# Biogeographic Patterns in the Indian Ocean and their Drivers

A thesis
submitted towards the partial fulfilment of the
BS-MS dual degree programme
by

Vikram Iyer


Date: 20 May 2024

under the guidance of

Devapriya Chattopadhyay

Department of Earth and Climate Sciences, Indian Institute of Science Education and Research Pune

from May 2023 to May 2024

Indian Institute of Science Education and Research Pune

# Certificate

This is to certify that this dissertation entitled "Biogeographic Patterns in the Indian Ocean and their Drivers" submitted towards the partial fulfillment of the BS-MS degree at the Indian Institute of Science Education and Research, Pune represents original research carried out by Vikram Iyer at the Indian Institute of Science Education and Research Pune, under the supervision of Prof. Devapriya Chattopadhyay during the academic year from May 2023 to May 2024.

20.5.24

19/05/2024.

Supervisor:
Devapriya Chattopadhyay
Professor
IISER Pune

Vikram Iyer
20181065
BS-MS
IISER Pune

# Declaration

I hereby declare that the matter embodied in this report titled "Biogeographic Patterns in the Indian Ocean and their Drivers" is the result of investigations carried out by me at the Indian Institute of Science Education and Research Pune from the period 01-05-2023 to 25-05-2024 under the supervision of Prof. Devapriya Chattopadhyay and the same has not been submitted elsewhere for any other degree.

20.5.24

19/05/2024

Supervisor:
Devapriya Chattopadhyay
Professor
IISER Pune

Vikram Iyer
20181065
BS-MS
IISER Pune

# Acknowledgements

There are several people to whom I shall be forever grateful for making me able enough to create this thesis in its present form. Firstly, the person who has directly influenced both my growth as an academic and this project in particular—Devapriya. I think my ability to plan and lead a scientific project has grown tremendously in this one year, and I owe a large part of that to you. You have pushed me, oftentimes to my limit, through this one year, and I think that has both made me a better colleague and a better academic. This is apart from the several academic skills like scientific presentation and documentation that made significant improvement through this one year. Thank you for this.

Secondly, I would like to thank my parents. Not everyone is privileged enough to be raised within academia, and this privilege has given me a large head start in my scientific temper. Not everyone is privileged enough to have an insider's vantage point of academia growing up, and several of my interpersonal skills in navigating the academic landscape I owe to this.

Lastly, I would like to thank IISER Pune and the people I met through it. This thesis is almost entirely data analysis. As someone who knew not even print("hello world") before my first year here, I would not have been able to do what follows in this thesis without the interdisciplinarity of IISER Pune and the flexibility it allowed me. Adding to this were some faculty members whose mentorship, alongside Devapriya's, has helped me tremendously. For this, I would like to thank Drs Sutirth, Sudha, and Shalini.

My peer group was one of the best I could have asked for. Through Ameya, Dhrubo, Jhelam, Krishna, Mihir, Rishika, Shikhara, Shivam, and Shivangi (and Biology Mimamsa Question Making Sessions!), I learned more biology than what many of the courses I have taken here have taught me; through Abhishek, Arjun, Jhelam, Prathith, and Vatsal, I learned much of the maths, physics, and programming to enable what I have done here. And lastly through Punya, I learned some earth science that was no doubt necessary to be in an earth science lab. But perhaps more importantly, many of these people have become lasting friends of mine and I am indebted to them for ensuring my sanity through this one year. To Achintya, Arun, Dhrubo, Jhelam, Pallav, Punya, and Sudeepta, who showed me that physical proximity or daily communication is not necessary for lasting friendship. To Arun, Prathith, and Shivani, for investing in me as much as I invested in their company. To Arun, Ipsa, and Kunjal, for being friends with me since

my first year here and to Avanthika, Dhairya, Jai, Lokesh, Megha, Prathith, and Sanhita, for giving me company through my loneliest last year here. To Amma, who has perhaps been my best friend throughout this year, and lastly, to Swayam, for teaching me caution.

# Contributions

| Contributor name | Contributor role |
|---|---|
| Vikram Iyer, Devapriya Chattopadhyay | Conceptualisation Ideas |
| Vikram Iyer, Devapriya Chattopadhyay | Methodology |
| Vikram Iyer | Software |
| Vikram Iyer | Validation |
| Vikram Iyer | Formal analysis & investigation |
| Freely Accessible | Resources |
| Vikram Iyer | Data Curation |
| Vikram Iyer | Writing: original draft preparation |
| Vikram Iyer, Devapriya Chattopadhyay | Writing: review and editing |
| Vikram Iyer | Visualization |
| Devapriya Chattopadhyay | Supervision |
| Devapriya Chattopadhyay | Project administration |
| Unfunded | Funding acquisition |

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy[1]

[1]https://journals.biologists.com/jcs/pages/author-contributions

# Abstract

Marine biogeographic provinces based on species composition are taxon and method-dependant and cannot be extrapolated. Dispersal is not the first barrier to a region's species composition in the neritic oceans. Environmental filtering does not universally dictate community composition. Using public species presence and oceanographic data, we create standardised biogeographic provinces for all data-rich faunal groups within the Indian Ocean. We find three provinces for several taxa: the Australian Bight, the Northern Indian Ocean, and the rest of the Indian Ocean. Next, we test if environmental filtering drives community composition within and across provinces. We find that environmental filtering pooly predicts community composition within the Indian Ocean. We also find that the former Tethys basin once forming a contiguous neritic region acts as a biogeographic axis within the Indian Ocean. Our results suggest that species composition in the Indian Ocean is based on past faunal patterns influenced by tectonic activity.

Plate tectonics regulates marine diversity patterns, but quantifications of the timescales involved are lacking. The former Tethys basin is an ideal system to attempt to quantify this. This is because it once formed a shallow marine basin stretching from today's Mediterranean in the West to today's Indo-Australian Archipelago (IAA) in the East and contained three different hotspots at different time periods within it arranged in a NW-SE direction. We create testable models of marine diversity using publicly available occurrence data incorporating hotspot location along with the oceanographic parameters dictating community composition for various present-day neritic faunal groups. We find that the present IAA hotspot regulates diversity patterns today for non-planktonic shallow marine groups, and has masked the effects of the hotspots preceding it. This allows us to place an upper bound on the timescales regulating marine diversity to when the demise of the Arabian hotspot preceding the current IAA hotspot occurred.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction and Motivation

## 1.1 Marine biogeography lacks consensus and consistency

In pioneering work, Wallace [1] partitions the terrestrial world into six regions based on species endemism and faunal assemblage distinctness that he calls "regions" (today usually called biogeographic realms). Each region has several "sub-regions" (today usually called biogeographic provinces) nested in it. To this end, he adopts Sclater's [2] partitioning scheme originally employed for birds and applies them to mammals as a model taxon after minor modification. He claims that this partitioning of the terrestrial world is applicable to other taxa like insects and molluscs with minor modifications explainable when their specific "conditions of existence" [1] (p. 58) are accounted for. Recent studies that partition the world into regions analogous to Wallace's obtain results similar to his. This holds true when non-overlapping, large datasets of distributional data are used across a variety of land taxa groupings such as mammals only [3], mammals and birds [4], mammals, birds, and amphibians [5], and angiosperms [6].

On the other hand for shallow marine regions, foundational texts such as Ekman [7] and Briggs [8] recognise that there is little evidence for such congruous partitionings across taxa. Classification schemes vary based on the choice of taxon. For example, delimiting biogeographic realms and provinces using fish [9] and molluscs [10] give different, non-comparable results. Figure 1 (A–E) compares such provinces in the Indian Ocean obtained by different studies. Moreover, different clustering methods for province-demarcation on the same dataset lead to different results [11].

As Wallace himself acknowleges, any biogegoraphic partitioning including Wallace's own biogeographic realms [1] is ultimately arbitrary [12]. However,

having one common scheme across different taxa like mammals, reptiles, and molluscs provides a common framework and acts as a reference point across disparate studies [4, 12]. In the absence of such a framework for shallow marine fauna, studies [13, 14] extrapolate classification schemes to different taxa without justification and attempts to arrive at a common classification scheme by comparing congruous patterns across literature [15] are arbitrary. At other times, studies demarcate their own provinces and use those for further analyses [16, 17]. Others use an existing province-demarcation scheme as a given [18–23]. Because of multiple, often arbitrary schemes with no clear consensus, results across studies are not comparable.

Costello et al. [24] use a large, global dataset and species occurrence records of taxonomically distantly related organisms (plants and animals) to arrive at a single global, nested, classification scheme for the oceans, thus resolving many such inconsistencies. However, because marine provinces are dissimilar across various taxa, their provinces are not useful for studies on finer taxonomic groups.

## 1.2 Does the environment filter in the neritic oceans?

Dispersal is not the first barrier to community composition in marine regions. Biodiverse marine regions affect the diversity of adjacent regions through dispersal, by acting as sources of biodiversity [30–33]. Although early evidence showed that shallow marine regions were completely unstructured with very high population connectivity, there is mounting evidence that this is not the case [34]. However, even then, dispersal distances for most shallow marine taxa range from 100s of meters to 100s of kilometers [35]. This relatively well-connected nature of shallow marine ecosystems along with the fact that most marine dispersal occurs in the pelagic larval stage [34] results in marine ecosystems being more spatially homogeneous than terrestrial ones.

For benthic organisms, this dispersal occurs in life-history stages which are planktic [34]. Motile organisms may disperse in adult stages as well. On the other hand, taxa which are fully planktonic may disperse throughout their lifespan [34]. Moreover, planktonic organisms inhabiting the water column are not limited to neritic regions where most benthic and motile taxon-diversity lies. Thus deep-ocean dynamics, which are distinct and relatively well-understood [34], regulate their diversity patterns as well. Because the mechanics of shallow and deep-water dynamics as well as their timescales are different, we group taxa into benthic, actively motile, and passively free-

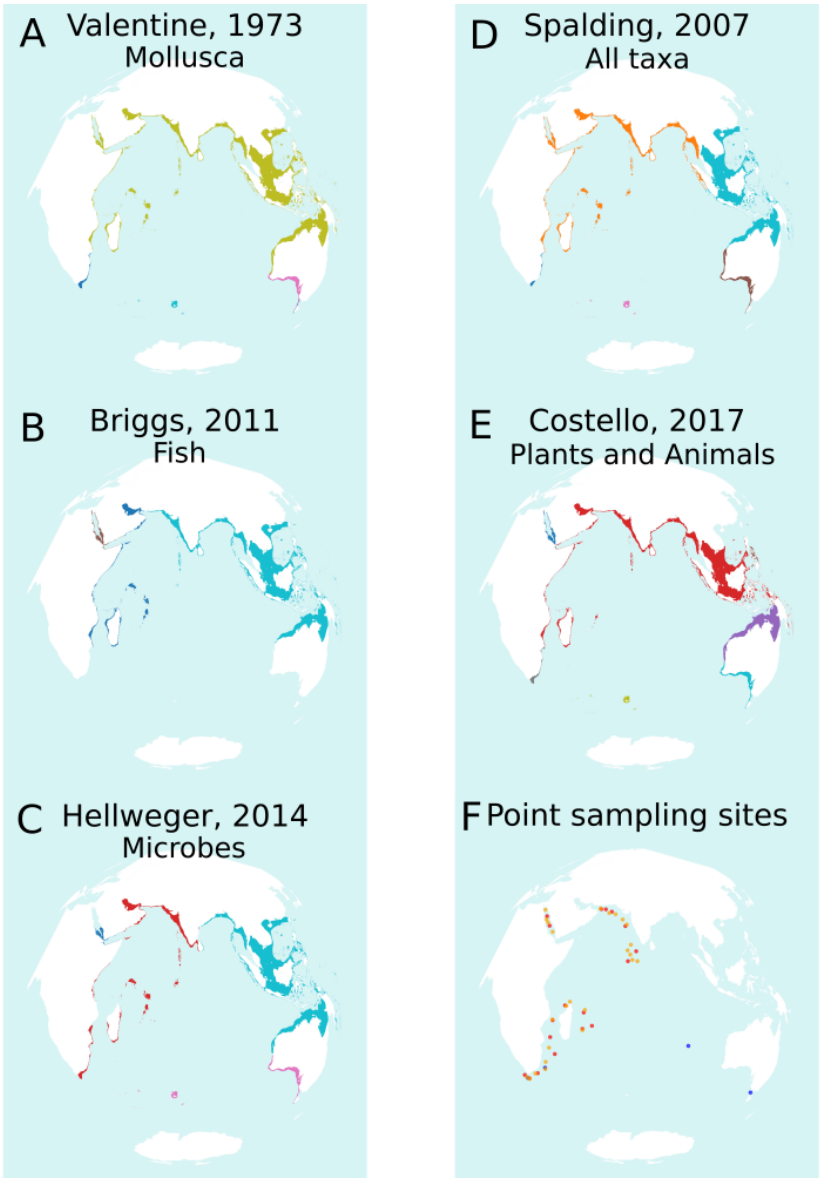



Figure 1.1: Provine demarcations clipped to the boundaries of the Indian Ocean and Indo-Pacific from available literature. Valentine, 1973 = [10]; Spalding, 2007 = [15]; Briggs, 2011 = [9]; Costello, 2017 = [25]; Hellweger, 2014 = [26]; Lower right = Metagenomic study sampling sites from the Indian Ocean including the Indo-Australian Archipelago compiled from available literature. Points are approximated. Red = Sommeria-Klein et al., 2021 [27], Blue = Sánchez et al., 2024 [28], Orange = Ibarbalz et al., 2019. [29]

floating life-modes and check for biogeographic patterns and their drivers in them.

In the absence of dispersal limitations primarily affecting community composition, environmental filtering—the second barrier to community composition on land—might be expected to dictate marine biogeographic patterns. While it has long been recognised that marine biogeographic partitionings are primarily dependant on oceanographic variables such as abiotic parameters and productivity [7, 10], recent studies show that this is not universally true. For neritic benthic and motile macrofauna, oceanographic parameters such as temperature and salinity govern species distribution [7, 14, 36], although this is not true for plankton [27].

## 1.3 The Indian Ocean housed a contiguous, neritic region

Mechanistically, over evolutionary timescales, origination rates, extinction rates, and migration determine the richness of a given region. For shallow marine regions which are a part of the continental shelf, however, plate tectonics is recognised as the primary regulator of marine diversity [37, 38]. Specifically for the Indian Ocean, it was once the Tethys Sea stretching from today's Mediterranean to today's Indo-West Pacific. The Tethys Sea had large and contiguous neritic regions. Because of the high connectivity of neritic regions, we expect the Tethys Sea to have had relatively homogeneous community composition. However, after the closure of the Tethys Sea during the mid-Miocence [39], by the formation of the *Gomphotherium* land bridge connecting the Eurasian and Afro-Arabian landmasses [40], this connectivity was disrupted. Vicariance emerging in populations after this event been shown across a wide range of neritic taxa [41–45]. Because of this, we expect community composition today to vary from one end to the other of this former Tethyian axis from the Mediterranean to the Indo-Pacific.

## 1.4 The timescales regulating marine diversity are poorly quantified

Plate tectonics drives marine diversity patterns [37, 38]. While this operates over several millions of years, there are few [46] attempts to quantify these timescales. Available results correlating marine diversity through time with oceanographic parameters show that both very long-term (200Myr) and

short-term ($\leq$ 50Myr) timescale processes regulate marine diversity patterns [46], but these do not provide finer-scale insight. The former Tethys basin is an ideal system for testing such finer-scale processes. This region stretches from the Mediterranean in the northwest to the Indo-Australian Archipelago (IAA) in the southeast. This is because over the past 50Myr, there have been three successive diversity hotspots within the Tethys basin. The fossil record shows that around 42–39 Mya, the West Tethys, corresponding to today's Mediterranean, was the region of maximum marine diversity. By 23–16 Mya, this region of maximum marine diversity was centred at Arabia with a secondary hotspot in the IAA [47]. This is followed by a hotspot centered at the IAA which continues till today [47, 48]. Since each of these datable hotspots appeared successively south-eastwards, testing the effects of these hotspots' locations on present-day diversity patterns allows us to put an upper bound on the residual effects of these hotspots in regulating marine richness patterns. This in turn puts an upper bound on the timescales regulating marine diversity.

Biodiverse marine regions affect the diversity of adjacent regions through dispersal, by acting as sources of biodiversity [30–33]. For benthic organisms, this occurs in life-history stages which are planktic [34]. Motile organisms may disperse in adult stages as well. On the other hand, taxa which are fully planktonic may disperse throughout their lifespan [34]. Moreover, planktonic organisms inhabiting the water column are not limited to the shallow ocean where most benthic and motile taxon-diversity lies. Thus deep-ocean dynamics, which are distinct and relatively well-understood [34], regulate their diversity patterns as well. Because the mechanics of shallow and deep-water dynamics as well as their timescales are different, we check for the factors driving marine diversity for benthic, actively motile, and passively free-floating organisms separately. For shallow benthic and motile macrofauna, oceanographic parameters such as temperature and salinity govern species distribution [7, 14, 36], although this is not true for plankton [27]. This motivates us to also incorporate oceanographic parameters into our models for marine richness.

# Chapter 2

# Objectives

We propose the following two sets of testable hypotheses:

1. For a region, benthic taxa composition is driven by oceanographic variables such as temperature and salinity.

2. Taxa closely associated with benthic taxa such as neritic fishes are also driven by oceanographic variables.

3. The length of the Tethys Sea from the Mediterranean to the Indo-Pacific is a biogeographic axis for benthic taxa and taxa associated with them.

4. The composition of free-floating neritic taxa is not affected by these drivers.

---

1. Richness patterns for benthic taxa and taxa associated with them such as neritic fishes are predicted by the distance from the current IAA hotspot.

2. Richness patterns for free-floating taxa are unrelated to distance from current or past hotspots.

# Chapter 3

# Methods

## 3.1 Species occurrence data processing

The Ocean Biodiversity Information System (OBIS) (`https://obis.org/`) is a freely accessible subset of the Global Biodiversity Information Facility (GBIF) (`https://www.gbif.org/`) providing presence-only data for marine taxa. We queried and downloaded all data from here with the following keywords to generate three datasets (henceforth called basins):

- Indian Ocean: Scientific name = "Animalia", "Plantae"; Areas = "Indian Ocean", "Laccadive Sea", "Andaman or Burma Sea", "Red Sea", "Persian Gulf", "Gulf of Oman", "Gulf of Aden", "Mozambique Channel", "Great Australian Bight", "Timor Sea", "Arafura Sea", "Bay of Bengal", "Arabian Sea". We classify these seas and oceans as within the Indian Ocean Basin. This region encompasses the margin of the Indian Ocean.

- Western Pacific: Scientific name = "Animalia", "Plantae"; Areas = "Philippine Sea", "Bismarck Sea", "Gulf of Thailand", "Java Sea", "Makassar Strait", "South China Sea", "Celebes Sea", "Banda Sea", "Molukka Sea", "Sulu Sea", "Solomon Sea", "Solomon Islands". This corresponds to all the seas and oceans we classify as the Western Pacific. This region contains the IAA as well as the Coral Triangle completely within it.

- Mediterranean Scientific name = "Animalia", "Plantae"; Areas = "Mediterranean Sea". This corresponds to the Mediterranean basin.

Verbatim query expressions are available in S1. Figure 3.1 displays the boundaries of these regions.

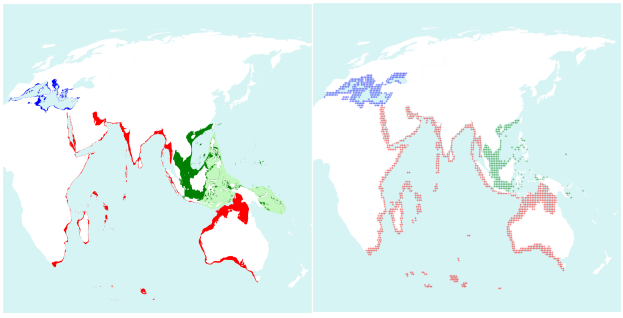


Figure 3.1: **Left:** Regions under consideration: Blue = Mediterranean, Red = Indian Ocean, Green = Western Pacific, light-green = The Coral Triangle as defined by Hoeksema, 2007 [48]. **Right:** Grid-cells considered

We filtered these datasets to retain only the columns corresponding to a recorded species' coordinates, scientific (binomial) name, genus, phylum, and class using custom scripts. We removed records without species-level classification. We also removed all records for birds (Aves) and mammals. This is because there are no completely marine Aves and this would include land-based taxa in our dataset. The only fully marine mammals (Cetaceans) are not not restricted to the shallow ocean and therefore do not follow shallow-marine faunal distribution patterns. We then rounded down each record's latitude and longitude to classify each record into one-by-one degree latitude-longitude grid-cells. These grid-cells have a size of approximately 111 km by 111 km at the equator. Kingdom Plantae was excluded by inspection because of uneven coverage (S2).

