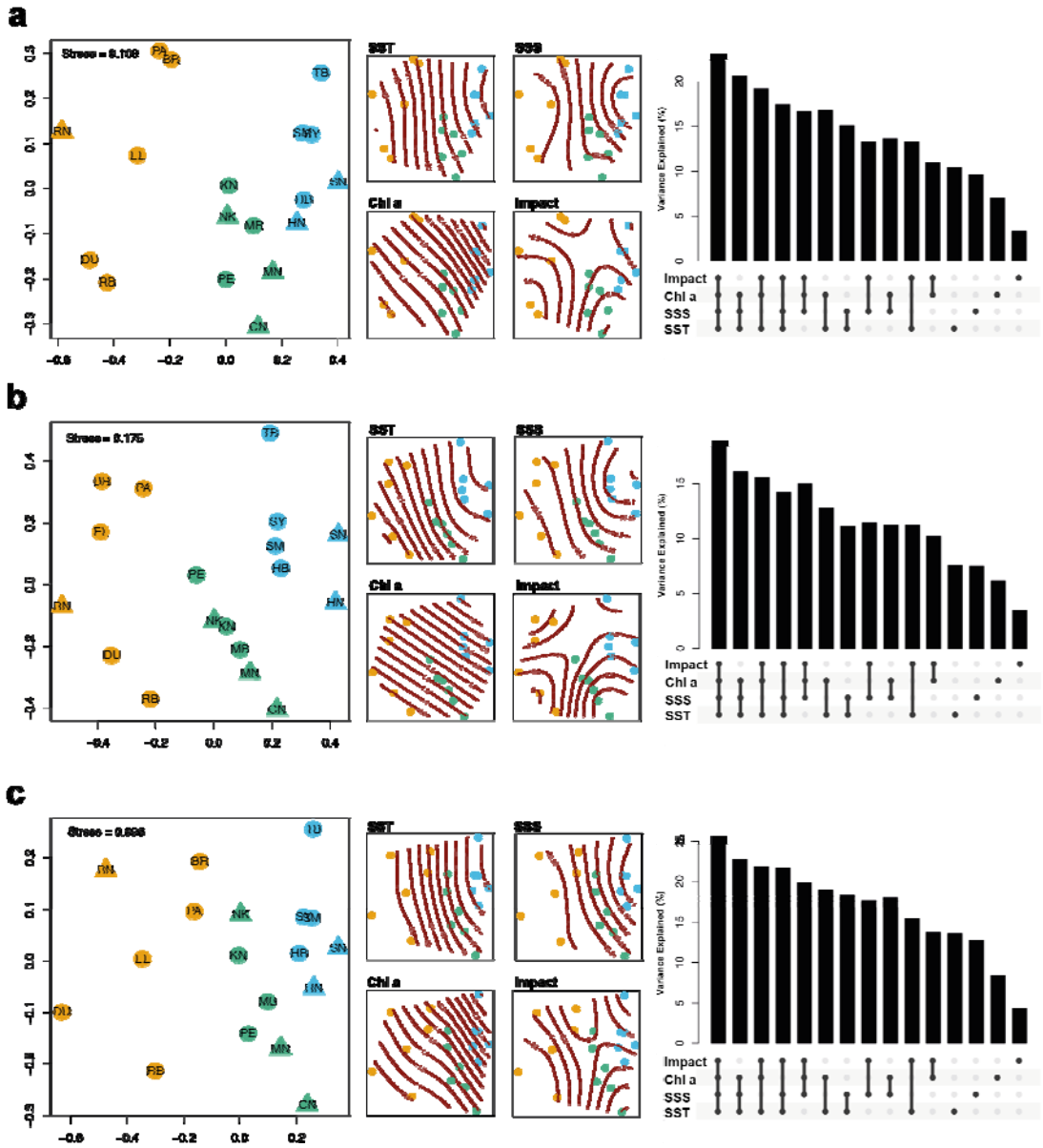


**Figure 2**. Observed patterns of β-diversity from environmental DNA metabarcoding of: **a** COI, **b** 18S and **c** 16S; based on Jaccard dissimilarities between amplicon sequence variants along the coast of South Africa. The first column of plots show non-metric multidimensional scaling (nMDS) ordinations. Coloured ellipses indicate the 95% confidence interval around the centroid and are grouped by coast with the east, south and west coasts denoted by orange, green and blue respectively. Site abbreviations correspond with Figure 1, natural sites are denoted with triangles and artificial sites with filled circles. The second column of plots show the same nMDS ordination as the first column including the output of a generalised additive model with a 2D smoothed function for each of the environmental / impact variables overlaid; SST – mean sea surface temperature (°C); SSS – mean sea surface salinity (parts per thousand); Chl a – chlorophyll a concentration (mg $m^{-3}$); impact – human marine impact score (unitless measurement, see details in text) against the two nMDS axes. The last column of plots depicts upset plots showing the variance (%) in Jaccard dissimilarities among sites, explained by different terms using a distance-based redundancy analysis. Each bar represents a model with the terms in the model indicated by filled dots with the most complex model on the left and the least on the right.