Next, we filtered this data to only retain shallow marine regions (<200 m deep). This was done in two steps using custom shapefiles that were created using the General bathymetric Chart of the Oceans (GEBCO) gridded bathymetry data (`https://download.gebco.net/`) to extract depth contours ranging from 0 to 200 meters for our three basins. These were used to create shapefile polygons of the shallow regions. This was done using QGIS ([http://www.qgis.org](http://www.qgis.org/)). Excluding deep observations was done in two steps. First, we removed all grid-cells which are fully in the deep ocean. Next, where a grid-cell has both shallow and deep regions within it, this results in also including observations from deep regions. We eliminated these in our second filtering step by using the precise coordiantes of each observation from OBIS.

Because GBIF data has well-documented biases such as undersampling in the tropics and undersampling of less charismatic taxa [49], we only use presence-absence data for all analyses. We confirm the data-completeness

of OBIS data for our three basins by creating occurrence based rarefaction curves at the species and genus levels (S4). Because of the undersampling of the Western Pacific, we do not combine its data with the data for the Indian Ocean even though it has some connectivity with the Indian Ocean. Figure 3.2 is a flowchart of the processing steps involved. Figure 3.3 visualises the final OBIS datasets generated.

Mode of life groupings were generated by pooling taxa with similar life-modes together:

- Organisms with a benthic non-larval stage: Phylum Porifera, class Anthozoa, class Pycnogonida, class Thecostraca, class Gastropoda, class Bivalvia, phylum Echinodermata, class Asicidacea
- Organisms with a motile non-larval stage: Class Elasmobranchii, class Teleostei, class Cephalopoda
- Organisms with a free-floating non-larval stage: Class Copepoda, class Scyphozoa, phylum Ctenophora

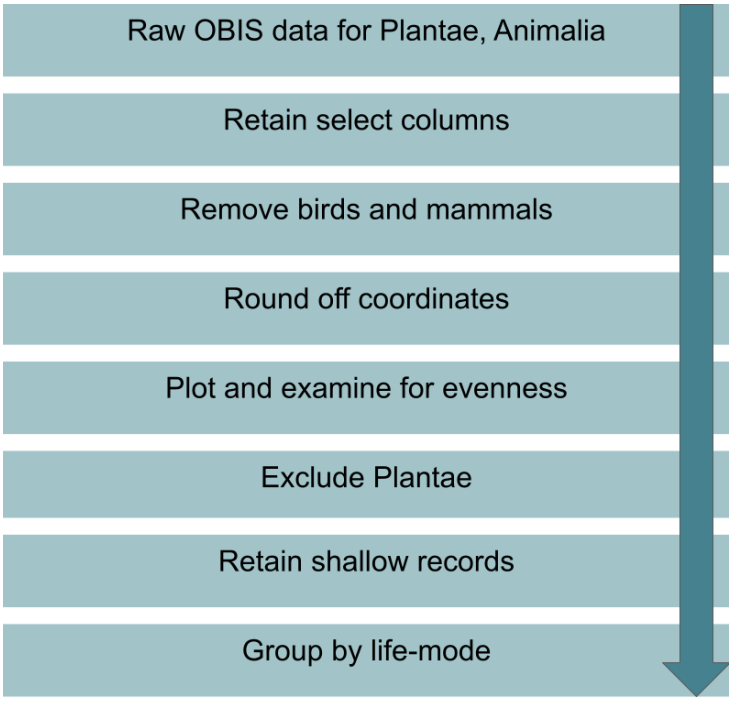



Figure 3.2: A flowchart depicting the processing done with the raw OBIS data.

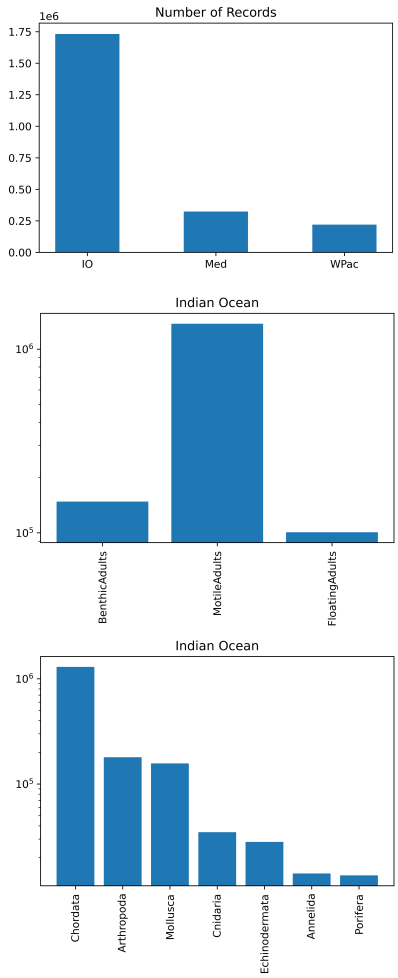



Figure 3.3: **Top:** The number of records from each ocean basin. **Middle:** Records from the Indian Ocean by life-mode. **Bottom:** Records from the Indian Ocean by phylum

## 3.2 Oceanographic data

We downloaded all available ocean-surface data for ocean temperature, salinity, dissolved Oxygen, silicate, phosphate, and nitrate for the year 2018 from NOAA (`https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/`). This data is available at a resolution of one-by-one degree grid-cells. For oceanic productivity, Behrenfeld-CbPM data were downloaded from [50], and subset for the date 01-01-2018. We further subset the productivity data to retain observations nearest to the midpoints of one-by-one degree latitude-longitude grid-cells. All data were subset to only retain records from shallow grid-cells for all basins, and were z-transformed to rescale them (S5). This allows us to compare between values recorded in different units and different precisions.

## 3.3 Calculating distances from the Coral Triangle and the Mediterranean

To calculate the distances of grid-cells from the Coral Triangle (CT), we first computed the centroid of the Coral Triangle (Figure 3.1). We then used the coordinates of this centroid (-0.85856°S, 125.67995°E) as a reference point to calculate distances from. We used the coordinates 31°N, 31°E as a reference point for calculating distances from the Mediterranean. We then used a custom code using GeoPandas [51] in Python and the following crs proj strings for the Mediterranean and the CT respectively:

```
crs="+proj=aeqd+lat_0=31+lon_0=31+x_0=0+y_0=0+datum=WGS84+units=m+no_defs"
```

```
crs="+proj=aeqd+lat_0=-0.85856+lon_0=125.67995+x_0=0+y_0=0+datum=WGS84+units=m+no_defs"
```

These proj strings represent an azimuthal equidistant projection centered at the reference points. We use these because azimuthal equidistant projections preserve globe-distances from their centres to any other point on the map.

## 3.4 Data analyses

### 3.4.1 Province creation and their environmental correlates

We non-hierarchically clustered grid-cells based on genus presences in the Indian Ocean using Principal Coordinate Analyses (PCoA). On obtaining a similar, two or three-pronged pattern for all faunal groupings considered, we used Partition Around Medoids (PAM), an independent clustering algorithm. The number of medoids to cluster around is user-defined, and we set this to two or three depending on the number of prongs visible in each PCoA for each faunal grouping. Because both PCoA and PAM are non-hierarchical clustering methods, this implies that the clusters are split at similar levels of dissimilarity. Next, we plotted these grid-cells on a map of the Indian Ocean and coloured them by the PAM cluster they belong to. Dissimilarity matrices for Principal Coordinate Analyses (PCoA) and Partition Around Medoids (PAM) were made using the packages "fossil" [52] and "vegan" [53] in R. PCoA vectors were generated using the library "ecodist" [54] in R. PAM clustering was done using the library "cluster" [55] in R.

In order to interpret PCoA axes, we correlated them with various oceanographic variables and distances from the Mediterranean and the CT. We calculated Pearsons' correlation coefficients and significances for each faunal group's PCoA axes with the oceanographic variables and distances considered using the Python package "scipy.stats" [56].

### 3.4.2 Comparing grid-cell tuples

For kingdom Animalia, for each cluster obtained by PAM clustering, we compared tuples of grid-cells within each cluster to see how biotically and abiotically different they are. We computed the biotic dissimilarity of two grid-cells as the Jaccard's distance between them. To compute the abiotic dissimilaity between two grid-cells, we calculated the Euclidean distance of the z-transformed oceanographic variables between them using the base library "stats" in R [57]. Grid-cells with missing values of any oceanographic variable were not considered. Mantel tests were run using the library "ecodist" [54] in R.

### 3.4.3 Implementing Generalised Linear Models (GLMs)

We implemented Generalised Linear Models (GLMs) for grid-cell genus richness in R using z-transformed oceanographic variables and the distance from

the Mediterranean and CT as predictor variables for the entire dataset, and for each life-mode. All grid-cells with missing values for any predictor variable were excluded. This does not alter the structure of our data upon inspection (S7). We then adapted Zuur et al.'s [58] protocol for data exploration before running GLMs (S5, S8). We do not exclude any outliers as per Zuur et al. [58].

Next, we check for highly correlated predictor variables using the R package "corrplot" [59] and exclude them (S8). Correlations of explanatory variables show that temperature is strongly negatively correlated with dissolved oxygen and silicate content is strongly positively correlated with DO and phosphorous content. Because temperature directly controls DO, we exclude the latter as a derived variable. Phosphate is limiting to life in the oceans [60], and although marine phosphorous dynamics are complex, ocanic phosphate's primary sink is precipitation in sediments [60]. Phosphate and other ions in the system like silicate influence each other's concentrations by competitive sorption and precipitation [61]. Because of this and phosphate's central role as a limiting nutrient, we choose to exclude silicate as an explanatory variable.

We calculate Akaike's Information Criterion (AIC) values for all single and multiple models considered. Finally, we run a negative binomial single and multiple GLMs for genus richness using the R package MASS [62] for the best model as indicated by AIC values. Results from a Poisson GLM using a log link showed extreme overdispersion and were not used.

# Chapter 4

# Results

## 4.1 Provinces in the Indian Ocean

For kingdom Animalia, the PCoA and PAM clusters correspond to each other (Figure 4.1). Even though there is agreement between the PCoA and PAM clusters, the first two principal components together explain a very small percentage of the total variation in the data (Table 4.1). These PAM clusters correspond to the Australian Bight along with parts of the southwestern coast of Australia and the southern tip of South Africa, the northern Indian Ocean, and the rest of the margin of the Indian Ocean (Figure 4.1). Similar results are obtained on repeating this for various phyla, classes, and life-modes (Figure 4.2 – 4.5). Some PCoA plots show two clusters instead of three. In these cases as well, the PAM clusters largely agree with the PCoA clusters. Porifera does not show any clustering. The Great Australian Bight emerges as a distinct region for most taxa, along with other parts of the Southern Ocean depending of the taxon. For taxa showing three clusters, the Northern Indian Ocean separates as a third cluster along with these two. This holds for groupings based on mode of life as well, except for planktonic species which spend their entire life cycle floating with the water column.

## 4.2 Environmental correlates

For all phyla considered, oceanographic variables are only weakly correlated to PCoA axes (Figure 4.6). This implies that no oceanographic variable alone explains the variation captured by the first two principal components. Echinodermata's first principal component shows the strongest correlations which are significant. These are negatively correlated to oceanographic variables. For various chordate groups and life modes as well, oceanographic

| S. No. | Taxon | PC1 | PC2 |
|---|---|---|---|
| 1 | Animalia | 4.46 | 2.93 |
| 2 | Porifera | 4.76 | 3.94 |
| 3 | Cnidaria | 6.31 | 3.98 |
| 4 | Annelida | 5.34 | 4.02 |
| 5 | Arthropoda | 6.83 | 4.03 |
| 6 | Mollusca | 5.87 | 3.67 |
| 7 | Echinodermata | 4.44 | 3.42 |
| 8 | Chordata | 7.26 | 4.48 |
| 9 | Demospongiae | 4.89 | 4.42 |
| 10 | Anthozoa | 7.11 | 4.97 |
| 11 | Hydrozoa | 8.56 | 7.26 |
| 12 | Polychaeta | 5.42 | 4.10 |
| 13 | Copepoda | 16.67 | 7.05 |
| 14 | Malacostraca | 4.03 | 2.68 |
| 15 | Ostracoda | 27.43 | 11.37 |
| 16 | Bivalvia | 5.33 | 4.17 |
| 17 | Cephalopoda | 18.49 | 10.35 |
| 18 | Gastropoda | 7.30 | 3.88 |
| 19 | Asteroidea | 9.35 | 6.21 |
| 20 | Crinoidea | 10.92 | 7.37 |
| 21 | Echinoidea | 6.81 | 6.26 |
| 22 | Holothuroidea | 14.49 | 6.68 |
| 23 | Ophiuroidea | 8.25 | 7.04 |
| 24 | nonChordata | 3.76 | 2.99 |
| 25 | Ascidiacea | 10.07 | 7.04 |
| 26 | Sagittoidea | 56.45 | 14.19 |
| 27 | Holocephali | 74.38 | 15.11 |
| 28 | Elasmobranchii | 11.06 | 7.61 |
| 29 | Teleostei | 7.50 | 4.48 |
| 30 | BenthicAdults | 3.33 | 2.51 |
| 31 | MotileAdults | 7.32 | 4.55 |
| 32 | FloatingAdults | 16.35 | 6.43 |

Table 4.1: The percentages of the total variation explained by PC1 and PC2 for each group considered.

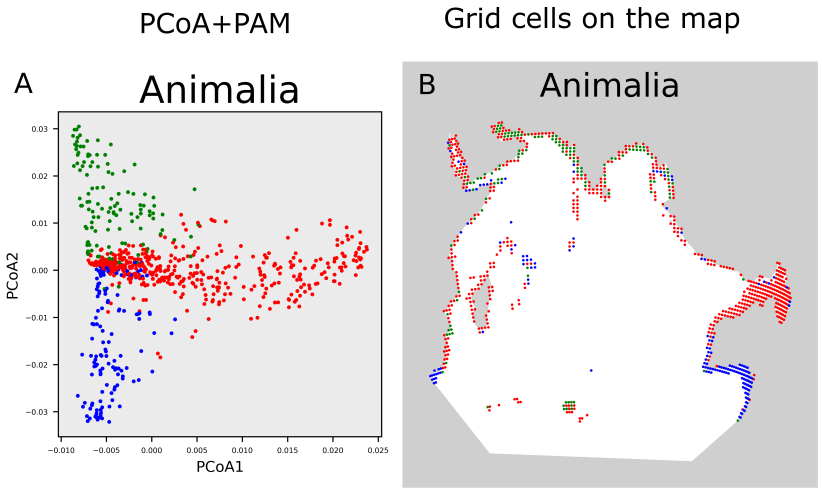



Figure 4.1: **A:** Principal Coordinate Analysis (PCoA) plots for kingdom Animalia at the genus level for the margins of the Indian Ocean. Each point represents one grid-cell. PCoA and PAM analyses were done using genus presences in a grid-cell. The colours represent medoid number after Partition Around Medoids (PAM) clustering (k=3). Points represent grid-cells. **B:** Grid-cells plotted on a map of the margins of the Indian Ocean coloured by medoid number. The colours correspond to the colours on the PCoA plots.

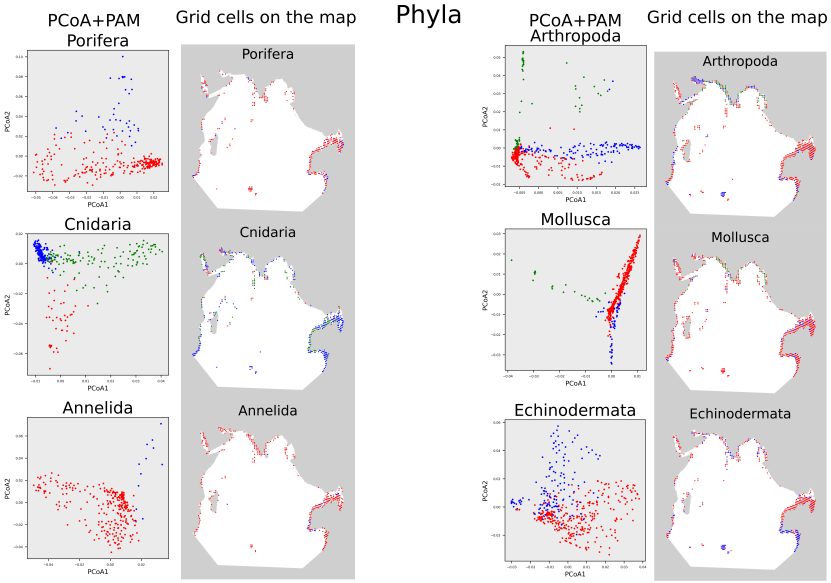



Figure 4.2: Plots for all phyla considered

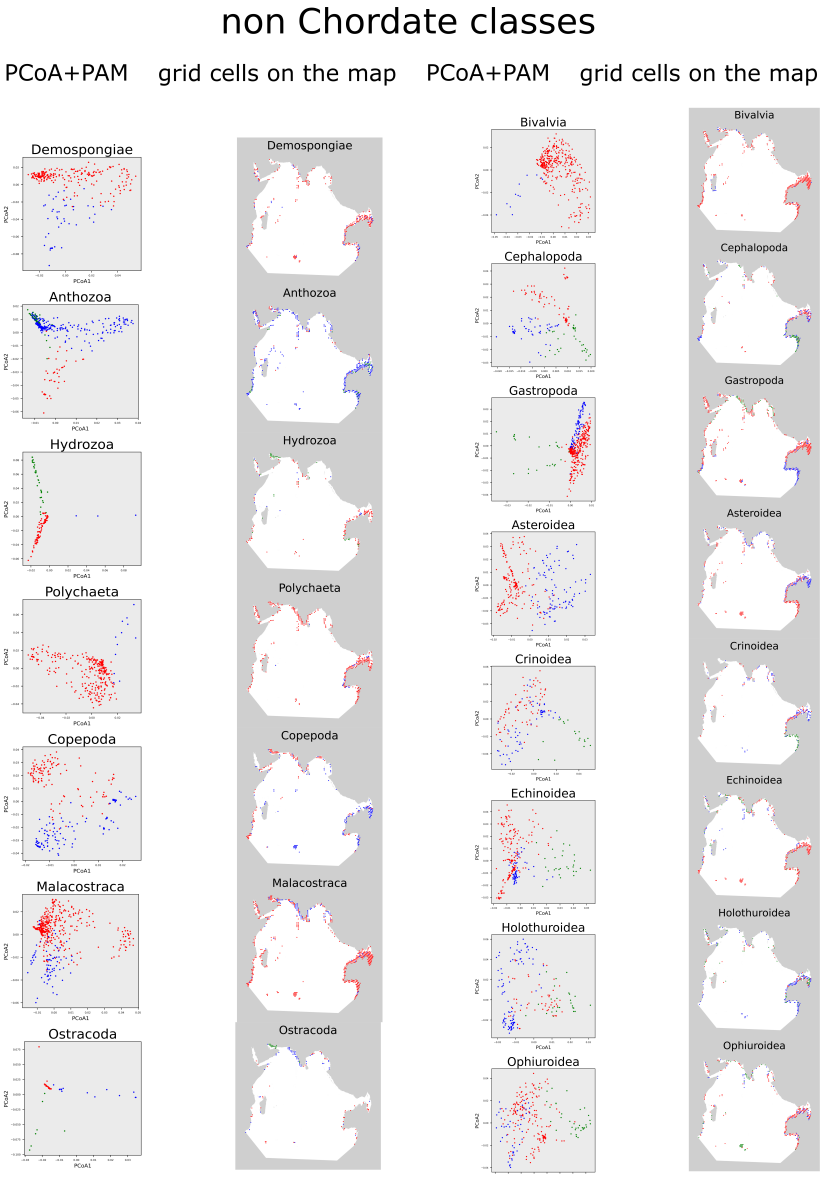



Figure 4.3: Plots for all non-chordate classes considered

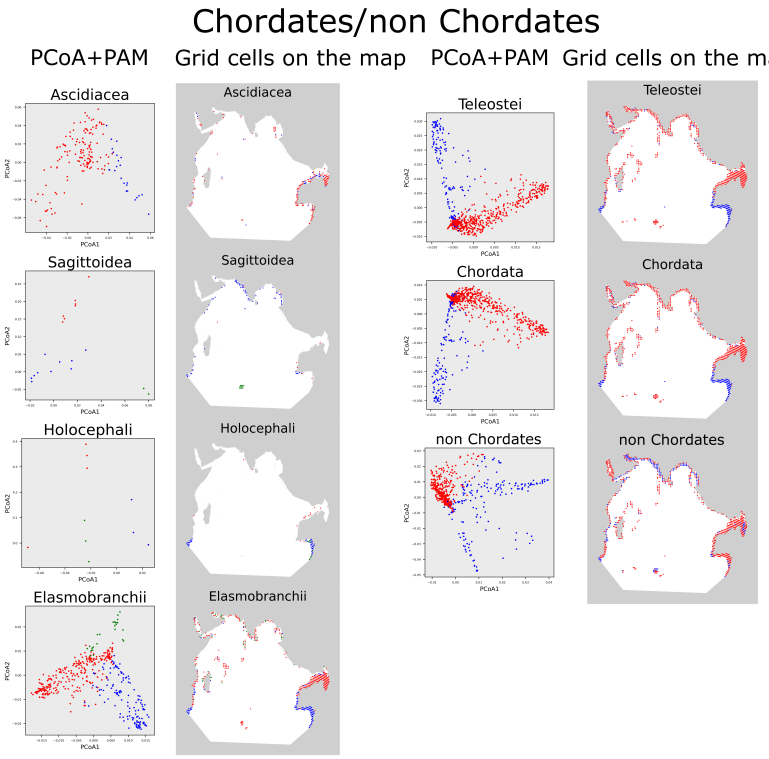



Figure 4.4: Plots for all chordate classes considered, along with all non-chordates together.

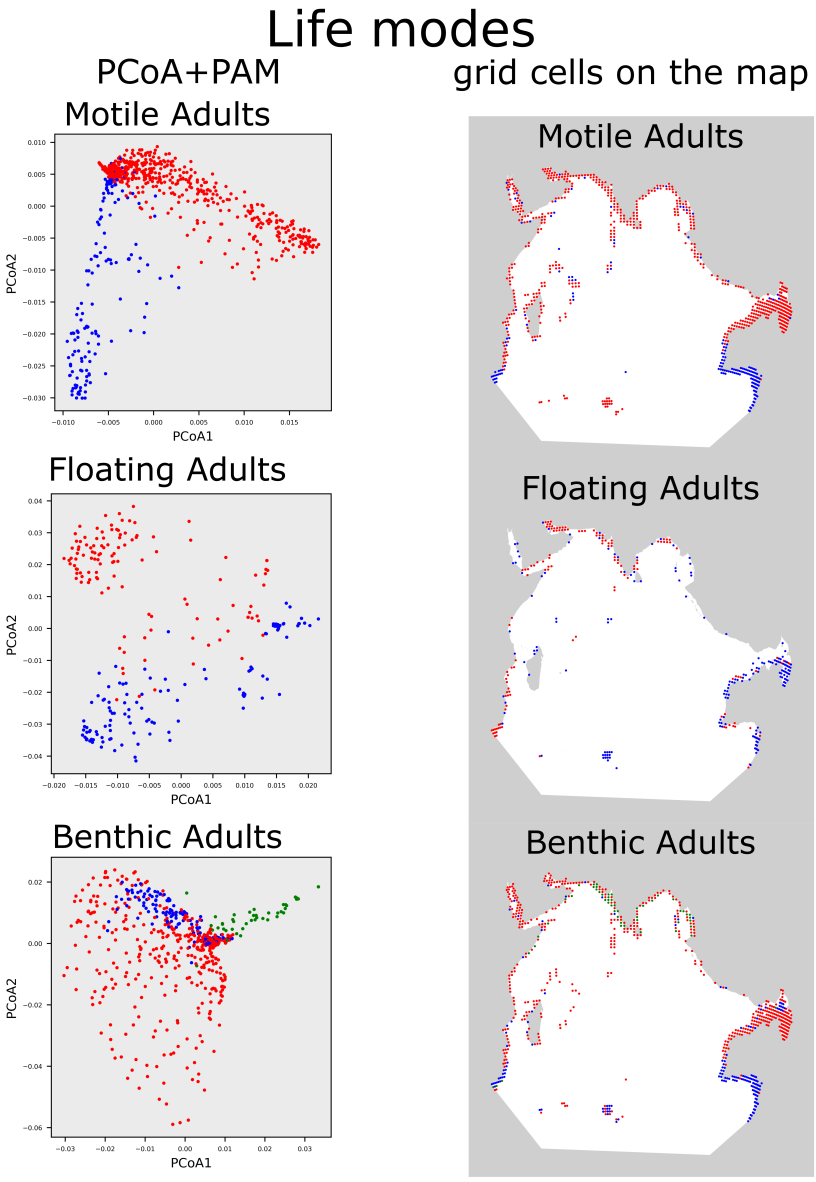



Figure 4.5: Plots for all life-modes considered.

variables are weakly correlated to PCoA axes, except for Sagittoidea. For non-chordate classes, Asteroidea shows patterns similar to Echinodermata, implying that it could account for its patterns. Ophiuroidea's first principal component correlates moderately positively and significantly with oceanographic variables, implying that its patterns are contrary to the patterns of the overall phylum Echinodermata.

Distances from the Mediterranean and Coral Triangle (CT) are weakly correlated to both PCoA axes as well. For most faunal groups, a positive correlation with distance from the Mediterranean has a corresponding negative correlation with distance from the CT. This implies that these two regions which correspond to the ends of the former Tethys basin, are axes of variation for genus similarities among grid-cells in the Indian Ocean. Notably, this pattern does not hold for organisms with a free-floating adult stage.

## 4.3 Environmental dissimilarity poorly predicts compositional dissimilarity

Environmental dissimilarity poorly predicts compositional dissimilarity for all provinces and ocean basins we consider (Figure 4.7). This is evidenced by the fact that most points cluster into the top-left of the figure. Most grid-cell tuples have very few species in common, resulting in biotic compositional dissimilarities close to one.

Medoid 1 (red) has the largest geographic spread and includes north Australia as well as parts of the African coast, the Red Sea, and the Indian coast. Thus, geographically far away points would be compared in tuples. This is reflected in its large values along the x-axis with large standard deviations. Medoid 2 (blue) is largely confined to the Australian Bight. In spite of this, it displays large values of abiotic compositional dissimilarity even exceeding that of medoid 1 at the genus level. On the other hand, the Mediterranean and Western Pacific are both geographically smaller regions and thus show smaller variations in abiotic parameters with smaller standard deviations. Medoid 3 (green) is largely confined to the northern Indian Ocean and displays an intermediate value.

The Western Pacific region contains the Coral Triangle and has high genus richnesses for several taxa. Moreover, dispersal is not the first barrier to community composition in shallow marine regions. However, tuples of grid-cells in it have compositional dissimilarity values close to one, implying very few species in common. This could be attributed to high spatial heterogeneity or incomplete lists of species/genus presences. Hughes et al. [49] show that the

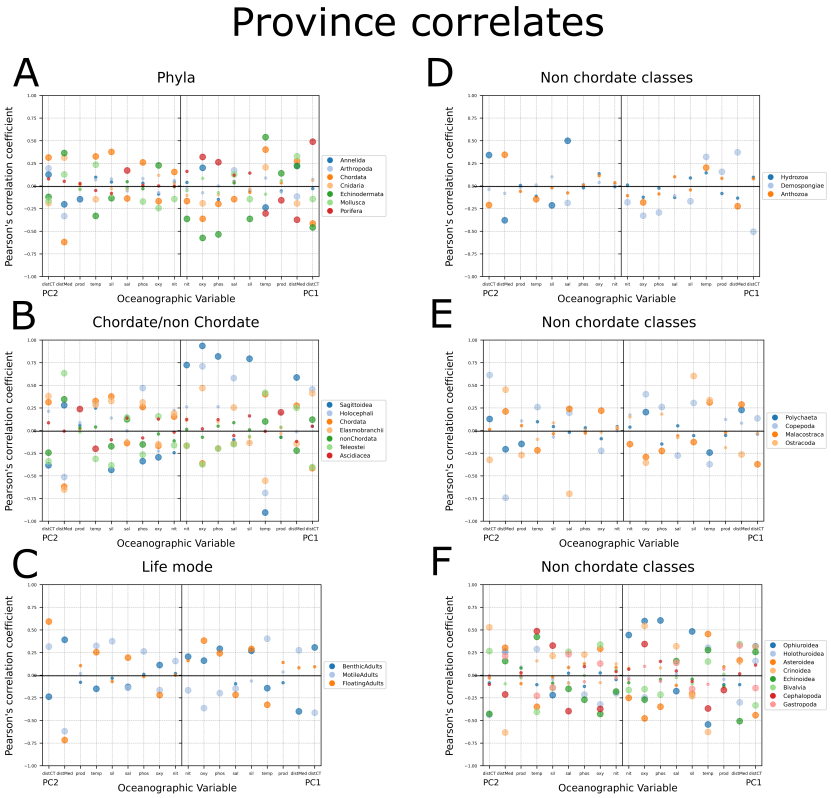



Figure 4.6: Correlations of PCoA axes of each subset of kingdom Animalia with predictor variables including the distance from the CT Small circle = insignificant correlation, large circle = significant correlation. nit = dissolved nitrate, oxy = dissolved oxygen, phosphate = dissolved phosphate, sal = salinity, sil = dissolved silicate, temp = temperature, prod = productivity, distCT = distance from the Coral Triangle, distMed = distance from the Mediterranean.

| Medoid | Mantel's r | p1 (H0 $r \leqslant 0$) | p2 (H0 $r \geqslant 0$) | p3 (H0 $r = 0$) |
|---|---|---|---|---|
| A | -0.068490 | 0.960 | **0.041** | 0.077 |
| B | 0.072355 | **0.050** | 0.951 | 0.100 |
| C | 0.062513 | 0.111 | 0.890 | 0.211 |

Table 4.2: Mantel test results for kingdom Animalia at the genus level. Bold p-values ($< 0.05$) are significant.

Mediterranean is well sampled with respect to the Indian Ocean and Indo-Pacific in OBIS. In spite of using species-presence data and binning our data into grid-cells, both of which reduce data disparity, we still obtain compositional dissimilarities between grid-cells in the Mediterranean comparable to that of the Western Pacific. This may be attributed to data incompleteness in the Western Pacific, as indicated by occurrence-based rarefaction curves (S4).

Medoid 3 including geographically faraway regions like the South African and Indian coasts, has the maximum compositional similarity of all. Because these regions have a high abiotic dissimilarity, especially compared to the Mediterranean and Western Pacific, this evidences a poor match between abiotic and biotic dissimilarities. That is, regardless of whether tuples of grid-cells are close or far apart by abiotic distance, they continue to remain compositionally dissimilar. Moreover, medoids 1–3 separate out into distinct provinces as shown by PCoA plots and PAM clustering (Figure 4.2). In spite of this, tuples of grid-cells within each medoid are abiotically far apart regardless of values of compositional dissimilarity. This is further evidence that compositionally similar regions need not be abiotically similar. We confirm this by running Mantel tests for medoids in the Indian Ocean (Table 4.2). Most correlations are very weak and not significant.

## 4.4 Predictive models with distance from the IAA

The distance from the Coral Triangle (CT) is the best single predictors of a grid-cell's genus richness and outperforms oceanographic variables for all faunal groups except free-floating adults (Table 4.3 – 4.6). AIC weight values indicate that this distance from the CT is by far the most parsimonious amongst all the models tested for all faunal groups except free-floating adults, where the values are comparable (similar orders of magnitude) (Table 4.6). The negative values of model coefficients for distance from the CT show that genus richness decreases on moving away from it (Table 4.7). This is

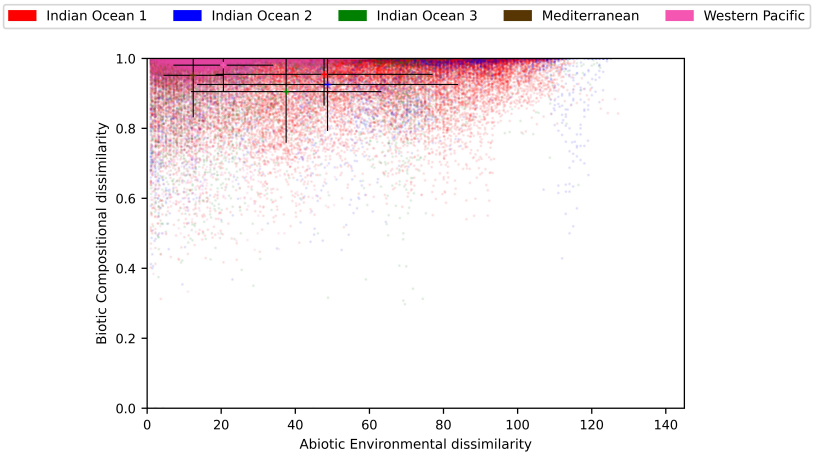



Figure 4.7: Comparing the abiotic and biotic dissimilarities for each grid-cell in the Indian Ocean. Each point = a tuple of two grid-cells belonging to the same medoid/basin. Abiotic environmental dissimilarity between two grid-cells = Euclidean distance between that pair of grid-cells computed using the z-transformed values of all oceanographic variables for the two grid-cells. Biotic compositional dissimilarity = Jaccard's dissimilarity between two grid-cells computed using genus presences in each grid-cell. Colours correspond to the medoids obtained by PAM clustering for each of the four datasets. "+" symbols are centroids and black lines are standard deviations for each axis.

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| scaledDistCT | 5480.305626 | 0.000000 | 9.999750e-01 | scaledDistMed | 5453.761709 | 0.000000 | 1.000000e+00 |
| Phos | 5502.484166 | 22.178539 | 1.530000e-05 | Phos | 5502.484166 | 48.722456 | 2.630000e-11 |
| Nit | 5503.391977 | 23.086351 | 9.700000e-06 | Nit | 5503.391977 | 49.630268 | 1.670000e-11 |
| Sal | 5518.166342 | 37.860716 | 6.010000e-09 | Sal | 5518.166342 | 64.404633 | 1.030000e-14 |
| Temp | 5521.206797 | 40.901170 | 1.310000e-09 | Temp | 5521.206797 | 67.445087 | 2.260000e-15 |
| Productivity | 5521.322196 | 41.016570 | 1.240000e-09 | Productivity | 5521.322196 | 67.560486 | 2.140000e-15 |

Table 4.3: AIC table for single GLMs for predicting grid-cell genus richness for kingdom Animalia.

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| scaledDistCT | 3658.331733 | 0.000000 | 0.603850 | scaledDistMed | 3638.977077 | 0.000000 | 9.999589e-01 |
| Phos | 3660.112990 | 1.781257 | 0.247819 | Phos | 3660.112990 | 21.135913 | 2.570000e-05 |
| Nit | 3661.509882 | 3.178149 | 0.123255 | Nit | 3661.509882 | 22.532805 | 1.280000e-05 |
| Sal | 3665.087383 | 6.755650 | 0.020604 | Sal | 3665.087383 | 26.110306 | 2.140000e-06 |
| Temp | 3669.500318 | 11.168585 | 0.002268 | Temp | 3669.500318 | 30.523241 | 2.350000e-07 |
| Productivity | 3669.558048 | 11.226315 | 0.002204 | Productivity | 3669.558048 | 30.580970 | 2.290000e-07 |

Table 4.4: AIC table for single GLMs for predicting grid-cell genus richness for organisms with a benthic adult stage.

consistent with the fact that the IAA is the most biodiverse region in the Indian Ocean, and correlations with latitude confirm this (Figure 4.8). This implies that the IAA hotspot is currently regulating marine richness patterns for all the life-modes we consider.

## 4.5 Predictive models with distance from the Mediterranean

The distance from the Mediterranean is also the best predictor of genus richness when compared to oceanographic variables for all faunal groups except free-floating adults (Table 4.3 – 4.6). For the latter, although the distance from the CT is the best predictor for genus richness, sea surface temperature is a better model than the distance from the Mediterranean although a ΔAIC value of less than 2 between the latter two shows that these models are not significantly different (Table 4.6). AIC weight values once again indicate that the distance from the Mediterranean is the most parsimonious model for all faunal groups except free-floating adults, where the values are comparable (similar orders of magnitude). The positive coefficients for the distance from the Mediterranean imply that there are no persisting effects

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| scaledDistCT | 4367.490239 | 0.000000 | 9.994755e-01 | scaledDistMed | 4322.106376 | 0.000000 | 1.000000e+00 |
| Phos | 4382.595370 | 15.105131 | 5.244868e-04 | Phos | 4382.595370 | 60.488994 | 7.330000e-14 |
| Nit | 4401.495031 | 34.004793 | 4.130000e-08 | Nit | 4401.495031 | 79.388655 | 5.770000e-18 |
| Temp | 4425.591040 | 58.100802 | 2.420000e-13 | Temp | 4425.591040 | 103.484664 | 3.380000e-23 |
| Productivity | 4428.205466 | 60.715227 | 6.540000e-14 | Productivity | 4428.205466 | 106.099089 | 9.140000e-24 |
| Sal | 4428.399559 | 60.909321 | 5.940000e-14 | Sal | 4428.399559 | 106.293183 | 8.290000e-24 |

Table 4.5: AIC table for single GLMs for predicting grid-cell genus richness for organisms with a motile adult stage.

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| scaledDistCT | 1319.687667 | 0.000000 | 0.738096 | Temp | 1323.184246 | 0.000000 | 0.407171 |
| Temp | 1323.184246 | 3.496579 | 0.128481 | scaledDistMed | 1324.931150 | 1.746905 | 0.169997 |
| Phos | 1325.177572 | 5.489905 | 0.047424 | Phos | 1325.177572 | 1.993326 | 0.150291 |
| Productivity | 1325.179794 | 5.492127 | 0.047371 | Productivity | 1325.179794 | 1.995548 | 0.150124 |
| Sal | 1326.772839 | 7.085172 | 0.021359 | Sal | 1326.772839 | 3.588593 | 0.067690 |
| Nit | 1327.197973 | 7.510307 | 0.017269 | Nit | 1327.197973 | 4.013728 | 0.054728 |

Table 4.6: AIC table for single GLMs for predicting grid-cell genus richness for organisms with a free-floating adult stage.

| Parameter | Estimate | Std. Error | z-value | p-value |
|---|---|---|---|---|
| **Animalia, distCT** | | | | |
| Intercept | 5.41340 | 0.10960 | 49.380 | < 2e-16 |
| scaledDistCT | -0.11540 | 0.01850 | -6.240 | 4.37E-10 |
| **Animalia, distMed** | | | | |
| Intercept | 3.76155 | 0.13176 | 28.548 | <2e-16 |
| scaledDistMed | 0.12642 | 0.01496 | 8.448 | <2e-16 |
| **Benthic adults, distCT** | | | | |
| Intercept | 4.32831 | 0.12380 | 34.963 | < 2e-16 |
| scaledDistCT | -0.07113 | 0.02155 | -3.301 | 0.000963 |
| **Benthic adults, distMed** | | | | |
| Intercept | 3.08343 | 0.16177 | 19.060 | < 2e-16 |
| scaledDistMed | 0.10105 | 0.01762 | 5.737 | 9.65E-09 |
| **Free floating adults, distCT** | | | | |
| Intercept | 2.99724 | 0.13207 | 22.694 | < 2e-16 |
| scaledDistCT | -0.06293 | 0.02178 | -2.889 | 0.00387 |
| **Free floating adults, distMed** | | | | |
| Intercept | 2.46398 | 0.15179 | 16.233 | <2e-16 |
| scaledDistMed | 0.02526 | 0.01705 | 1.481 | 0.139 |
| **Motile adults, distCT** | | | | |
| Intercept | 4.78096 | 0.11194 | 42.710 | < 2e-16 |
| scaledDistCT | -0.14317 | 0.01917 | -7.468 | 8.13E-14 |
| **Motile adults, distMed** | NaN | NaN | NaN | NaN |
| Intercept | 2.57290 | 0.13560 | 18.970 | <2e-16 |
| scaledDistMed | 0.17330 | 0.01520 | 11.400 | <2e-16 |

Table 4.7: Single GLM model coefficients for the optimal models run for all life-modes.