*Composition-distance-decay relationship*

Distance-decay slopes for all observations showed an exponential decrease in compositional similarity as the distance between sites increased (Fig. 3). Regression models of log10 transformed compositional similarity indicated that this slope was statistically significant in all cases ($P<0.005$ for all markers, full model output in Supplementary Table 10). In the COI data the model showed a significant difference in both the slope and intercept between artificial and natural sites ($F_{3,77}=77.92$, additive $p=0.018$, interactive $p=0.004$, Supplementary Table 10 for full model). No statistically significant difference was found between site types in the 18S or 16S data. As shown in Figure 3b-d, there is much greater variance in the residuals for the 18S and 16S data compared to the COI data and the $R^2$ values from the models support these observations (COI $R^2=0.730$, 18S $R^2=0.224$, 16S $R^2=0.473$).

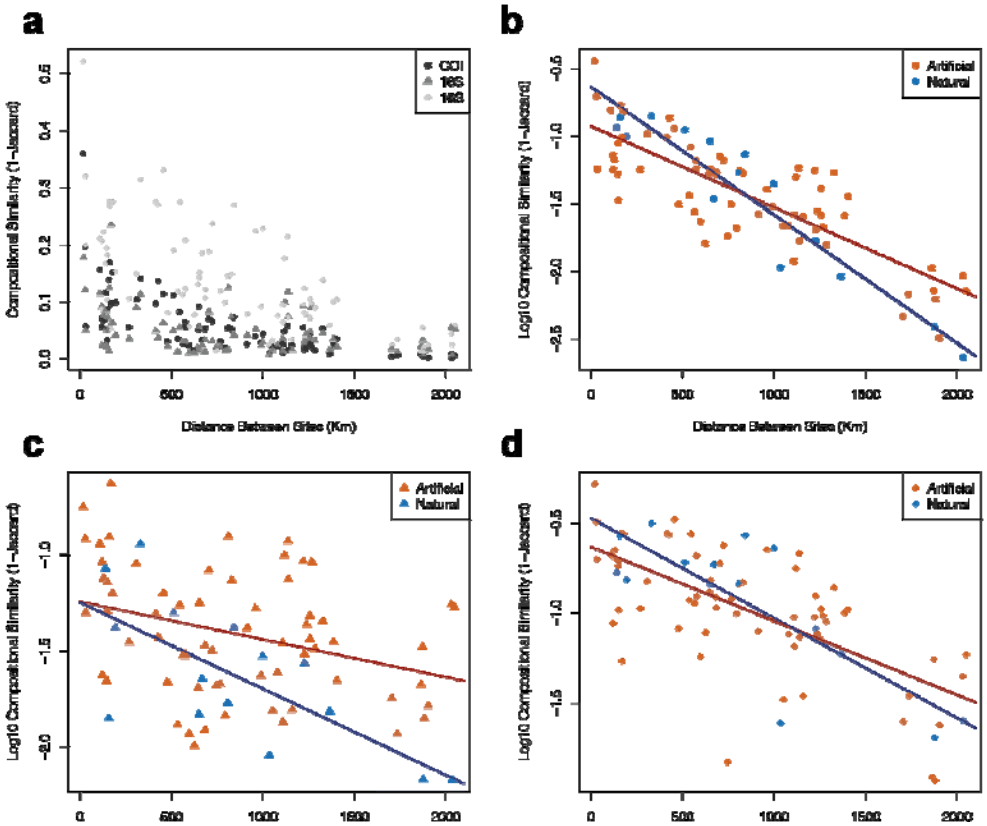


**Figure 3**. Plots showing distance between sites and community dissimilarity measured by environmental DNA metabarcoding across South Africa. a) Untransformed data for all gene regions. Logarithmically (base 10) transformed compositional similarity against distance is shown for b) COI, c) 18S and d) 16S data with comparisons between artificial sites coloured red and natural sites coloured blue. 95% confidence intervals from the least-squares linear regression models are shown as light shaded areas around each regression slope.

*Phylum level analyses*

Removing phyla with fewer than 20 ASVs resulted in 9, 11 and 6 Phyla remaining for the COI, 18S and 16S data respectively with a mean of 152 ASVs per phyla. All but two phyla showed a significant difference between coastlines in PERMANOVA models, with *Chordata* and *Cryptophyta* in the 18S data having no global statistical difference between coasts. In contrast only three phyla showed a difference in the intercept or slope between artificial and natural sites. Regression models for the *Bacillariophyta* from the COI data and *Epsilonbacteraeota* from the 16S data showed a significant ($p < 0.05$) difference in intercept and the *Proteobacteria* from the 16S data showed a difference in both the intercept and slope. Full model outputs for both PERMANOVA and regression models are shown in Supplementary Table 11.