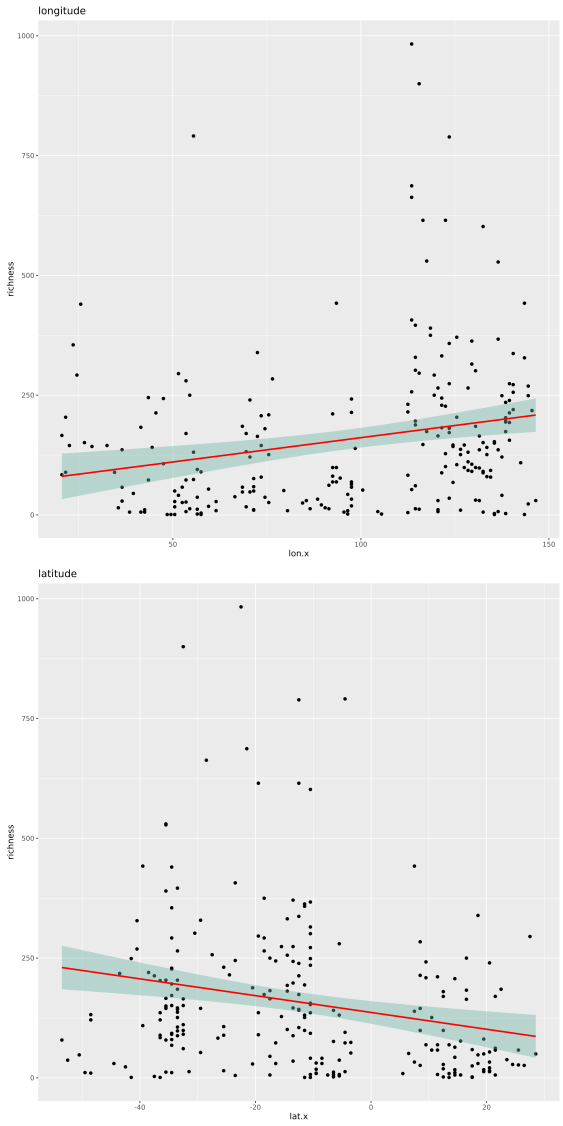



Figure 4.8: Correlations of grid-cell genus richness with latitude and longitude

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| Nit + Sal + Temp + scaledDistCT | 5449.884377 | 0.000000 | 3.225354e-01 | Nit + Sal + Phos + Temp + Productivity + scaledDistMed | 5419.965598 | 0.000000 | 2.455898e-01 |
| Nit + Sal + Temp + Productivity + scaledDistCT | 5450.020575 | 0.136198 | 3.013022e-01 | Nit + Sal + Phos + Temp + scaledDistMed | 5420.240680 | 0.275081 | 2.140312e-01 |
| Nit + Sal + Phos + Temp + scaledDistCT | 5451.094246 | 1.209869 | 1.761399e-01 | Nit + Sal + Temp + scaledDistMed | 5420.459326 | 0.493728 | 1.918663e-01 |
| Nit + Sal + Phos + Temp + Productivity + scaledDistCT | 5451.582626 | 1.698250 | 1.379772e-01 | Nit + Sal + Temp + Productivity + scaledDistMed | 5421.188187 | 1.222589 | 1.332688e-01 |
| Sal + Phos + Temp + scaledDistCT | 5455.841731 | 5.957355 | 1.640417e-02 | Sal + Temp + scaledDistMed | 5423.011743 | 3.046144 | 5.354864e-02 |
| Sal + Phos + Temp + Productivity + scaledDistCT | 5456.522594 | 6.638217 | 1.167097e-02 | Sal + Temp + Productivity + scaledDistMed | 5423.253453 | 3.287854 | 4.745279e-02 |
| Nit + Sal + scaledDistCT | 5456.805677 | 6.921300 | 1.013062e-02 | Nit + Sal + scaledDistMed | 5424.013480 | 4.047882 | 3.245069e-02 |
| Nit + Sal + Productivity + scaledDistCT | 5456.987229 | 7.102852 | 9.251511e-03 | Sal + Phos + Temp + Productivity + scaledDistMed | 5424.936532 | 4.970934 | 2.045435e-02 |
| Nit + Sal + Phos + Productivity + scaledDistCT | 5458.262156 | 8.377780 | 4.890639e-03 | Sal + Phos + Temp + scaledDistMed | 5424.968699 | 5.003101 | 2.012800e-02 |
| Nit + Sal + Phos + scaledDistCT | 5458.419682 | 8.535305 | 4.520219e-03 | Nit + Sal + Productivity + scaledDistMed | 5425.031636 | 5.066037 | 1.950447e-02 |
| Sal + Temp + Productivity + scaledDistCT | 5460.785755 | 10.901378 | 1.384756e-03 | Nit + Sal + Phos + scaledDistMed | 5425.940220 | 5.974622 | 1.238334e-02 |
| Sal + Temp + scaledDistCT | 5461.640063 | 11.755687 | 9.033633e-04 | Nit + Sal + Phos + Productivity + scaledDistMed | 5427.015998 | 7.050400 | 7.231618e-03 |
| Sal + Productivity + scaledDistCT | 5461.748317 | 11.863941 | 8.557668e-04 | Sal + Phos + scaledDistMed | 5430.687098 | 10.721500 | 1.153629e-03 |
| Sal + scaledDistCT | 5462.507095 | 12.622719 | 5.855836e-04 | Sal + Phos + Productivity + scaledDistMed | 5432.003479 | 12.037881 | 5.973346e-04 |
| Sal + Phos + Productivity + scaledDistCT | 5462.672442 | 12.788066 | 5.391185e-04 | Sal + scaledDistMed | 5434.340937 | 14.375338 | 1.856287e-04 |
| Sal + Phos + scaledDistCT | 5462.711674 | 12.827297 | 5.286463e-04 | Sal + Productivity + scaledDistMed | 5434.959089 | 14.993490 | 1.362747e-04 |
| Nit + Phos + Temp + scaledDistCT | 5465.613311 | 15.728934 | 1.239033e-04 | Nit + Temp + scaledDistMed | 5441.685281 | 21.719682 | 4.720000e-06 |
| Nit + Temp + scaledDistCT | 5465.844900 | 15.960523 | 1.103555e-04 | Nit + scaledDistMed | 5442.376996 | 22.411398 | 3.340000e-06 |
| Nit + Temp + Productivity + scaledDistCT | 5467.296360 | 17.411983 | 5.340000e-05 | Nit + Temp + Productivity + scaledDistMed | 5443.416810 | 23.451211 | 1.990000e-06 |
| Nit + Phos + Temp + Productivity + scaledDistCT | 5467.356135 | 17.471759 | 5.180000e-05 | Nit + Phos + Temp + scaledDistMed | 5443.632759 | 23.667160 | 1.780000e-06 |
| Phos + Temp + scaledDistCT | 5469.813606 | 19.929230 | 1.520000e-05 | Nit + Phos + scaledDistMed | 5443.899637 | 23.934039 | 1.560000e-06 |
| Nit + scaledDistCT | 5470.900569 | 21.016193 | 8.810000e-06 | Nit + Productivity + scaledDistMed | 5444.195876 | 24.230278 | 1.340000e-06 |
| Phos + Temp + Productivity + scaledDistCT | 5471.645140 | 21.760764 | 6.070000e-06 | Nit + Phos + Temp + Productivity + scaledDistMed | 5445.306548 | 25.340950 | 7.720000e-07 |
| Nit + Productivity + scaledDistCT | 5472.265627 | 22.381250 | 4.450000e-06 | Nit + Phos + Productivity + scaledDistMed | 5445.785691 | 25.820093 | 6.070000e-07 |
| Nit + Phos + scaledDistCT | 5472.823460 | 22.939083 | 3.370000e-06 | Temp + scaledDistMed | 5447.251319 | 27.285721 | 2.920000e-07 |
| Nit + Phos + Productivity + scaledDistCT | 5474.236490 | 24.352113 | 1.660000e-06 | Phos + Temp + scaledDistMed | 5447.669960 | 27.704362 | 2.370000e-07 |
| Phos + scaledDistCT | 5476.529919 | 26.645542 | 5.280000e-07 | Phos + scaledDistMed | 5448.051811 | 28.086213 | 1.960000e-07 |
| Phos + Productivity + scaledDistCT | 5478.022430 | 28.138054 | 2.500000e-07 | Temp + Productivity + scaledDistMed | 5448.713404 | 28.747806 | 1.410000e-07 |
| Temp + scaledDistCT | 5480.233940 | 30.349563 | 8.280000e-08 | Phos + Temp + Productivity + scaledDistMed | 5449.437205 | 29.471606 | 9.780000e-08 |
| Productivity + scaledDistCT | 5481.138235 | 31.253859 | 5.270000e-08 | Phos + Productivity + scaledDistMed | 5450.005722 | 30.040124 | 7.360000e-08 |
| Temp + Productivity + scaledDistCT | 5481.147210 | 31.262834 | 5.250000e-08 | Productivity + scaledDistMed | 5455.421534 | 35.455936 | 4.910000e-09 |
| Nit + Phos + Temp | 5498.054417 | 48.170040 | 1.120000e-11 | Nit + Phos + Temp | 5498.054417 | 78.088818 | 2.710000e-18 |
| Nit + Sal + Phos + Temp | 5499.012206 | 49.127830 | 6.930000e-12 | Nit + Sal + Phos + Temp | 5499.012206 | 79.046608 | 1.680000e-18 |
| Nit + Phos + Temp + Productivity | 5499.612484 | 49.728107 | 5.130000e-12 | Nit + Phos + Temp + Productivity | 5499.612484 | 79.646885 | 1.240000e-18 |
| Phos + Temp | 5499.946883 | 50.062506 | 4.340000e-12 | Phos + Temp | 5499.946883 | 79.981284 | 1.050000e-18 |
| Nit + Sal + Phos + Temp + Productivity | 5500.660380 | 50.776003 | 3.040000e-12 | Nit + Sal + Phos + Temp + Productivity | 5500.660380 | 80.694782 | 7.370000e-19 |
| Sal + Phos + Temp | 5500.696467 | 50.812090 | 2.980000e-12 | Sal + Phos + Temp | 5500.696467 | 80.730869 | 7.240000e-19 |
| Nit + Phos | 5500.951376 | 51.067000 | 2.630000e-12 | Nit + Phos | 5500.951376 | 80.985778 | 6.370000e-19 |
| Nit + Sal + Phos | 5501.339262 | 51.454885 | 2.160000e-12 | Nit + Sal + Phos | 5501.339262 | 81.373664 | 5.250000e-19 |
| Phos + Temp + Productivity | 5501.447202 | 51.562826 | 2.050000e-12 | Phos + Temp + Productivity | 5501.447202 | 81.481604 | 4.970000e-19 |
| Sal + Phos + Temp + Productivity | 5502.316922 | 52.432545 | 1.330000e-12 | Sal + Phos + Temp + Productivity | 5502.316922 | 82.351323 | 3.220000e-19 |
| Sal + Phos | 5502.680539 | 52.796163 | 1.110000e-12 | Sal + Phos | 5502.680539 | 82.714941 | 2.680000e-19 |
| Nit + Phos + Productivity | 5502.717817 | 52.833441 | 1.090000e-12 | Nit + Phos + Productivity | 5502.717817 | 82.752219 | 2.640000e-19 |
| Nit + Sal + Phos + Productivity | 5503.169188 | 53.284812 | 8.670000e-13 | Nit + Sal + Phos + Productivity | 5503.169188 | 83.203590 | 2.100000e-19 |
| Nit + Sal | 5503.446367 | 53.561991 | 7.550000e-13 | Nit + Sal | 5503.446367 | 83.480769 | 1.830000e-19 |
| Phos + Productivity | 5504.238359 | 54.353982 | 5.080000e-13 | Phos + Productivity | 5504.238359 | 84.272761 | 1.230000e-19 |
| Nit + Temp | 5504.240170 | 54.355794 | 5.070000e-13 | Nit + Temp | 5504.240170 | 84.274572 | 1.230000e-19 |
| Sal + Phos + Productivity | 5504.512596 | 54.628219 | 4.430000e-13 | Sal + Phos + Productivity | 5504.512596 | 84.546998 | 1.070000e-19 |
| Nit + Sal + Temp | 5504.648569 | 54.764193 | 4.140000e-13 | Nit + Sal + Temp | 5504.648569 | 84.682971 | 1.000000e-19 |
| Nit + Productivity | 5505.282404 | 55.398027 | 3.010000e-13 | Nit + Productivity | 5505.282404 | 85.316806 | 7.310000e-20 |
| Nit + Sal + Productivity | 5505.375186 | 55.490810 | 2.880000e-13 | Nit + Sal + Productivity | 5505.375186 | 85.409588 | 6.980000e-20 |
| Nit + Temp + Productivity | 5506.105777 | 56.221401 | 2.000000e-13 | Nit + Temp + Productivity | 5506.105777 | 86.140179 | 4.840000e-20 |
| Nit + Sal + Temp + Productivity | 5506.558496 | 56.674119 | 1.590000e-13 | Nit + Sal + Temp + Productivity | 5506.558496 | 86.592898 | 3.860000e-20 |
| Sal + Temp | 5519.718772 | 69.834396 | 2.210000e-16 | Sal + Temp | 5519.718772 | 99.753174 | 5.360000e-23 |
| Sal + Productivity | 5520.165820 | 70.281443 | 1.770000e-16 | Sal + Productivity | 5520.165820 | 100.200221 | 4.290000e-23 |
| Sal + Temp + Productivity | 5521.718727 | 71.834350 | 8.130000e-17 | Sal + Temp + Productivity | 5521.718727 | 101.753129 | 1.970000e-23 |
| Temp + Productivity | 5523.196244 | 73.311868 | 3.880000e-17 | Temp + Productivity | 5523.196244 | 103.230646 | 9.420000e-24 |

Table 4.8: AIC table for multiple GLMs for predicting grid-cell genus richness for kingdom Animalia.

of either the West Tethyian or Arabian diversity hotspots today and that the IAA hotspot, after its emergence, has masked the influences of these past hotspots. AIC values show that the distance from the Mediterranean is a better predictor of richness than the distance from the CT for all faunal groups except free-floating adults. This is corroborated by the higher magnitudes of the model coefficients for the distance from the Mediterranean over the distance from the CT except for free-floating adults.

## 4.6 Incorporating oceanographic variables improves models' predictive powers

Combining oceanographic parameters along with our two distance parameters improves our models' predictive powers significantly (Table 4.13). Model predictions improve the least for free-floating adults (Table 4.11). For kingdom Animalia, any model including a distance parameter is better than all models without it (Table 4.8). This pattern holds for life-modes except for benthic adults for distance from the CT and free-floating adults for distance from the Mediterranean. Models with all variables included are often significantly worse than models with fewer parameters. Model coefficients' p-values show that the distance parameters are the most significant for all groups except free-floating adults (Table 4.12).