## Discussion

Here, we showed that ecoregions are consistently delimited across the tree of life, with both eukaryotes and prokaryotes showing similar biogeographic patterns along an extensive and heterogenous coastline. On one hand, these observations were in line with our initial expectations that abiotic factors would lead to the maintenance of the heterogeneity of large-scale biodiversity patterns, and our findings clearly showed that seawater temperature was a particularly important factor. On the other hand, our expectation of a more homogeneous distribution along the studied coastline for microbial species was not met; suggesting that abiotic forces are more important drivers than dispersal ability and other life-history characteristics of niche occupancy. Finally, we expected a disrupting effect of anthropogenic impact on species distributions, and although we found that it played a role in shaping the geographic arrangement of some taxa reflected in the COI dataset, our results suggested limited effect of biotic homogenisation (see comparisons between natural and artificial sites in Fig. 3). These findings contradicted our expectations but highlighted the inherent difficulties of predicting the effects of human impact. In addition, they showed that although anthropogenic impacts in ecosystems can be pervasive, they may not be the central force structuring biogeographic patterns along extensive coastlines.

Prokaryotes and eukaryotes diverged billions of years ago and have since evolved to inhabit a vast range of ecological niches. Previous work has shown both similar[58,59] and dissimilar[60] patterns of β diversity between macro- and microscopic species across environmental and geographic gradients. This evidence suggests that different ecological processes drive a number of taxon-specific responses to produce patterns that are not universal across systems at different spatial scales[13]. Indeed evidence from metazoan species demonstrates that biogeography can differ among different classes[6], which is a relatively shallow taxonomic level. Therefore, lack of any universal biogeographic structure across taxa should be considered the null hypothesis for the study of any system, particularly in the ocean where there are comparatively fewer barriers to dispersal and migration. In our study, we observed remarkably similar biogeographic patterns across kingdoms (Fig. 2), rejecting the above null hypothesis, and providing an example of cross-phyla biogeographic congruence.

Our analyses suggested that temperature was the primary environmental variable structuring marine communities across the study region (Fig. 2). In line with this, global studies of biogeographic patterns have shown a central role of temperature in the structuring of both microbial[12] and larger planktonic life[28,61] across the ocean. There is growing evidence that the range boundaries of marine organisms closely track their thermal limits[4]. Therefore, a general expectation was that regardless of taxonomic kingdom, 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 for the global ocean[12], with exceptions found in microbial[62] and meiofaunal[63] life in regions with unusually strong salinity gradients (e.g. Baltic Sea). The SSS range across our study system was very narrow (35.04 - 35.38ppt) and so the negligible observed effect was expected. Similarly, primary productivity (measured here as chlorophyll a concentration) showed little effect in structuring biogeographic patterns (Fig. 1), in line with previous research showing little or no role of productivity in driving coastal and oceanic scale biodiversity patterns[28,61].

Anthropogenic activities are known to alter both the physico-chemical properties of the marine environment and the trophic and ecological properties of ecosystems[64]. In our study system,

human impact provided some explanatory power to understand the observed community structure, but to a much lesser extent compared to temperature (Fig. 2). The human impact index used here[50] covered a large number of different types of impact (e.g. pollution, shipping intensity) but even this granular approach explained the variation in ASVs observed among sampling sites. Previous work on marine metazoans has shown a strong effect of proximate urbanisation[35] and the ecological drivers produced through anthropogenic activities are well documented[64]. Interestingly, the pervasive and conspicuous urbanisation of the marine environment in the study area showed a much weaker effect on biogeographic patterns than abiotic factors (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 human activities are altering evolutionary trajectories[64], either through extinction, speciation or range expansion. However, centuries of human impacts in our study system have not yet demonstrably altered the main biogeographic patterns across kingdoms of life.

Previous work on biotic homogenisation has shown a dramatic effect on whole communities at both regional[17,65] and global scales[23,24]. Here, we found limited evidence for biotic homogenisation along the South African coastline. More specifically, we found no major differences in the distance-decay relationship between artificial and natural sites. This pattern was consistent for the vast majority of gene regions considered (Fig. 3) and subsets of the genes when individual phyla were analysed (see Supplementary Table 11). However, we did find a significant difference in the distance-decay relationship between site types in the COI dataset as a whole (Fig. 3), with more homogenous community composition between artificial sites. As the majority of COI phyla did not individually provide evidence of biotic homogenisation, the difference in the distance decay relationship observed across the whole COI dataset must be driven by the portion of the community not assigned to taxonomy, making interpretation difficult. The remaining markers and phyla showed no evidence of biotic homogenisation when sites were compared, likely a result of consistent species introductions and extirpations in both natural and artificial sites, despite the higher rate of species introductions in anthropogenic habitats. 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. Additionally, seasonal effects may reveal some variation that might explain biotic homogenisation patterns along highly dynamic coastlines.