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| Sal + Phos + Temp + scaledDistCT | 3639.449036 | 0.000000 | 2.550799e-01 | Nit + Sal + scaledDistMed | 3621.083912 | 0.000000 | 2.424605e-01 |
| Nit + Sal + Phos + Temp + scaledDistCT | 3639.490122 | 0.041086 | 2.498933e-01 | Nit + Sal + Temp + scaledDistMed | 3622.812970 | 1.729058 | 1.021364e-01 |
| Nit + Sal + Temp + scaledDistCT | 3640.134305 | 0.685269 | 1.810806e-01 | Nit + Sal + Phos + scaledDistMed | 3622.987265 | 1.903353 | 9.361232e-02 |
| Sal + Phos + Temp + Productivity + scaledDistCT | 3641.422316 | 1.973280 | 9.510073e-02 | Nit + Sal + Productivity + scaledDistMed | 3623.066626 | 1.982714 | 8.997050e-02 |
| Nit + Sal + Phos + Temp + Productivity + scaledDistCT | 3641.459563 | 2.010527 | 9.334600e-02 | Sal + Phos + scaledDistMed | 3623.137281 | 2.053369 | 8.684755e-02 |
| Nit + Sal + Temp + Productivity + scaledDistCT | 3642.133610 | 2.684575 | 6.663896e-02 | Sal + Temp + scaledDistMed | 3623.568322 | 2.484410 | 7.000969e-02 |
| Sal + Temp + scaledDistCT | 3645.431633 | 5.982597 | 1.281067e-02 | Sal + scaledDistMed | 3623.890129 | 2.806217 | 5.960444e-02 |
| Nit + Sal + scaledDistCT | 3646.711996 | 7.262960 | 6.753743e-03 | Nit + Sal + Temp + Productivity + scaledDistMed | 3624.786730 | 3.702818 | 3.807011e-02 |
| Phos + Temp + scaledDistCT | 3647.206215 | 7.757179 | 5.275044e-03 | Nit + Sal + Phos + Temp + scaledDistMed | 3624.808820 | 3.724908 | 3.765195e-02 |
| Sal + Temp + Productivity + scaledDistCT | 3647.337933 | 7.888897 | 4.938828e-03 | Nit + Sal + Phos + Productivity + scaledDistMed | 3624.976349 | 3.892437 | 3.462653e-02 |
| Sal + scaledDistCT | 3647.353403 | 7.904367 | 4.900773e-03 | Sal + Phos + Temp + scaledDistMed | 3624.988934 | 3.905022 | 3.440933e-02 |
| Nit + Phos + Temp + scaledDistCT | 3647.440318 | 7.991282 | 4.692359e-03 | Sal + Phos + Productivity + scaledDistMed | 3625.123651 | 4.039739 | 3.216790e-02 |
| Sal + Phos + scaledDistCT | 3648.595998 | 9.146962 | 2.632925e-03 | Sal + Temp + Productivity + scaledDistMed | 3625.448350 | 4.364438 | 2.734735e-02 |
| Nit + Sal + Phos + scaledDistCT | 3648.652804 | 9.203768 | 2.559194e-03 | Sal + Productivity + scaledDistMed | 3625.791489 | 4.707577 | 2.303582e-02 |
| Nit + Sal + Productivity + scaledDistCT | 3648.711331 | 9.262295 | 2.485389e-03 | Nit + Sal + Phos + Temp + Productivity + scaledDistMed | 3626.777809 | 5.693897 | 1.406783e-02 |
| Phos + Temp + Productivity + scaledDistCT | 3648.802001 | 9.352965 | 2.375230e-03 | Sal + Phos + Temp + Productivity + scaledDistMed | 3626.953819 | 5.869907 | 1.288270e-02 |
| Nit + Phos + Temp + Productivity + scaledDistCT | 3649.043366 | 9.594330 | 2.105203e-03 | Nit + scaledDistMed | 3634.723267 | 13.639355 | 2.647845e-04 |
| Sal + Productivity + scaledDistCT | 3649.293304 | 9.844268 | 1.857892e-03 | Phos + scaledDistMed | 3635.985322 | 14.901411 | 1.408772e-04 |
| Nit + Temp + scaledDistCT | 3650.338187 | 10.889151 | 1.101862e-03 | Nit + Phos + scaledDistMed | 3636.058492 | 14.974580 | 1.358164e-04 |
| Sal + Phos + Productivity + scaledDistCT | 3650.586413 | 11.137378 | 9.732521e-04 | Nit + Productivity + scaledDistMed | 3636.626734 | 15.542822 | 1.022257e-04 |
| Nit + Sal + Phos + Productivity + scaledDistCT | 3650.649904 | 11.200868 | 9.428411e-04 | Nit + Temp + scaledDistMed | 3636.719402 | 15.635490 | 9.760000e-05 |
| Nit + Temp + Productivity + scaledDistCT | 3652.199390 | 12.750354 | 4.344822e-04 | Phos + Temp + scaledDistMed | 3637.654189 | 16.570277 | 6.120000e-05 |
| Sal + Phos + Temp | 3652.926213 | 13.477177 | 3.020956e-04 | Nit + Phos + Temp + scaledDistMed | 3637.746105 | 16.662193 | 5.840000e-05 |
| Phos + Temp | 3652.971361 | 13.522325 | 2.953524e-04 | Phos + Productivity + scaledDistMed | 3637.852150 | 16.768239 | 5.540000e-05 |
| Nit + Phos + Temp | 3653.717647 | 14.268611 | 2.033699e-04 | Nit + Phos + Productivity + scaledDistMed | 3637.921193 | 16.837281 | 5.350000e-05 |
| Nit + Sal + Phos + Temp | 3653.777838 | 14.328802 | 1.973405e-04 | Nit + Temp + Productivity + scaledDistMed | 3638.624866 | 17.540954 | 3.760000e-05 |
| Phos + Temp + Productivity | 3653.934803 | 14.485826 | 1.824395e-04 | Phos + Temp + Productivity + scaledDistMed | 3639.446118 | 18.362206 | 2.500000e-05 |
| Sal + Phos + Temp + Productivity | 3654.081599 | 14.632563 | 1.695335e-04 | Nit + Phos + Temp + Productivity + scaledDistMed | 3639.541851 | 18.457939 | 2.380000e-05 |
| Nit + Phos + Temp + Productivity | 3654.662752 | 15.213716 | 1.267826e-04 | Temp + scaledDistMed | 3639.641525 | 18.557613 | 2.260000e-05 |
| Nit + Sal + Phos + Temp + Productivity | 3654.893866 | 15.444830 | 1.129468e-04 | Productivity + scaledDistMed | 3640.954578 | 19.870667 | 1.170000e-05 |
| Nit + scaledDistCT | 3654.939985 | 15.490949 | 1.103721e-04 | Temp + Productivity + scaledDistMed | 3641.631125 | 20.547213 | 8.370000e-06 |
| Phos + scaledDistCT | 3656.163807 | 16.714771 | 5.990000e-05 | Sal + Phos + Temp | 3652.926213 | 31.842301 | 2.950000e-08 |
| Nit + Phos + scaledDistCT | 3656.475914 | 17.026878 | 5.120000e-05 | Phos + Temp | 3652.971361 | 31.887449 | 2.890000e-08 |
| Nit + Productivity + scaledDistCT | 3656.848404 | 17.399368 | 4.250000e-05 | Nit + Phos + Temp | 3653.717647 | 32.633735 | 1.990000e-08 |
| Temp + scaledDistCT | 3657.985613 | 18.536577 | 2.410000e-05 | Nit + Sal + Phos + Temp | 3653.777838 | 32.693926 | 1.930000e-08 |
| Phos + Productivity + scaledDistCT | 3658.052011 | 18.602975 | 2.330000e-05 | Phos + Temp + Productivity | 3653.934862 | 32.850950 | 1.780000e-08 |
| Nit + Phos + Productivity + scaledDistCT | 3658.344025 | 18.894989 | 2.010000e-05 | Sal + Phos + Temp + Productivity | 3654.081599 | 32.997687 | 1.660000e-08 |
| Sal + Phos | 3659.324995 | 19.875959 | 1.230000e-05 | Nit + Phos + Temp + Productivity | 3654.662752 | 33.578840 | 1.240000e-08 |
| Nit + Sal + Temp | 3659.546198 | 20.097162 | 1.100000e-05 | Nit + Sal + Phos + Temp + Productivity | 3654.893866 | 33.809954 | 1.100000e-08 |
| Temp + Productivity + scaledDistCT | 3659.975558 | 20.526522 | 8.900000e-06 | Sal + Phos | 3659.324995 | 38.241083 | 1.200000e-09 |
| Nit + Sal | 3660.204651 | 20.755615 | 7.940000e-06 | Nit + Sal + Temp | 3659.546198 | 38.462286 | 1.080000e-09 |
| Nit + Sal + Phos | 3660.228593 | 20.779557 | 7.840000e-06 | Nit + Sal | 3660.204651 | 39.120739 | 7.760000e-10 |
| Productivity + scaledDistCT | 3660.326638 | 20.877602 | 7.470000e-06 | Nit + Sal + Phos | 3660.228593 | 39.144681 | 7.660000e-10 |
| Nit + Temp | 3660.339161 | 20.890125 | 7.420000e-06 | Nit + Temp | 3660.339161 | 39.255249 | 7.250000e-10 |
| Sal + Phos + Productivity | 3660.903610 | 21.454574 | 5.600000e-06 | Sal + Phos + Productivity | 3660.903610 | 39.819698 | 5.470000e-10 |
| Nit + Phos | 3660.930592 | 21.481557 | 5.520000e-06 | Nit + Phos | 3660.930592 | 39.846681 | 5.400000e-10 |
| Nit + Sal + Temp + Productivity | 3661.036306 | 21.587270 | 5.240000e-06 | Nit + Sal + Temp + Productivity | 3661.036306 | 39.952394 | 5.120000e-10 |
| Phos + Productivity | 3661.577444 | 22.128408 | 4.000000e-06 | Phos + Productivity | 3661.577444 | 40.493532 | 3.900000e-10 |
| Nit + Temp + Productivity | 3661.719113 | 22.270077 | 3.720000e-06 | Nit + Temp + Productivity | 3661.719113 | 40.635201 | 3.640000e-10 |
| Nit + Sal + Phos + Productivity | 3661.742589 | 22.293553 | 3.680000e-06 | Nit + Sal + Phos + Productivity | 3661.742589 | 40.658677 | 3.600000e-10 |
| Nit + Sal + Productivity | 3661.794237 | 22.345201 | 3.580000e-06 | Nit + Sal + Productivity | 3661.794237 | 40.710325 | 3.500000e-10 |
| Nit + Phos + Productivity | 3662.336917 | 22.887881 | 2.730000e-06 | Nit + Phos + Productivity | 3662.336917 | 41.253005 | 2.670000e-10 |
| Nit + Productivity | 3663.011402 | 23.562366 | 1.950000e-06 | Nit + Productivity | 3663.011402 | 41.927490 | 1.910000e-10 |
| Sal + Productivity | 3666.926116 | 27.477080 | 2.750000e-07 | Sal + Productivity | 3666.926116 | 45.842204 | 2.690000e-11 |
| Sal + Temp | 3666.966865 | 27.517829 | 2.700000e-07 | Sal + Temp | 3666.966865 | 45.882953 | 2.640000e-11 |
| Sal + Temp + Productivity | 3668.797854 | 29.348818 | 1.080000e-07 | Sal + Temp + Productivity | 3668.797854 | 47.713942 | 1.060000e-11 |
| Temp + Productivity | 3671.243621 | 31.794585 | 3.180000e-08 | Temp + Productivity | 3671.243621 | 50.159709 | 3.110000e-12 |

Table 4.9: AIC table for multiple GLMs for predicting grid-cell genus richness for organisms with a benthic adult stage.

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| Nit + Sal + Phos + Temp + scaledDistCT | 4335.101608 | 0.000000 | 2.767549e-01 | Nit + Sal + Temp + scaledDistMed | 4274.798594 | 0.000000 | 2.884969e-01 |
| Nit + Sal + Phos + Temp + Productivity + scaledDistCT | 4335.826555 | 0.724946 | 1.926084e-01 | Nit + Sal + Temp + Productivity + scaledDistMed | 4274.967707 | 0.169113 | 2.651055e-01 |
| Sal + Phos + Temp + scaledDistCT | 4336.432180 | 1.330572 | 1.422870e-01 | Nit + Sal + Phos + Temp + scaledDistMed | 4276.698471 | 1.899877 | 1.115805e-01 |
| Sal + Phos + Temp + Productivity + scaledDistCT | 4337.264205 | 2.162597 | 9.380278e-02 | Nit + Sal + Phos + Temp + Productivity + scaledDistMed | 4276.967705 | 2.169111 | 9.752695e-02 |
| Nit + Sal + Phos + scaledDistCT | 4338.799213 | 3.697605 | 4.356830e-02 | Sal + Temp + Productivity + scaledDistMed | 4277.620718 | 2.822124 | 7.035982e-02 |
| Nit + Phos + Temp + scaledDistCT | 4338.842091 | 3.740483 | 4.264418e-02 | Sal + Phos + Temp + scaledDistMed | 4277.819002 | 3.020408 | 6.371883e-02 |
| Nit + Sal + Phos + Productivity + scaledDistCT | 4338.858210 | 3.756601 | 4.230188e-02 | Sal + Phos + Temp + Productivity + scaledDistMed | 4278.241806 | 3.443212 | 5.157713e-02 |
| Sal + Phos + scaledDistCT | 4339.721211 | 4.619603 | 2.747649e-02 | Sal + Temp + scaledDistMed | 4278.256237 | 3.457643 | 5.120631e-02 |
| Sal + Phos + Productivity + scaledDistCT | 4339.862601 | 4.760993 | 2.560111e-02 | Nit + Sal + Phos + scaledDistMed | 4291.356236 | 16.557642 | 7.320000e-05 |
| Phos + Temp + scaledDistCT | 4340.108787 | 5.007179 | 2.263603e-02 | Nit + Temp + scaledDistMed | 4291.398877 | 16.600283 | 7.170000e-05 |
| Nit + Phos + Temp + Productivity + scaledDistCT | 4340.371340 | 5.269732 | 1.985124e-02 | Nit + Phos + Temp + scaledDistMed | 4291.607459 | 16.808865 | 6.460000e-05 |
| Nit + Sal + Productivity + scaledDistCT | 4341.010610 | 5.909002 | 1.442022e-02 | Phos + Temp + scaledDistMed | 4292.521421 | 17.722827 | 4.090000e-05 |
| Nit + Sal + Temp + Productivity + scaledDistCT | 4341.239521 | 6.137913 | 1.286070e-02 | Nit + Temp + Productivity + scaledDistMed | 4292.672424 | 17.873830 | 3.790000e-05 |
| Phos + Temp + Productivity + scaledDistCT | 4341.716757 | 6.615149 | 1.013057e-02 | Nit + Sal + Phos + scaledDistMed | 4292.760213 | 17.961619 | 3.630000e-05 |
| Nit + Sal + Temp + scaledDistCT | 4341.857341 | 6.755733 | 9.442924e-03 | Nit + Sal + Phos + Productivity + scaledDistMed | 4293.224870 | 18.426276 | 2.880000e-05 |
| Nit + Sal + scaledDistCT | 4341.895689 | 6.794081 | 9.263591e-03 | Nit + Phos + Temp + Productivity + scaledDistMed | 4293.238915 | 18.440321 | 2.860000e-05 |
| Nit + Phos + scaledDistCT | 4343.151594 | 8.049986 | 4.943824e-03 | Phos + Temp + Productivity + scaledDistMed | 4294.225770 | 19.427176 | 1.740000e-05 |
| Phos + scaledDistCT | 4343.998709 | 8.897100 | 3.236790e-03 | Sal + Phos + Productivity + scaledDistMed | 4294.703623 | 19.905029 | 1.370000e-05 |
| Nit + Phos + Productivity + scaledDistCT | 4344.253246 | 9.151638 | 2.849983e-03 | Nit + Sal + scaledDistMed | 4296.535546 | 21.736952 | 5.500000e-06 |
| Phos + Productivity + scaledDistCT | 4345.169383 | 10.067775 | 1.802626e-03 | Nit + Sal + Productivity + scaledDistMed | 4297.966897 | 23.168303 | 2.690000e-06 |
| Nit + Temp + scaledDistCT | 4347.935340 | 12.833732 | 4.521534e-04 | Temp + scaledDistMed | 4298.716822 | 23.918228 | 1.850000e-06 |
| Nit + scaledDistCT | 4348.406865 | 13.305257 | 3.571868e-04 | Nit + Phos + scaledDistMed | 4298.884678 | 24.086084 | 1.700000e-06 |
| Nit + Temp + Productivity + scaledDistCT | 4348.541903 | 13.440294 | 3.338662e-04 | Temp + Productivity + scaledDistMed | 4299.420305 | 24.621711 | 1.300000e-06 |
| Nit + Productivity + scaledDistCT | 4348.791590 | 13.689982 | 2.946819e-04 | Phos + scaledDistMed | 4300.060830 | 25.262236 | 9.430000e-07 |
| Sal + Productivity + scaledDistCT | 4355.519641 | 20.418033 | 1.020000e-05 | Nit + Phos + Productivity + scaledDistMed | 4300.884211 | 26.085617 | 6.250000e-07 |
| Sal + Temp + Productivity + scaledDistCT | 4357.519626 | 22.418017 | 3.750000e-06 | Phos + Productivity + scaledDistMed | 4302.052221 | 27.253627 | 3.480000e-07 |
| Sal + scaledDistCT | 4357.739840 | 22.638232 | 3.360000e-06 | Nit + scaledDistMed | 4305.328323 | 30.529729 | 6.770000e-08 |
| Sal + Temp + scaledDistCT | 4359.710795 | 24.609187 | 1.250000e-06 | Nit + Productivity + scaledDistMed | 4307.158620 | 32.360026 | 2.710000e-08 |
| Productivity + scaledDistCT | 4367.024030 | 31.922421 | 3.240000e-08 | Sal + scaledDistMed | 4312.537578 | 37.738984 | 1.840000e-09 |
| Temp + Productivity + scaledDistCT | 4368.837083 | 33.735475 | 1.310000e-08 | Sal + Productivity + scaledDistMed | 4313.816153 | 39.017560 | 9.720000e-10 |
| Temp + scaledDistCT | 4369.152948 | 34.051340 | 1.120000e-08 | Productivity + scaledDistMed | 4323.869601 | 49.071007 | 6.380000e-12 |
| Nit + Phos | 4383.375197 | 48.273589 | 9.110000e-12 | Nit + Phos | 4383.375197 | 108.576603 | 7.640000e-25 |
| Phos + Productivity | 4383.733121 | 48.631512 | 7.620000e-12 | Phos + Productivity | 4383.733121 | 108.934527 | 6.390000e-25 |
| Phos + Temp | 4383.848476 | 48.746867 | 7.190000e-12 | Phos + Temp | 4383.848476 | 109.049882 | 6.030000e-25 |
| Sal + Phos | 4384.425231 | 49.323623 | 5.390000e-12 | Sal + Phos | 4384.425231 | 109.626637 | 4.520000e-25 |
| Nit + Phos + Temp | 4384.524402 | 49.422794 | 5.130000e-12 | Nit + Phos + Temp | 4384.524402 | 109.725808 | 4.300000e-25 |
| Nit + Phos + Productivity | 4384.529388 | 49.427780 | 5.120000e-12 | Nit + Phos + Productivity | 4384.529388 | 109.730794 | 4.290000e-25 |
| Phos + Temp + Productivity | 4384.778782 | 49.677174 | 4.520000e-12 | Phos + Temp + Productivity | 4384.778782 | 109.980188 | 3.790000e-25 |
| Nit + Sal + Phos | 4385.150864 | 50.049256 | 3.750000e-12 | Nit + Sal + Phos | 4385.150864 | 110.352270 | 3.140000e-25 |
| Nit + Phos + Temp + Productivity | 4385.482017 | 50.380409 | 3.180000e-12 | Nit + Phos + Temp + Productivity | 4385.482017 | 110.683424 | 2.660000e-25 |
| Sal + Phos + Temp | 4385.493360 | 50.391752 | 3.160000e-12 | Sal + Phos + Temp | 4385.493360 | 110.694767 | 2.650000e-25 |
| Sal + Phos + Productivity | 4385.527899 | 50.426290 | 3.110000e-12 | Sal + Phos + Productivity | 4385.527899 | 110.729305 | 2.600000e-25 |
| Nit + Sal + Phos + Temp | 4386.068388 | 50.966780 | 2.370000e-12 | Nit + Sal + Phos + Temp | 4386.068388 | 111.269794 | 1.990000e-25 |
| Nit + Sal + Phos + Productivity | 4386.269917 | 51.168308 | 2.140000e-12 | Nit + Sal + Phos + Productivity | 4386.269917 | 111.471323 | 1.800000e-25 |
| Sal + Phos + Temp + Productivity | 4386.327939 | 51.226330 | 2.080000e-12 | Sal + Phos + Temp + Productivity | 4386.327939 | 111.529345 | 1.750000e-25 |
| Nit + Sal + Phos + Temp + Productivity | 4386.931613 | 51.830005 | 1.540000e-12 | Nit + Sal + Phos + Temp + Productivity | 4386.931613 | 112.133020 | 1.290000e-25 |
| Nit + Productivity | 4403.078323 | 67.976715 | 4.800000e-16 | Nit + Productivity | 4403.078323 | 128.279730 | 4.020000e-29 |
| Nit + Temp | 4403.085590 | 67.983982 | 4.780000e-16 | Nit + Temp | 4403.085590 | 128.286996 | 4.010000e-29 |
| Nit + Sal | 4403.401519 | 68.299911 | 4.080000e-16 | Nit + Sal | 4403.401519 | 128.602925 | 3.420000e-29 |
| Nit + Temp + Productivity | 4404.713938 | 69.612330 | 2.120000e-16 | Nit + Temp + Productivity | 4404.713938 | 129.915345 | 1.780000e-29 |
| Nit + Sal + Productivity | 4404.974499 | 69.872891 | 1.860000e-16 | Nit + Sal + Productivity | 4404.974499 | 130.175905 | 1.560000e-29 |
| Nit + Sal + Temp | 4405.066276 | 69.964668 | 1.780000e-16 | Nit + Sal + Temp | 4405.066276 | 130.267682 | 1.490000e-29 |
| Nit + Sal + Temp + Productivity | 4406.686823 | 71.585215 | 7.900000e-17 | Nit + Sal + Temp + Productivity | 4406.686823 | 131.888230 | 6.620000e-30 |
| Sal + Temp | 4426.783146 | 91.681538 | 3.420000e-21 | Sal + Temp | 4426.783146 | 151.984553 | 2.860000e-34 |
| Temp + Productivity | 4427.462664 | 92.361056 | 2.430000e-21 | Temp + Productivity | 4427.462664 | 152.664070 | 2.040000e-34 |
| Sal + Temp + Productivity | 4428.691649 | 93.590040 | 1.320000e-21 | Sal + Temp + Productivity | 4428.691649 | 153.893055 | 1.100000e-34 |
| Sal + Productivity | 4430.140458 | 95.038850 | 6.380000e-22 | Sal + Productivity | 4430.140458 | 155.341864 | 5.350000e-35 |

Table 4.10: AIC table for multiple GLMs for predicting grid-cell genus richness for organisms with a motile adult stage.

| Model names | AIC | ΔAIC | AICWt | Model names | AIC | ΔAIC | AICWt |
|---|---|---|---|---|---|---|---|
| Sal + scaledDistCT | 1317.184359 | 0.000000 | 0.114717 | Sal + Phos + Temp + scaledDistMed | 1319.056439 | 0.000000 | 0.110428 |
| Sal + Productivity + scaledDistCT | 1317.260618 | 0.076258 | 0.110426 | Sal + Temp + Productivity + scaledDistMed | 1319.113242 | 0.056803 | 0.107336 |
| Sal + Temp + Productivity + scaledDistCT | 1318.000135 | 0.815776 | 0.076293 | Sal + Phos + Temp + Productivity + scaledDistMed | 1319.348896 | 0.292457 | 0.095405 |
| Sal + Temp + scaledDistCT | 1318.099153 | 0.914794 | 0.072608 | Sal + Temp + scaledDistMed | 1319.527035 | 0.470596 | 0.087275 |
| Nit + Sal + scaledDistCT | 1319.114478 | 1.930119 | 0.043703 | Temp + scaledDistMed | 1320.349668 | 1.293229 | 0.057844 |
| Sal + Phos + scaledDistCT | 1319.184324 | 1.999965 | 0.042203 | Nit + Sal + Temp + Productivity + scaledDistMed | 1320.392502 | 1.336062 | 0.056618 |
| Sal + Phos + Productivity + scaledDistCT | 1319.217141 | 2.032782 | 0.041516 | Temp + Productivity + scaledDistMed | 1320.455230 | 1.398790 | 0.054870 |
| Nit + Sal + Productivity + scaledDistCT | 1319.223776 | 2.039416 | 0.041379 | Nit + Sal + Temp + scaledDistMed | 1320.703764 | 1.647325 | 0.048458 |
| Sal + Phos + Temp + scaledDistCT | 1319.260414 | 2.076055 | 0.040627 | Nit + Sal + Phos + Temp + scaledDistMed | 1320.985332 | 1.928893 | 0.042094 |
| Sal + Phos + Temp + Productivity + scaledDistCT | 1319.498678 | 2.314318 | 0.036065 | Nit + Sal + Phos + Temp + Productivity + scaled... | 1321.251952 | 2.195513 | 0.036841 |
| Temp + scaledDistCT | 1319.618251 | 2.433892 | 0.033972 | Phos + Temp + scaledDistMed | 1321.605898 | 2.549459 | 0.030865 |
| Nit + Sal + Temp + Productivity + scaledDistCT | 1319.641241 | 2.456881 | 0.033583 | Nit + Temp + scaledDistMed | 1321.779467 | 2.723028 | 0.028300 |
| Nit + Sal + Temp + scaledDistCT | 1319.676555 | 2.492195 | 0.032996 | Nit + Temp + Productivity + scaledDistMed | 1321.975950 | 2.919511 | 0.025652 |
| Temp + Productivity + scaledDistCT | 1319.965580 | 2.781221 | 0.028556 | Phos + Temp + Productivity + scaledDistMed | 1322.089642 | 3.033203 | 0.024234 |
| Productivity + scaledDistCT | 1320.292482 | 3.108122 | 0.024250 | Temp + Productivity | 1322.670930 | 3.614491 | 0.018122 |
| Nit + Sal + Phos + Productivity + scaledDistCT | 1321.084970 | 3.900611 | 0.016316 | Nit + Phos + Temp + scaledDistMed | 1323.421332 | 4.364893 | 0.012452 |
| Nit + Sal + Phos + scaledDistCT | 1321.087789 | 3.903429 | 0.016293 | Sal + Temp + Productivity | 1323.839841 | 4.783402 | 0.010101 |
| Nit + Sal + Phos + Temp + scaledDistCT | 1321.195565 | 4.011205 | 0.015439 | Nit + Phos + Temp + Productivity + scaledDistMed | 1323.859308 | 4.802869 | 0.010003 |
| Phos + scaledDistCT | 1321.294807 | 4.110447 | 0.014691 | Sal + Productivity + scaledDistMed | 1324.282951 | 5.226512 | 0.008094 |
| Nit + Temp + scaledDistCT | 1321.309400 | 4.125040 | 0.014584 | Phos + Productivity | 1324.304078 | 5.247639 | 0.008009 |
| Nit + Sal + Phos + Temp + Productivity + scaled... | 1321.406173 | 4.221813 | 0.013895 | Sal + scaledDistMed | 1324.495328 | 5.438889 | 0.007278 |
| Phos + Temp + scaledDistCT | 1321.451793 | 4.267433 | 0.013582 | Phos + Temp + Productivity | 1324.533915 | 5.477476 | 0.007139 |
| Phos + Productivity + scaledDistCT | 1321.571290 | 4.386930 | 0.012794 | Nit + Temp + Productivity | 1324.621895 | 5.565456 | 0.006832 |
| Nit + scaledDistCT | 1321.683900 | 4.499541 | 0.012094 | Sal + Temp | 1324.699975 | 5.643536 | 0.006571 |
| Nit + Temp + Productivity + scaledDistCT | 1321.710802 | 4.526443 | 0.011932 | Sal + Phos + Productivity + scaledDistMed | 1324.747205 | 5.690766 | 0.006417 |
| Phos + Temp + Productivity + scaledDistCT | 1321.928646 | 4.744286 | 0.010701 | Phos + Productivity + scaledDistMed | 1324.869875 | 5.813436 | 0.006035 |
| Nit + Productivity + scaledDistCT | 1322.292413 | 5.108053 | 0.008921 | Nit + Temp | 1325.120068 | 6.063629 | 0.005326 |
| Temp + Productivity | 1322.670930 | 5.486571 | 0.007383 | Productivity + scaledDistMed | 1325.150960 | 6.094521 | 0.005244 |
| Nit + Phos + scaledDistCT | 1323.078163 | 5.893803 | 0.006023 | Phos + Temp | 1325.154052 | 6.097613 | 0.005236 |
| Nit + Phos + Temp + scaledDistCT | 1323.279892 | 6.095533 | 0.005445 | Phos + scaledDistMed | 1325.247201 | 6.190762 | 0.004998 |
| Nit + Phos + Productivity + scaledDistCT | 1323.297353 | 6.112993 | 0.005398 | Sal + Phos + scaledDistMed | 1325.430401 | 6.373962 | 0.004560 |
| Nit + Phos + Temp + Productivity + scaledDistCT | 1323.708668 | 6.524308 | 0.004394 | Sal + Phos + Productivity | 1325.624331 | 6.567892 | 0.004139 |
| Sal + Temp + Productivity | 1323.839841 | 6.655481 | 0.004115 | Sal + Phos + Temp + Productivity | 1325.777068 | 6.720629 | 0.003835 |
| Phos + Productivity | 1324.304078 | 7.119718 | 0.003263 | Nit + Sal + Temp + Productivity | 1325.786637 | 6.730198 | 0.003816 |
| Phos + Temp + Productivity | 1324.533915 | 7.349556 | 0.002909 | Nit + Phos + Productivity | 1326.056270 | 6.999831 | 0.003335 |
| Nit + Temp + Productivity | 1324.621895 | 7.437535 | 0.002783 | Sal + Productivity | 1326.158150 | 7.101711 | 0.003169 |
| Sal + Temp | 1324.699975 | 7.515616 | 0.002677 | Nit + Sal + Productivity + scaledDistMed | 1326.184352 | 7.127913 | 0.003128 |
| Nit + Temp | 1325.120068 | 7.935709 | 0.002170 | Nit + Phos + Temp + Productivity | 1326.329568 | 7.273129 | 0.002909 |
| Phos + Temp | 1325.154052 | 7.969693 | 0.002133 | Nit + Sal + scaledDistMed | 1326.434353 | 7.377914 | 0.002760 |
| Sal + Phos + Productivity | 1325.624331 | 8.439971 | 0.001686 | Nit + Sal + Phos + Productivity + scaledDistMed | 1326.554629 | 7.498190 | 0.002599 |
| Sal + Phos + Temp + Productivity | 1325.777068 | 8.592709 | 0.001562 | Nit + Phos + Productivity + scaledDistMed | 1326.588483 | 7.532044 | 0.002556 |
| Nit + Sal + Temp + Productivity | 1325.786637 | 8.602278 | 0.001555 | Nit + Sal + Temp | 1326.631015 | 7.574575 | 0.002502 |
| Nit + Phos + Productivity | 1326.056270 | 8.871910 | 0.001359 | Sal + Phos + Temp | 1326.694215 | 7.637776 | 0.002424 |
| Sal + Productivity | 1326.158150 | 8.973791 | 0.001291 | Sal + Phos | 1326.782778 | 7.726339 | 0.002319 |
| Nit + Phos + Temp + Productivity | 1326.329568 | 9.145209 | 0.001185 | Nit + scaledDistMed | 1326.835721 | 7.779281 | 0.002259 |
| Nit + Sal + Temp | 1326.631015 | 9.446655 | 0.001019 | Nit + Productivity | 1326.904169 | 7.847730 | 0.002183 |
| Sal + Phos + Temp | 1326.694215 | 9.509856 | 0.000988 | Nit + Phos | 1327.001425 | 7.944986 | 0.002079 |
| Sal + Phos | 1326.782778 | 9.598419 | 0.000945 | Nit + Phos + Temp | 1327.009224 | 7.952785 | 0.002071 |
| Nit + Productivity | 1326.904169 | 9.719810 | 0.000889 | Nit + Productivity + scaledDistMed | 1327.012491 | 7.956052 | 0.002067 |
| Nit + Phos | 1327.001425 | 9.817065 | 0.000847 | Nit + Phos + scaledDistMed | 1327.033337 | 7.976898 | 0.002046 |
| Nit + Phos + Temp | 1327.009224 | 9.824864 | 0.000844 | Nit + Sal + Phos + scaledDistMed | 1327.288281 | 8.231841 | 0.001801 |
| Nit + Sal + Phos + Productivity | 1327.423691 | 10.239331 | 0.000686 | Nit + Sal + Phos + Productivity | 1327.423691 | 8.367252 | 0.001683 |
| Nit + Sal + Phos + Temp + Productivity | 1327.618078 | 10.433719 | 0.000622 | Nit + Sal + Phos + Temp + Productivity | 1327.618078 | 8.561639 | 0.001527 |
| Nit + Sal + Productivity | 1327.886959 | 10.702600 | 0.000544 | Nit + Sal + Productivity | 1327.886959 | 8.830520 | 0.001335 |
| Nit + Sal | 1328.553413 | 11.369054 | 0.000390 | Nit + Sal | 1328.553413 | 9.496974 | 0.000957 |
| Nit + Sal + Phos + Temp | 1328.579162 | 11.394802 | 0.000385 | Nit + Sal + Phos + Temp | 1328.579162 | 9.522723 | 0.000945 |
| Nit + Sal + Phos | 1328.637135 | 11.452776 | 0.000374 | Nit + Sal + Phos | 1328.637135 | 9.580696 | 0.000918 |

Table 4.11: AIC table for multiple GLMs for predicting grid-cell genus richness for organisms with a free-floating adult stage.

| Parameter | Estimate | Std. Error | z value | p value |
|---|---|---|---|---|
| **Animalia, distCT** | | | | |
| (Intercept) | 5.59933 | 0.15652 | 35.774 | < 2e-16 |
| Nit | -0.44140 | 0.12138 | -3.637 | 2.76E-04 |
| Sal | 0.47936 | 0.09103 | 5.266 | 1.40E-07 |
| Temp | -0.33909 | 0.09955 | -3.406 | 6.58E-04 |
| scaledDistCT | -0.15139 | 0.02047 | -7.395 | 1.42E-13 |
| **Animalia, distMed** | | | | |
| (Intercept) | 2.62287 | 0.24059 | 10.902 | < 2e-16 |
| Nit | -0.35449 | 0.14023 | -2.528 | 0.01147 |
| Sal | 0.51994 | 0.08913 | 5.833 | 5.43E-09 |
| Phos | 0.26238 | 0.13641 | 1.924 | 0.05442 |
| Temp | 0.45955 | 0.13829 | 3.323 | 0.00089 |
| Productivity | 0.06898 | 0.04555 | 1.514 | 0.12999 |
| scaledDistMed | 0.18222 | 0.01837 | 9.918 | < 2e-16 |
| **Benthic adults, distCT** | | | | |
| (Intercept) | 4.59747 | 0.17694 | 25.984 | < 2e-16 |
| Sal | 0.45609 | 0.11860 | 3.846 | 0.00012 |
| Phos | -0.36271 | 0.12741 | -2.847 | 0.004415 |
| Temp | -0.48596 | 0.13632 | -3.565 | 0.000364 |
| scaledDistCT | -0.09666 | 0.02513 | -3.846 | 0.00012 |
| **Benthic adults, distMed** | | | | |
| (Intercept) | 2.60818 | 0.17620 | 14.802 | < 2e-16 |
| Nit | -0.25872 | 0.13012 | -1.988 | 0.0468 |
| Sal | 0.50076 | 0.10840 | 4.620 | 3.84E-06 |
| scaledDistMed | 0.12698 | 0.01813 | 7.004 | 2.48E-12 |
| **Free floating adults, distCT** | | | | |
| (Intercept) | 3.03240 | 0.13190 | 22.991 | <2e-16 |
| Sal | 0.19754 | 0.09110 | 2.168 | 0.0301 |
| scaledDistCT | -0.08284 | 0.02340 | -3.540 | 0.0004 |
| **Free floating adults, distMed** | | | | |
| (Intercept) | 1.70254 | 0.27477 | 6.196 | 5.78E-10 |
| Sal | 0.20502 | 0.09538 | 2.150 | 0.03159 |
| Phos | 0.21720 | 0.13244 | 1.640 | 0.101 |
| Temp | 0.48848 | 0.16109 | 3.032 | 0.00243 |
| scaledDistMed | 0.06472 | 0.02170 | 2.983 | 0.00285 |
| **Motile adults, distCT** | | | | |
| (Intercept) | 4.71232 | 0.16197 | 29.094 | < 2e-16 |
| Nit | -0.28208 | 0.14840 | -1.901 | 0.057322 |
| Sal | 0.31300 | 0.09599 | 3.261 | 0.001112 |
| Phos | -0.47138 | 0.13695 | -3.442 | 0.000577 |
| Temp | -0.31829 | 0.12379 | -2.571 | 0.010135 |
| scaledDistCT | -0.15107 | 0.02168 | -6.969 | 3.19E-12 |
| **Motile adults, distMed** | | | | |
| (Intercept) | 1.36720 | 0.20672 | 6.614 | 3.75E-11 |
| Nit | -0.27796 | 0.12102 | -2.297 | 0.0216 |
| Sal | 0.43939 | 0.08680 | 5.062 | 4.14E-07 |
| Temp | 0.63063 | 0.10887 | 5.792 | 6.94E-09 |
| scaledDistMed | 0.21476 | 0.01676 | 12.816 | < 2e-16 |

Table 4.12: Multiple GLM model coefficients for the optimal models run for all life-modes.

| Dataset | CT/Med | Model | AICsingle | AICmultiple | ΔAIC |
|---|---|---|---|---|---|
| Animalia | CT | Nit + Sal + Temp + scaledDistCT | 5480.305626 | 5449.884377 | 30.421250 |
| Animalia | Med | Nit + Sal + Phos + Temp + Productivity + scaledDistMed | 5453.761709 | 5419.965598 | 33.796111 |
| Benthic Adults | CT | Sal + Phos + Temp + scaledDistCT | 3658.331733 | 3639.449036 | 18.882697 |
| Benthic Adults | Med | Nit + Sal + scaledDistMed | 3638.977077 | 3621.083912 | 17.893165 |
| Motile Adults | CT | Nit + Sal + Phos + Temp + scaledDistCT | 4367.490239 | 4335.101608 | 32.388630 |
| Motile Adults | Med | Nit + Sal + Temp + scaledDistMed | 4322.106376 | 4274.798594 | 47.307782 |
| Floating Adults | CT | Sal + scaledDistCT | 1319.687667 | 1317.184359 | 2.503307 |
| Floating Adults | Med | Sal + Phos + Temp + scaledDistMed | 1324.931150 | 1319.056439 | 5.874711 |

Table 4.13: Comparing between the best single and multiple GLMs run for each life-mode.

# Chapter 5

# Discussion

## 5.1 Provinces in the Indian Ocean

The province-patterns we obtain are similar to published province-demarcations [15, 24, 36] within the Indian Ocean in that the Australian Bight emerges as a distinct province, but are similar to only Spalding's provinces [15] in dividing the northern Indian into a distinct province. Although broadly similar, by using a standardised procedure across taxa and functional groups, we enable our results to be replicable, comparable, and objective.

The first two principal components for all our faunal groups explain a very small proportion of the total variation in the data (Table 4.1). This implies that the Indian Ocean is overall heterogeneous with respect to genus composition. This is unlikely to be because of poor data coverage, because we consider only genus-presences and the Indian Ocean is relatively well sampled (S4). This heterogeneity implies that overall, there are no regions of distinct faunal composition within the Indian Ocean. Because global province schemes place large parts of the neritic Indian Ocean but not the Atlantic or Pacific oceans together [10, 15, 24], this could imply that the Indian Ocean is overall more unstructured and heterogeneous than they are.

## 5.2 The Tethys could regulate diversity patterns in the Indian Ocean

Dispersal is not limiting in the oceans and we find environmental filtering to poorly predict community composition in the Indian Ocean (Figure 4.7). This suggests that other processes such as past faunal patterns from the Tethys Sea could govern community composition in the Indian Ocean today.

The fact that distances from the CT and the Mediterranean are moderately and oppositely correlated to the principal components suggests that relict populations from the former Tethys Sea could dictate community composition in the Indian Ocean.

The IAA has the maximum richnesses for several taxa, and descriptive contours of diversity with a maximum in the IAA have been generated for some taxa such as stomatopods [63] and corals [64]. This combined with the fact that large parts of the margins of the Indian Ocean including the IAA form a single province suggests that the IAA houses a superset of the taxa in the rest of the Indian Ocean's margins. Plate tectonics regulates marine diversity [37, 38], and the break up of the Tethys Sea could be the possible process operating here. This implies that the fauna along the margins of the Indian Ocean derives from Tethyian fauna. The segregation of the North Indian Ocean including the coast of India into a province for some faunal groups suggests that these could derive from fauna on the Indian plate before its collision with the Eurasian plate.

Our results look at diversity patterns in a systematic manner across faunal groups within the Indian Ocean for the first time. We suggest that our standardised provinces be used in future studies for comparability. Although our results suggest that Tethyian compositional patterns govern composition in the Indian Ocean, further studies directly testing this using fossil and genetic evidence are possible.

## 5.3 The timescales regulating marine richness

We formulate testable equivalents for descriptive contours of richness from the IAA for the first time. We find the distance from the CT to the best predictor of the genus richness of a given location (grid-cell) regardless of faunal group. Broadly, the two points taken as references to measure distances from the Mediterranean and the CT are on opposite ends of the former Tethys basin. The marine biodiversity hotspot migrated within the Tethys basin from one end of this to the other, and so these two points can be considered inverses of each other to measure distances from. This is evidenced by negative coefficient values for distance from the CT and positive values for distance from the Mediterranean.

Results for benthic and motile groups are comparable. This is corroborated by the fact that shallow organisms from these groups have similar scales of dispersal [35] and are rarely found in the deep ocean as well. Such shallow-water taxa have been used to study the Latitudinal Diversity Gradient (LDG) and propose the Out of The Tropics (OTT) model of biodiversity

several taxa [65–67] (but also see [68] for a Gastropod counterexample). Although the CT and IAA are tropical regions and are therefore consistent with this, the longitudinal diversity gradient shows that this is not true for parts of the tropical African coast and the North Indian Ocean. This could be because the IAA today has a much larger tropical, shallow-marine area and therefore would dominate in global studies considering the tropics. Larger, contiguous, shallow-tropical regions are more likely to be preserved in part with climactic fluctuations. Since such refugia preserved shallow-marine diversity [22] which can then emigrate elsewhere [33], this could be a possible mechanism by which the IAA has regulated marine diversity patterns after its emergence. Note that we do not address the mechanisms by which the current IAA hotspot emerged from the former West Tethyian and Arabian hotspots in the first place. Hotspot generation is tectonically driven [37, 38]. However, we show that once established, the effects of a hotspot overrule the effects of any former hotspots in regulating diversity.