Environmental parameters have a clear-cut effect on the community structure across kingdoms of life, but the proportion of the total observed variance explained in the best RDA model was less than half. Therefore, the comparative role of deterministic (environmental filtering, niche processes, etc.) and stochastic processes (ecological drift, random extinction, etc.) in explaining the observed patterns remains uncertain. The classical deterministic theory (Baas Becking hypothesis[66]) of microbial biogeography (often summarised as 'everything is everywhere') 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[67]. In line with previous efforts studying deterministic and stochastic processes across taxonomic kingdoms[68,69], we found that the majority of the observed variation could not be explained for both prokaryotic and eukaryotic species (Fig 2.). Indeed recent biogeographic research in the oceans has provided both theoretical[70] and empirical[28] 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 view of the observed variance between communities, the interactions between members of these communities, 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[71]. By rewinding the ecological clock, the approach provides evidence to evaluate the role of deterministic processes in reconstructions of community composition through time relative to changes in environmental proxies. In conjunction, the analysis of networks from molecular data facilitate the testing of species interaction hypotheses that may further elucidate the role of ecological interactions (e.g. Djurhuus, et al. [14]) in structuring biogeographic patterns. Finally, very high-resolution multi-

spectral remote sensing data (e.g. WorldView-3, <100m) will provide unparalleled insights into the role of environmental forces on the distribution of ecological communities[72]. Taken together, recently developed technological and analytical techniques will further advance our understanding of the mechanisms underpinning the biogeography of ecological communities across kingdoms of life.

**Acknowledgments**

We acknowledge the Environmental Sequencing Facility from the National Oceanography Centre, Southampton for access and sequencing assistance. We thank IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, for access. LH was supported by the Natural Environmental Research Council (grant number NE/L002531/1). The UK Research and Innovation Newton Fund (grant: ES/N013913/1) supported LH's research stay in South Africa. LH acknowledges Shirley Parker-Nance and the Elwandle Node of the South African Environmental Observation Network for assistance and in-country logistics. LH acknowledges the assistance of Molly Czachur and Thomas Grevesse during field surveys, Sophie von der Heyden for lab consumables and both Jamie Hudson and Ivan Haigh for assistance with remote sensing data. All marina owners and operators are acknowledged for field site access.

**Data Availability**

Raw Illumina sequencing data are available from the European Nucleotide Archive under study accession number PRJEB38452. Associated metadata, R scripts and intermediate files are available at the following DOI:10.5281/zenodo.3906304 .

**Code Availability**

All code used in the current study can be found at the following DOI:10.5281/zenodo.3906304 .

## 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 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).

16 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).

17 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).

18 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).

19 Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296-299, doi:10.1126/science.1248484 (2014).
20 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).
21 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).
22 Mouillot, D. *et al.* Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biology* **11**, e1001569, doi:10.1371/journal.pbio.1001569 (2013).
23 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).
24 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).
25 Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nature Reviews Microbiology* **5**, 782-791, doi:10.1038/nrmicro1747 (2007).
26 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).
27 Pedersen, M. W. *et al.* Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20130383, doi:10.1098/rstb.2013.0383 (2015).
28 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).
29 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).
30 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).
31 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).
32 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).
33 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).
34 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).
35 Kelly, R. P. *et al.* Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* **4**, e2444, doi:10.7717/peerj.2444 (2016).

36 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).

37 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).

38 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).

39 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).

40 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**, 10-12, doi:10.14806/ej.17.1.200 (2011).

41 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).

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

43 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).

44 Mahmoudabadi, G. & Phillips, R. A comprehensive and quantitative exploration of thousands of viral genomes. *Elife* **7**, doi:10.7554/eLife.31955 (2018).

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

46 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).

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

48 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).

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

50 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).

51 Anderson, M. J. in *Wiley Stats Ref: Statistics Reference Online* (eds N. Balakrishnan *et al.*) Ch. 1 Permutational multivariate analysis of variance (PERMANOVA), 1-15 (John Wiley & Sons, Ltd, 2014).

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

53 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).

54 Martinez Arbizu, P. PairwiseAdonis: Pairwise multilevel comparison using adonis. *R package, Version 0.3* (2020).

55 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).

56 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).

57 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).

58 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).

59 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).

60 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).

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

62 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).

63 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).

64 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).

65 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).

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

67 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).

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

69 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).

70 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).

71 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).

72 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).