# Bibliography


1. 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* 586 pp. ISBN: 978-1-4021-1656-8. Google Books: `8lAPAAAAYAAJ` (Macmillan and Company, 1876).
2. Sclater, P. L. On the General Geographical Distribution of the Members of the Class Aves. *Zoological Journal of the Linnean Society* **2,** 130–136. ISSN: 0024-4082. `https://doi.org/10.1111/j.1096-3642.1858.tb02549.x` (2023) (Feb. 1, 1858).
3. Procheş, Ş. & Ramdhani, S. The World's Zoogeographical Regions Confirmed by Cross-Taxon Analyses. *BioScience* **62,** 260–270. ISSN: 0006-3568. `https://doi.org/10.1525/bio.2012.62.3.7` (2023) (Mar. 1, 2012).
4. Holt, B. G. *et al.* An Update of Wallace's Zoogeographic Regions of the World. *Science* **339,** 74–78. `https://www.science.org/doi/full/10.1126/science.1228282` (2023) (Jan. 4, 2013).
5. Rueda, M., Rodríguez, M. Á. & Hawkins, B. A. Identifying Global Zoogeographical Regions: Lessons from Wallace. *Journal of Biogeography* **40,** 2215–2225. ISSN: 1365-2699. `https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.12214` (2023) (2013).
6. Liu, Y. *et al.* An Updated Floristic Map of the World. *Nature Communications* **14,** 2990. ISSN: 2041-1723. `https://www.nature.com/articles/s41467-023-38375-y` (2023) (1 May 30, 2023).
7. Ekman, S. *Zoogeography of the Sea* 444 pp. Google Books: `dma2AAAAIAAJ` (Sidgwick and Jackson, 1953).
8. Briggs, J. C. *Global Biogeography* 473 pp. ISBN: 978-0-08-053254-7. Google Books: `AxDpURo9vnkC` (Elsevier, Oct. 13, 1995).

9. Briggs, J. C. & Bowen, B. W. A Realignment of Marine Biogeographic Provinces with Particular Reference to Fish Distributions. *Journal of Biogeography* **39,** 12–30. ISSN: 1365-2699. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2011.02613.x (2023) (2012).

10. Valentine, J. W. *Evolutionary Paleoecology of the Marine Biosphere* 544 pp. ISBN: 978-0-13-293720-7. Google Books: s1m1AAAAIAAJ (Prentice-Hall, 1973).

11. Mouillot, D. *et al.* The Challenge of Delineating Biogeographical Regions: Nestedness Matters for Indo-Pacific Coral Reef Fishes. *Journal of Biogeography* **40,** 2228–2237. ISSN: 1365-2699. https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.12194 (2023) (2013).

12. Wallace, A. R. What Are Zoological Regions?1. *Nature* **49,** 610–613. ISSN: 1476-4687. https://www.nature.com/articles/049610a0 (2023) (1278 Apr. 1, 1894).

13. Rosa, R. *et al.* Global Patterns of Species Richness in Coastal Cephalopods. *Frontiers in Marine Science* **6.** ISSN: 2296-7745. https://www.frontiersin.org/articles/10.3389/fmars.2019.00469 (2023) (2019).

14. Belanger, C. L. *et al.* Global Environmental Predictors of Benthic Marine Biogeographic Structure. *Proceedings of the National Academy of Sciences* **109,** 14046–14051. https://www.pnas.org/doi/full/10.1073/pnas.1212381109 (2023) (Aug. 28, 2012).

15. Spalding, M. D. *et al.* Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience* **57,** 573–583. ISSN: 1525-3244, 0006-3568. https://academic.oup.com/bioscience/article/57/7/573/238419 (2023) (July 1, 2007).

16. Keith, S. A., Baird, A. H., Hughes, T. P., Madin, J. S. & Connolly, S. R. Faunal Breaks and Species Composition of Indo-Pacific Corals: The Role of Plate Tectonics, Environment and Habitat Distribution. *Proceedings of the Royal Society B: Biological Sciences* **280,** 20130818. https://royalsocietypublishing.org/doi/full/10.1098/rspb.2013.0818 (2023) (July 22, 2013).

17. Kerr, M. R. & Alroy, J. Marine Diversity Patterns in Australia Are Filtered through Biogeography. *Proceedings of the Royal Society B: Biological Sciences* **288,** 20211534. https://royalsocietypublishing.org/doi/full/10.1098/rspb.2021.1534 (2023) (Nov. 10, 2021).

18. Samoilys, M. A., Halford, A. & Osuka, K. Disentangling Drivers of the Abundance of Coral Reef Fishes in the Western Indian Ocean. *Ecology and Evolution* **9,** 4149–4167. ISSN: 2045-7758. https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5044 (2023) (2019).

19. DiBattista, J. D. *et al.* When Biogeographical Provinces Collide: Hybridization of Reef Fishes at the Crossroads of Marine Biogeographical Provinces in the Arabian Sea. *Journal of Biogeography* **42,** 1601–1614. ISSN: 1365-2699. https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.12526 (2023) (2015).

20. Hobbs, J.-P. A., Frisch, A. J., Allen, G. R. & Van Herwerden, L. Marine Hybrid Hotspot at Indo-Pacific Biogeographic Border. *Biology Letters* **5,** 258–261. https://royalsocietypublishing.org/doi/10.1098/rsbl.2008.0561 (2023) (Dec. 23, 2008).

21. Huang, D., Goldberg, E. E., Chou, L. M. & Roy, K. The Origin and Evolution of Coral Species Richness in a Marine Biodiversity Hotspot*. *Evolution* **72,** 288–302. ISSN: 0014-3820. https://doi.org/10.1111/evo.13402 (2023) (Feb. 1, 2018).

22. Pellissier, L. *et al.* Quaternary Coral Reef Refugia Preserved Fish Diversity. *Science* **344,** 1016–1019. https://www.science.org/doi/full/10.1126/science.1249853 (2023) (May 30, 2014).

23. Holman, L. E. *et al.* Animals, Protists and Bacteria Share Marine Biogeographic Patterns. *Nature Ecology & Evolution* **5,** 738–746. ISSN: 2397-334X. https://www.nature.com/articles/s41559-021-01439-7 (2023) (6 June 2021).

24. Costello, M. J. *et al.* Marine Biogeographic Realms and Species Endemicity. *Nature Communications* **8,** 1–10. ISSN: 2041-1723. https://www.nature.com/articles/s41467-017-01121-2 (2023) (1 Oct. 20, 2017).

25. Costello, M. J. & Chaudhary, C. Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current Biology* **27,** R511–R527. ISSN: 0960-9822. https://www.sciencedirect.com/science/article/pii/S0960982217305055 (2023) (June 5, 2017).

26. 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. https://www.science.org/doi/full/10.1126/science.1254421 (2023) (Sept. 12, 2014).

27. Sommeria-Klein, G. *et al.* Global Drivers of Eukaryotic Plankton Biogeography in the Sunlit Ocean. *Science* **374,** 594–599. https://www.science.org/doi/full/10.1126/science.abb3717 (2023) (Oct. 29, 2021).

28. Sánchez, P. *et al.* Marine Picoplankton Metagenomes and MAGs from Eleven Vertical Profiles Obtained by the Malaspina Expedition. *Scientific Data* **11,** 154. ISSN: 2052-4463. https://www.nature.com/articles/s41597-024-02974-1 (2024) (1 Feb. 1, 2024).

29. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* **179,** 1084–1097.e21. ISSN: 0092-8674. https://www.sciencedirect.com/science/article/pii/S0092867419311249 (2023) (Nov. 14, 2019).

30. Topor, Z. M., Rasher, D. B., Duffy, J. E. & Brandl, S. J. Marine Protected Areas Enhance Coral Reef Functioning by Promoting Fish Biodiversity. *Conservation Letters* **12,** e12638. ISSN: 1755-263X. https://onlinelibrary.wiley.com/doi/abs/10.1111/conl.12638 (2023) (2019).

31. Harrison, H. B., Bode, M., Williamson, D. H., Berumen, M. L. & Jones, G. P. A Connectivity Portfolio Effect Stabilizes Marine Reserve Performance. *Proceedings of the National Academy of Sciences* **117,** 25595–25600. https://www.pnas.org/doi/abs/10.1073/pnas.1920580117 (2023) (Oct. 13, 2020).

32. Botsford, L. W. *et al.* Connectivity and Resilience of Coral Reef Metapopulations in Marine Protected Areas: Matching Empirical Efforts to Predictive Needs. *Coral Reefs* **28,** 327–337. ISSN: 1432-0975. https://doi.org/10.1007/s00338-009-0466-z (2024) (June 1, 2009).

33. Kiessling, W., Simpson, C. & Foote, M. Reefs as Cradles of Evolution and Sources of Biodiversity in the Phanerozoic. *Science* **327,** 196–198. https://www.science.org/doi/10.1126/science.1182241 (2024) (Jan. 8, 2010).

34. Cowen, R. K. & Sponaugle, S. Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science* **1,** 443–466. pmid: 21141044. https://doi.org/10.1146/annurev.marine.010908.163757 (2023) (2009).

35. Kinlan, B. P. & Gaines, S. D. Propagule Dispersal in Marine and Terrestrial Environments: A Community Perspective. *Ecology* **84,** 2007–2020. ISSN: 1939-9170. https://onlinelibrary.wiley.com/doi/abs/10.1890/01-0622 (2023) (2003).

36. Bretsky, P. W. Evolutionary Paleoecology of the Marine Biosphere. James W. Valentine. *The Quarterly Review of Biology* **49,** 252–253. ISSN: 0033-5770. https://www.journals.uchicago.edu/doi/10.1086/408094 (2023) (Sept. 1974).

37. Leprieur, F. *et al.* Plate Tectonics Drive Tropical Reef Biodiversity Dynamics. *Nature Communications* **7,** 11461. ISSN: 2041-1723. https://www.nature.com/articles/ncomms11461 (2023) (1 May 6, 2016).

38. Zaffos, A., Finnegan, S. & Peters, S. E. Plate Tectonic Regulation of Global Marine Animal Diversity. *Proceedings of the National Academy of Sciences* **114,** 5653–5658. https://www.pnas.org/doi/abs/10.1073/pnas.1702297114 (2024) (May 30, 2017).

39. Torfstein, A. & Steinberg, J. The Oligo–Miocene Closure of the Tethys Ocean and Evolution of the Proto-Mediterranean Sea. *Scientific Reports* **10,** 13817. ISSN: 2045-2322. https://www.nature.com/articles/s41598-020-70652-4 (2024) (1 Aug. 14, 2020).

40. Rögl, F. MEDITERRANEAN AND PARATETHYS. FACTS AND HYPOTHESES OF AN OLIGOCENE TO MIOCENE PALEOGEOGRAPHY (SHORT OVERVIEW). *Geologica Carpathica* **50,** 339–349 (1999).

41. Liu, H., Li, S., Ugolini, A., Momtazi, F. & Hou, Z. Tethyan Closure Drove Tropical Marine Biodiversity: Vicariant Diversification of Intertidal Crustaceans. *Journal of Biogeography* **45,** 941–951. ISSN: 1365-2699. https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.13183 (2024) (2018).

42. Barber, P. H. & Bellwood, D. R. Biodiversity Hotspots: Evolutionary Origins of Biodiversity in Wrasses (*Halichoeres*: Labridae) in the Indo-Pacific and New World Tropics. *Molecular Phylogenetics and Evolution* **35,** 235–253. ISSN: 1055-7903. https://www.sciencedirect.com/science/article/pii/S1055790304002969 (2024) (Apr. 1, 2005).

43. Bellwood, D. R., Herwerden, L. van & Konow, N. Evolution and Biogeography of Marine Angelfishes (Pisces: Pomacanthidae). *Molecular Phylogenetics and Evolution* **33,** 140–155. ISSN: 1055-7903. https://www.sciencedirect.com/science/article/pii/S1055790304001460 (2024) (Oct. 1, 2004).

44. Malaquias, M. A. E. & Reid, D. G. Tethyan Vicariance, Relictualism and Speciation: Evidence from a Global Molecular Phylogeny of the Opisthobranch Genus Bulla. *Journal of Biogeography* **36,** 1760–1777. ISSN: 1365-2699. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2009.02118.x (2024) (2009).

45. Read, C. I., Bellwood, D. R. & van Herwerden, L. Ancient Origins of Indo-Pacific Coral Reef Fish Biodiversity: A Case Study of the Leopard Wrasses (Labridae: Macropharyngodon). *Molecular Phylogenetics and Evolution* **38,** 808–819. ISSN: 1055-7903. https://www.sciencedirect.com/science/article/pii/S1055790305002538 (2023) (Mar. 1, 2006).

46. Roberts, G. G. & Mannion, P. D. Timing and Periodicity of Phanerozoic Marine Biodiversity and Environmental Change. *Scientific Reports* **9,** 6116. ISSN: 2045-2322. https://www.nature.com/articles/s41598-019-42538-7 (2023) (1 Apr. 16, 2019).

47. Renema, W. *et al.* Hopping Hotspots: Global Shifts in Marine Biodiversity. *Science.* https://www.science.org/doi/10.1126/science.1155674 (2023) (Aug. 1, 2008).

48. Hoeksema, B. W. in *Biogeography, Time, and Place: Distributions, Barriers, and Islands* (ed Renema, W.) 117–178 (Springer Netherlands, Dordrecht, 2007). ISBN: 978-1-4020-6374-9. https://doi.org/10.1007/978-1-4020-6374-9_5 (2023).

49. Hughes, A. C. *et al.* Sampling Biases Shape Our View of the Natural World. *Ecography* **44,** 1259–1269. ISSN: 1600-0587. https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.05926 (2023) (2021).

50. Ryan-Keogh, T., Thomalla, S., Chang, N. & Maolusi, T. Net Primary Production from the Behrenfeld-CbPM, Westberry-CbPM and Silsbe-CAFE Algorithms - HADLEY MLD 0.125 Criterion. https://zenodo.org/records/8320875 (2023).

51. Jordahl, K. *et al. Geopandas/Geopandas: V0.8.1* version v0.8.1. Zenodo, July 2020. https://doi.org/10.5281/zenodo.3946761.

52. Vavrek, M. J. Fossil: Palaeoecological and Palaeogeographical Analysis Tools.

53. Dixon, P. VEGAN, A Package of R Functions for Community Ecology. *Journal of Vegetation Science* **14,** 927–930. ISSN: 1100-9233. JSTOR: 3236992. https://www.jstor.org/stable/3236992 (2024) (2003).

54. Goslee, S. C. & Urban, D. L. The Ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software* **22,** 1–19. ISSN: 1548-7660. https://doi.org/10.18637/jss.v022.i07 (2024) (Sept. 30, 2007).

55. Mächler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. *Cluster: Cluster Analysis Basics and Extensions* (Jan. 1, 2012).

56. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

57. R Core Team. *R: A Language and Environment for Statistical Computing* manual (R Foundation for Statistical Computing, Vienna, Austria, 2020). https://www.R-project.org/.

58. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A Protocol for Data Exploration to Avoid Common Statistical Problems. *Methods in Ecology and Evolution* **1,** 3–14. ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2009.00001.x (2024) (2010).

59. Wei, T. *et al. Corrplot: Visualization of a Correlation Matrix* version 0.92. Nov. 18, 2021. https://cran.r-project.org/web/packages/corrplot/index.html (2024).

60. Paytan, A. & McLaughlin, K. The Oceanic Phosphorus Cycle. *Chemical Reviews* **107,** 563–576. ISSN: 0009-2665, 1520-6890. https://pubs.acs.org/doi/10.1021/cr0503613 (2024) (Feb. 1, 2007).

61. Sabur, M. A., Parsons, C. T., Maavara, T. & Van Cappellen, P. Effects of pH and Dissolved Silicate on Phosphate Mineral-Water Partitioning with Goethite. *ACS Earth and Space Chemistry* **6,** 34–43. https://doi.org/10.1021/acsearthspacechem.1c00197 (2024) (Jan. 20, 2022).

62. Ripley, B. *et al. MASS: Support Functions and Datasets for Venables and Ripley's MASS* version 7.3-60.0.1. Jan. 13, 2024. https://cran.r-project.org/web/packages/MASS/index.html (2024).

63. Reaka, M. L., Rodgers, P. J. & Kudla, A. U. Patterns of Biodiversity and Endemism on Indo-West Pacific Coral Reefs. *Proceedings of the National Academy of Sciences* **105,** 11474–11481. https://www.pnas.org/doi/abs/10.1073/pnas.0802594105 (2023) (supplement_1 Aug. 12, 2008).

64. Bellwood, D. R., Hughes, T. P., Connolly, S. R. & Tanner, J. Environmental and Geometric Constraints on Indo-Pacific Coral Reef Biodiversity. *Ecology Letters* **8,** 643–651. ISSN: 1461-0248. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2005.00763.x (2024) (2005).

65. Jablonski, D., Roy, K. & Valentine, J. W. Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient. *Science* **314,** 102–106. https://www.science.org/doi/full/10.1126/science.1130880 (2024) (Oct. 6, 2006).

66. Jablonski, D. *et al.* Out of the Tropics, but How? Fossils, Bridge Species, and Thermal Ranges in the Dynamics of the Marine Latitudinal Diversity Gradient. *Proceedings of the National Academy of Sciences* **110,** 10487–10494 (2013).

67. Krug, A. Z., Jablonski, D. & Valentine, J. W. Contrarian Clade Confirms the Ubiquity of Spatial Origination Patterns in the Production of Latitudinal Diversity Gradients. *Proceedings of the National Academy of Sciences* **104,** 18129–18134. https://www.pnas.org/doi/abs/10.1073/pnas.0709202104 (2024) (Nov. 13, 2007).

68. WILLIAMS, S. T. Origins and Diversification of Indo-West Pacific Marine Fauna: Evolutionary History and Biogeography of Turban Shells (Gastropoda, Turbinidae). *Biological Journal of the Linnean Society* **92,** 573–592. ISSN: 0024-4066. https://doi.org/10.1111/j.1095-8312.2007.00854.x (2023) (Nov. 1, 2007).

# Appendix A

# Supplement

## A.1 S1: OBIS query expressions

- "taxonid": "2", "areaid": "31904,34269,34274,34264,34266,34267,34265,10161,34276,34344,34347,34273,34268"
- "taxonid": "3", "areaid": "31904,34269,34274,34264,34266,34267,34265,10161,34276,34344,34347,34273,34268"
- "taxonid": "2", "areaid": "40026"
- "taxonid": "3", "areaid": "40026"
- "taxonid": "2", "areaid": "34360,34334,34338,34339,34332,34359,34349,34354,34358,34300,34361,216"
- "taxonid": "3", "areaid": "34300,34360,34334,34338,34339,34332,34359,34349,34354,34358,34361,216"

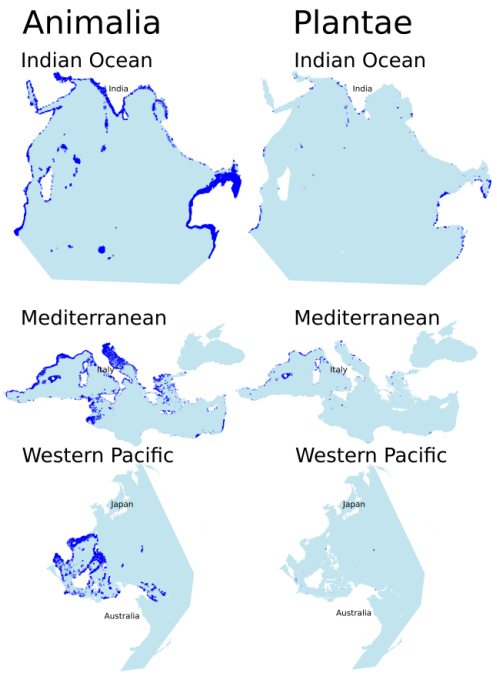



Figure A.1: The filtered OBIS datasets plotted on geographic maps. Kingdom Plantae has extremely uneven coverage.

## A.2 S2: OBIS Records

| phylum | IORecords | MedRecords | WPacRecords | %datasetIO | %datasetMed | %datasetWPac |
|---|---|---|---|---|---|---|
| **Annelida** | 14110 | 43752 | 6947 | 0.814704 | 13.505621 | 3.159566 |
| **Arthropoda** | 179683 | 55877 | 18482 | 10.374798 | 17.248437 | 8.405800 |
| Brachiopoda | 251 | 525 | 37 | 0.014493 | 0.162060 | 0.016828 |
| Bryozoa | 2517 | 1658 | 525 | 0.145330 | 0.511801 | 0.238775 |
| Chaetognatha | 3418 | 1772 | 64 | 0.197353 | 0.546991 | 0.029108 |
| **Chordata** | 1295678 | 128177 | 138805 | 74.811741 | 39.566420 | 63.129912 |
| **Cnidaria** | 34830 | 27684 | 18476 | 2.011065 | 8.545658 | 8.403071 |
| Ctenophora | 241 | 909 | 4 | 0.013915 | 0.280595 | 0.001819 |
| **Echinodermata** | 28195 | 13504 | 12565 | 1.627964 | 4.168493 | 5.714689 |
| Hemichordata | 175 | 71 | 2 | 0.010104 | 0.021917 | 0.000910 |
| Kinorhyncha | 11 | 17 | 4 | 0.000635 | 0.005248 | 0.001819 |
| **Mollusca** | 156811 | 44808 | 22529 | 9.054182 | 13.831593 | 10.246416 |
| Nematoda | 1590 | 521 | 37 | 0.091806 | 0.160825 | 0.016828 |
| Nemertea | 409 | 19 | 7 | 0.023615 | 0.005865 | 0.003184 |
| Platyhelminthes | 429 | 1 | 4 | 0.024770 | 0.000309 | 0.001819 |
| **Porifera** | 13554 | 4630 | 1381 | 0.782601 | 1.429215 | 0.628093 |
| Tardigrada | 16 | 29 | 3 | 0.000924 | 0.008952 | 0.001364 |

Table A.1: A summary of the filtered OBIS dataset for kingdom Animalia, subset by phylum. Bold phyla are data-rich phyla considered for further analyses.

## A.3 S3: The Filtered OBIS datasets

| phylum | IORecords | MedRecords | WPacRecords | %datasetIO | %datasetMed | %datasetWPac |
|---|---|---|---|---|---|---|
| Allomalorhagida | 4 | 8 | 2 | 0.000231 | 0.002472 | 0.000910 |
| **Anthozoa** | 30629 | 11490 | 17993 | 1.769264 | 3.550548 | 8.186376 |
| Appendicularia | 114 | 171 | 44 | 0.006585 | 0.052841 | 0.020019 |
| Arachnida | 8 | 4 | 2 | 0.000462 | 0.001236 | 0.000910 |
| **Ascidiacea** | 3722 | 3736 | 668 | 0.214999 | 1.154469 | 0.303924 |
| **Asteroidea** | 8780 | 2863 | 3344 | 0.507171 | 0.884701 | 1.521438 |
| **Bivalvia** | 21750 | 16818 | 7335 | 1.256374 | 5.196964 | 3.337246 |
| Branchiopoda | 831 | 1769 | 21 | 0.048002 | 0.546642 | 0.009554 |
| Calcarea | 1219 | 97 | 23 | 0.070415 | 0.029974 | 0.010464 |
| **Cephalopoda** | 85552 | 14168 | 549 | 4.941854 | 4.378082 | 0.249782 |
| Chromadorea | 1063 | 415 | 21 | 0.061403 | 0.128240 | 0.009554 |
| Clitellata | 72 | 26 | 4 | 0.004159 | 0.008034 | 0.001820 |
| **Copepoda** | 100496 | 19375 | 1127 | 5.805085 | 5.987108 | 0.512758 |
| **Crinoidea** | 2985 | 225 | 1334 | 0.172427 | 0.069528 | 0.606937 |
| Cubozoa | 158 | 328 | 15 | 0.009127 | 0.101356 | 0.006825 |
| Cyclorhagida | 7 | 9 | 2 | 0.000404 | 0.002781 | 0.000910 |
| **Demospongiae** | 12204 | 4420 | 1294 | 0.704956 | 1.365833 | 0.588738 |
| **Echinoidea** | 5594 | 3513 | 2768 | 0.323134 | 1.085559 | 1.259372 |
| **Elasmobranchii** | 235321 | 8724 | 3485 | 13.593161 | 2.695821 | 1.585590 |
| Enoplea | 527 | 106 | 16 | 0.030442 | 0.032755 | 0.007280 |
| Enteropneusta | 165 | 71 | 1 | 0.009531 | 0.021940 | 0.000455 |
| **Gastropoda** | 48226 | 12587 | 14208 | 2.785743 | 3.889534 | 6.464294 |
| Gymnolaemata | 1854 | 1558 | 496 | 0.107095 | 0.481441 | 0.225668 |
| Heterotardigrada | 16 | 29 | 3 | 0.000924 | 0.008961 | 0.001365 |
| Hexactinellida | 55 | 1 | 37 | 0.003177 | 0.000309 | 0.016834 |
| **Holocephali** | 2508 | 8 | 12 | 0.144873 | 0.002472 | 0.005460 |
| **Holothuroidea** | 3195 | 2672 | 2774 | 0.184557 | 0.825680 | 1.262102 |
| Homoscleromorpha | 76 | 112 | 27 | 0.004390 | 0.034609 | 0.012284 |
| Hoplonemertea | 151 | 3 | 2 | 0.008722 | 0.000927 | 0.000910 |
| **Hydrozoa** | 3962 | 12211 | 420 | 0.228862 | 3.773346 | 0.191090 |
| **Malacostraca** | 73524 | 34151 | 16801 | 4.247065 | 10.553070 | 7.644045 |
| Monogenea | 72 | 1 | 4 | 0.004159 | 0.000309 | 0.001820 |
| Myxini | 66 | 1 | 31 | 0.003812 | 0.000309 | 0.014104 |
| **Ophiuroidea** | 7641 | 4231 | 2345 | 0.441377 | 1.307430 | 1.066918 |
| **Ostracoda** | 3940 | 175 | 13 | 0.227591 | 0.054077 | 0.005915 |
| Pilidiophora | 55 | 16 | 5 | 0.003177 | 0.004944 | 0.002275 |
| **Polychaeta** | 14038 | 43726 | 6943 | 0.810896 | 13.511860 | 3.158896 |
| Polyplacophora | 601 | 328 | 119 | 0.034716 | 0.101356 | 0.054142 |
| Pycnogonida | 377 | 64 | 93 | 0.021777 | 0.019777 | 0.042313 |
| Rhynchonellata | 212 | 486 | 31 | 0.012246 | 0.150180 | 0.014104 |
| **Sagittoidea** | 3418 | 1772 | 64 | 0.197438 | 0.547569 | 0.029118 |
| Scaphopoda | 682 | 807 | 318 | 0.039395 | 0.249373 | 0.144682 |
| Scyphozoa | 75 | 3655 | 47 | 0.004332 | 1.129439 | 0.021384 |
| Stenolaemata | 663 | 100 | 29 | 0.038298 | 0.030901 | 0.013194 |
| **Teleostei** | 1053418 | 112763 | 134533 | 60.849991 | 34.845123 | 61.209234 |
| Tentaculata | 168 | 737 | 4 | 0.009704 | 0.227742 | 0.001820 |
| Thaliacea | 498 | 2743 | 29 | 0.028767 | 0.847620 | 0.013194 |
| Thecostraca | 480 | 339 | 356 | 0.027727 | 0.104755 | 0.161971 |

Table A.2: A summary of the filtered OBIS dataset for kingdom Animalia, subset by class. Bold classes are data-rich classes considered for further analyses, defined as having more than 2500 records in the Indian Ocean.

## A.4 S4: Occurrence-based rarefaction curves for each ocean basin

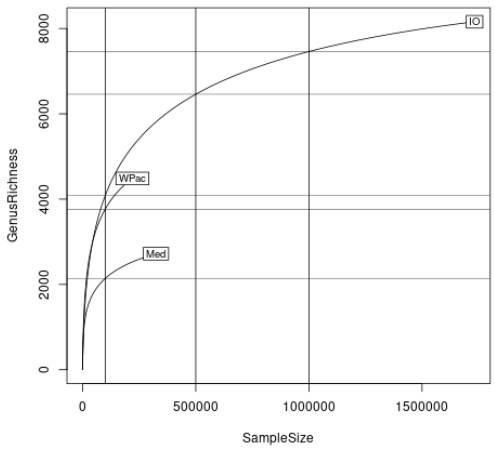



Figure A.2: Occurrence-based rarefaction curves for each ocean basin. While the Indian Ocean and Mediterranean are well-sampled, the Western Pacific is not.

## A.5 S5: Box-plots of z-transformed oceanographic variables

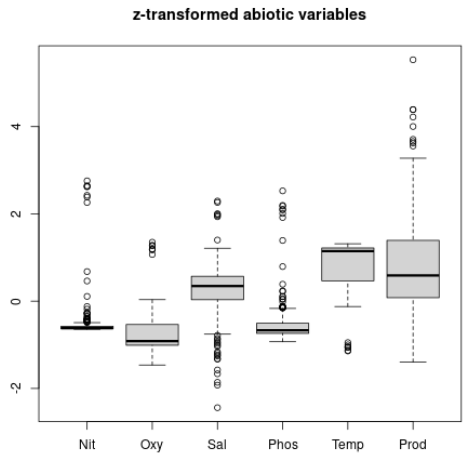



Figure A.3: Box-plots of oceanographic variables. Nitrate, dissolved oxygen (DO), phosphate, and productivity values are right-skewed while temperature and salinity are left-skewed.

## A.6 S6: Taxa reported from "n" grid-cells

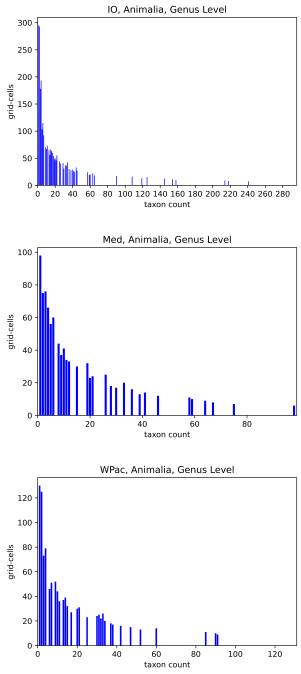



Figure A.4: x-axis = number of genera. y-axis = the number of grid-cells that report those many taxa. Most genera are reported from very few grid-cells.

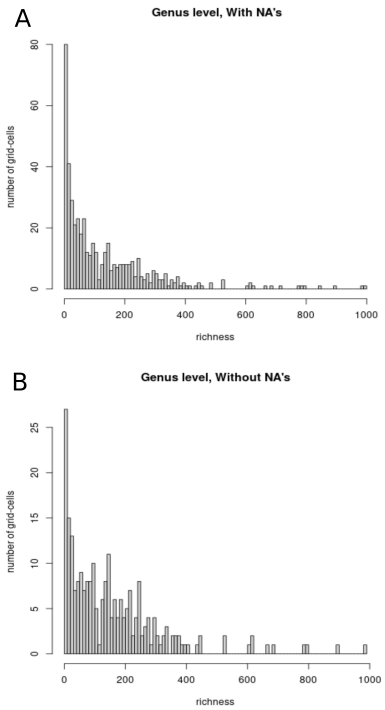



Figure A.5: Histograms of genus richness with the number of grid-cells reporting a given genus richness value. The data does not change qualitatively after dropping grid-cells with missing values.

## A.7 S7: Data structure before and after dropping grid-cells with missing values

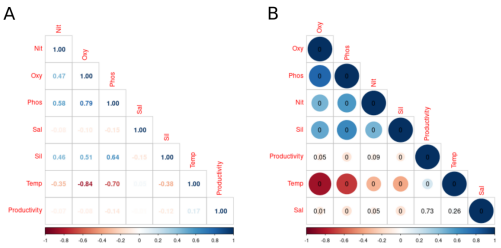



Figure A.6: Correlations between oceanographic parameters **A:** Numbers and colour intensities indicate Pearson's correlation coefficients. **B:** Circle sizes indicate the strength of correlation, numbers are p-values.

## A.8 S8: Correlations of oceanographic variables with each other

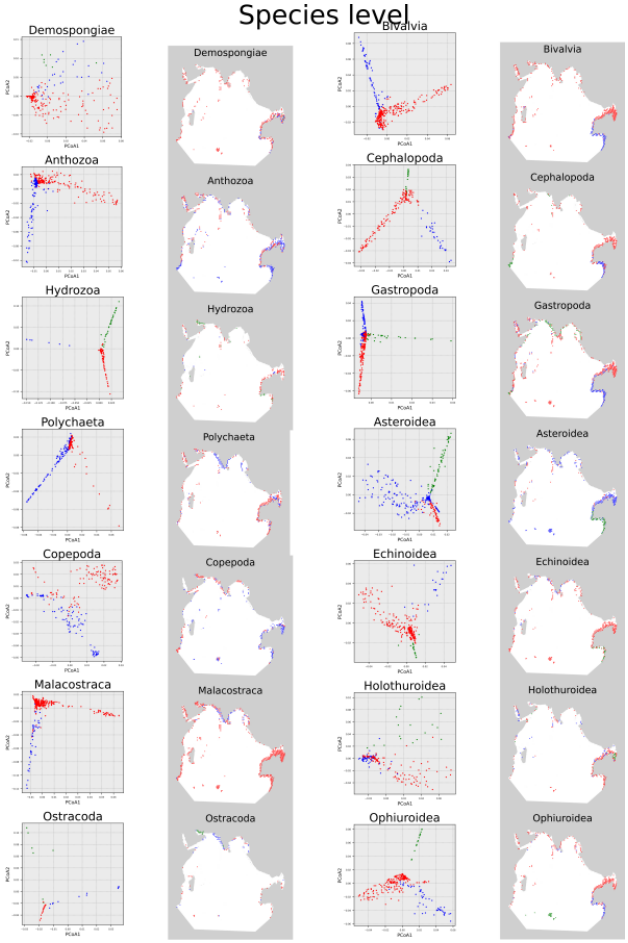
# Appendix B

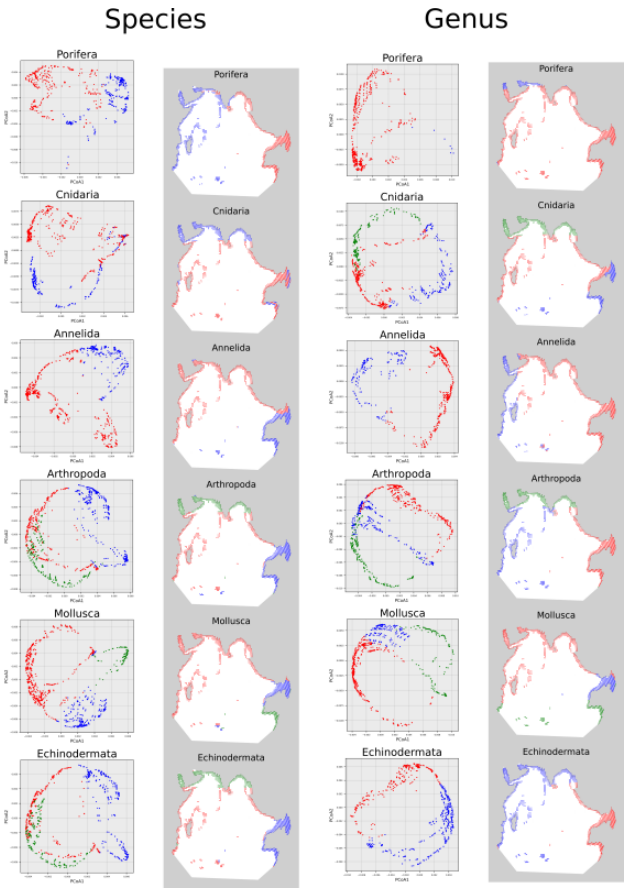
# Appendix

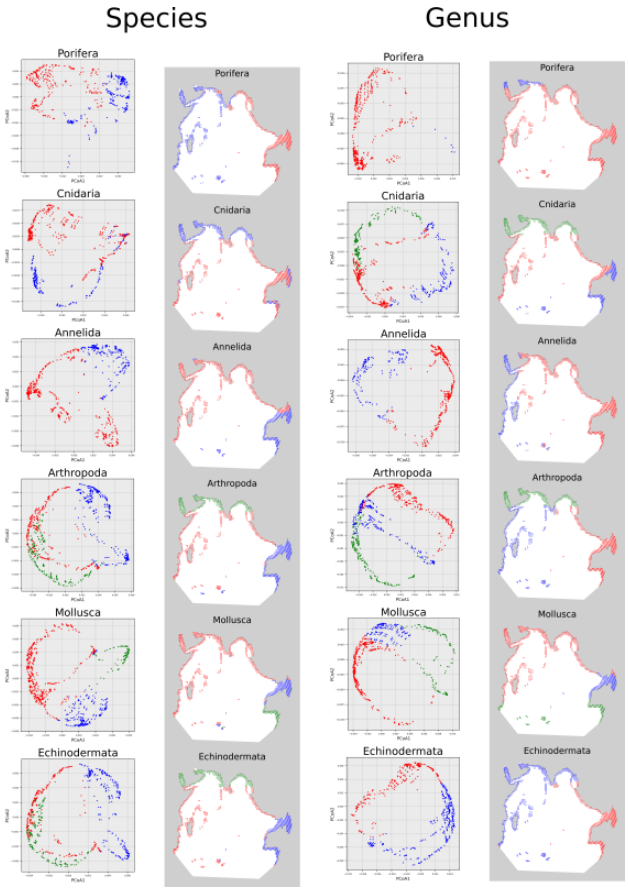
This appendix contains details of how spatial rangethrough was implemented with its analyses along with all analyses done at the species level.

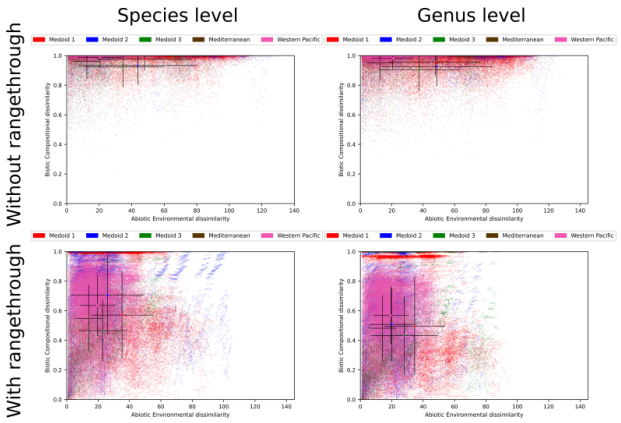
## B.1 Implementing spatial rangethrough

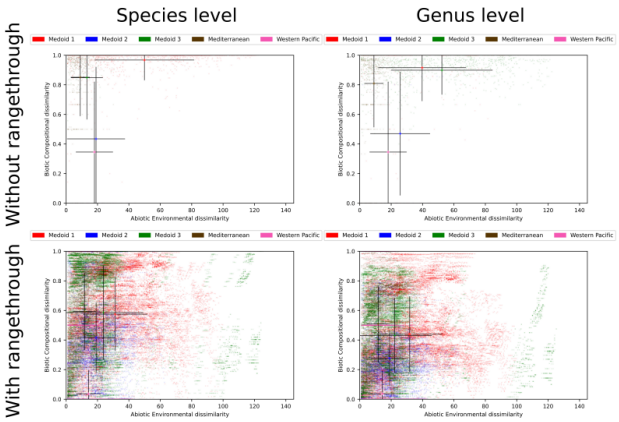
For a given species, we implemented this by listing all the grid-cells reporting said species and took the cartesian product of the minimum and maximum latitudes and longitudes it was reported from to generate an augmented, assumed range for each species. Then, only the grid-cells intersecting with the 200m shapefile were retained from these. Analyses were repeated using these augmented ranges alongside the species/genus occurrences as recorded in OBIS.


Species level
Demospongiae
Anthozoa
Hydrozoa
Polychaeta
Copepoda
Malacostraca
Ostracoda
Bivalvia
Cephalopoda
Gastropoda
Asteroidea
Echinoidea
Holothuroidea
Ophiuroidea


Species
Genus
Porifera
Cnidaria
Annelida
Arthropoda
Mollusca
Echinodermata


Species
Genus
Porifera
Cnidaria
Annelida
Arthropoda
Mollusca
Echinodermata


Species level
Genus level
Without rangethrough
With rangethrough
Medoid 1
Medoid 2
Medoid 3
Mediterranean
Western Pacific
Biotic Compositional dissimilarity
Abiotic Environmental dissimilarity


Species level
Genus level
Without rangethrough
With rangethrough
Medoid 1
Medoid 2
Medoid 3
Mediterranean
Western Pacific
Biotic Compositional dissimilarity
Abiotic Environmental dissimilarity