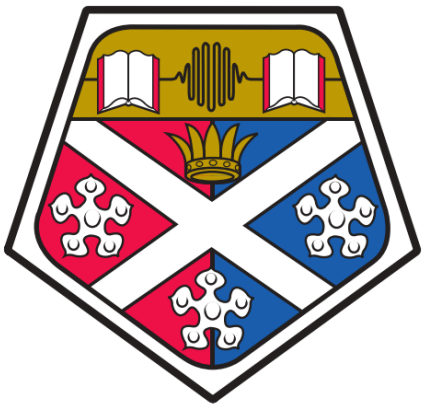
Latitudinal diversity gradient of marine phytoplankton through the lens of similarity

Matthew Coster Fleet Gray



Department of Mathematics and Statistics

University of Strathclyde

Glasgow, Scotland

A thesis submitted for the degree of Doctor of Philosophy

May 2025

Abstract

Phytoplankton are one of the oldest living groups of organisms on Earth. All of life owes its start to these oxygenic photoautotrophs arising around 2.7 billion years ago. Their existence oxygenated the Earth's atmosphere and their expansion from the ocean to land paved the way for modern plant-based agriculture and the advancement of humanity. Today, we know they are a speciose group inhabiting most aquatic ecosystems including the global ocean. Yet, general ecological understanding of their diversity distributions at the macroecological scale remains unclear despite their crucial role in global biogeochemical cycling and ecosystem productivity.

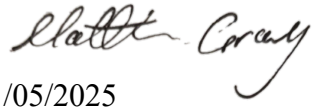
This thesis focuses on examining the existence of the latitudinal diversity gradient (LDG) of marine phytoplankton and how this pattern responds to changes in how diversity is mathematically defined. While the LDG is observed across fauna and flora, exceptions are observed in both marine and terrestrial ecosystems. Alignment of a group to the LDG leads itself to predictable environmental relationships such as temperature which covary with latitude, enhancing predictive modelling approaches required for biological modelling of ecological function. Deviation from the expected LDG relationship can identify evolutionary adaptions to environmental conditions or the role of time and location in species distributions. The knowledge gained with respect to alignment and deviation from this exploration of the LDG pattern of marine phytoplankton, informs future modelling and sampling directions globally by characterising three different components of diversity.

Here I use a recently compiled global dataset of marine phytoplankton to assess the global distribution of marine phytoplankton diversity. Typically, at macroecological scales, the definition of diversity is limited to species richness, particularly for latitudinal diversity gradients (LDGs). Here in Chapter 2, I found that the inclusion of community relative abundances using the Hill number index resulted in maximal diversity in the sub-tropical latitudinal regions. Following this, I further expand the mathematical definition of diversity in Chapter 3 to include taxonomic similarity using the Leinster and Cobbold diversity index (LCI) resulted in a singular peak of diversity in the southern hemisphere. Diversity assessments are always affected by sampling effort, and here this is particularly the case, rarefaction is used in Chapter 2 and 3 to assess the effect of effort on global distributions of marine phytoplankton diversity. Following the observational data research performed in Chapter 2 and 3, Chapter 4 focuses on modelling spatial distributions with limited observational data. In addition, in Chapter 4 I examine the diversity-productivity relationship of marine phytoplankton in the context of limited observational data availability and the modelling approach applied. All diversity distributions of marine phytoplankton are reviewed with respect to observed macroecological patterns.

This thesis contributes the first descriptions of global marine phytoplankton diversity using the Hill number diversity index and the LCI while assessing the sampling effort issues therein. The novel methodological development of the code required to assess this global dataset will contribute to new research focused on assessing diversity beyond the species richness paradigm. The species distribution model (SDM) research adds to the ongoing evaluation of SDM applicability and in the field of ecology. While the diversity-productivity relationship observed in the modelling research, suggests that SDMs may provide a novel avenue of predictive modelling of chlorophyll-a while capturing community composition.

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: 

Date: 01/05/2025

Acknowledgements

I would like to dedicate this thesis, to my Auntie, June Guild and Grandmother, Phyllis Gray. Your love, support and belief in me, as I am and who you see me to be, gave me the ability to chase my dreams, wherever they led. Your love of nature, grown from the beautiful gardens you both lovingly tended to season to season, decade to decade, taught me the enduring resilience of women, nature and family.

Thank you to my two supervisors, Dr. Bingzhang Chen and Dr. Michael Grinfeld, without your patience and the space to grow afforded to me, I would not be the independent researcher I am today.

With extreme gratitude, the fiscal support of this research from the University of Strathclyde (UOS) and Hong Kong University of Science and Technology (HKUST) for the opportunity to conduct this PhD via the research excellence award. The Scotia-Canadian Ocean Research Exchanges (SCORE) scheme support from partnership of Marine Alliance of Science and Technology Scotland (MASTS) and the Ocean Frontier Institute (OFI) in Canada was an incredible opportunity to develop research skills and form lifelong scientific collaborations.

I can not thank enough the support, friendship and scientific kinship found at the Future of Marine Ecosystems (FOME) lab at Dalhousie University during and beyond my research visit through the SCORE scheme. I am indebted to the leadership and guidance of Prof Derek Tittensor, for not only research directed guidance during SCORE but the continued life mentorship beyond the SCORE scheme period. To Dr Gabriel Reygondeau and Dr Yulia Egorova, I am endlessly grateful for your time and expertise guidance in this work. To Dr. Andrea Bryndum-Buchholz, my dear friend, to the fast friendship and support given to a fresh arrival, the long hours working virtually together across the Atlantic Ocean and the endless encouragement reminding me the light at the end of the tunnel does indeed exist, thank you. To Dr. Kristina Boerder, thank you for taking a chance on a chancer like me, your belief in my abilities as a researcher has given me confidence. I am endlessly inspired by the work you do and the person you are.

To the trans-Atlantic friends bonded together through the trials of fire that is life, without our open-hearted support through thick and thin, I am a better person for knowing your friendship. Thank you Alex for extending the frond of your friendship throughout and beyond the PhD experience. To Ashton, Maddie and Hania, I thank you for reminding me I am more than a PhD student, that we can make art, pottery and find joy outside of our work. To Ann, thank you for being an endless supporter. To Aislinn and Greg, thank you for roping me into climbing and the wider Strathclyde University Mountaineering Club community, a community to find joy in the pursuit of the outdoors, in laughter and chaos.

To the musicians, artists and creatives of the world, the solace I have found in your creations has given me space to reflect, dance and feel. Without imagination beyond what exists, we can not be more than we are. It is with great appreciation, I want to say thank you to you all.

Glossary

Fine-grain scale Diversity
A measure of species diversity within a particular area or ecosystem, usually expressed as species richness or using diversity indices.

Chlorophyll a (Chl-a)
A pigment found in phytoplankton used as a proxy for estimating biomass and primary productivity in oceanographic studies, often derived from satellite remote sensing.

Diversity Index
A mathematical measure that reflects how many different species are in a dataset, and in some cases, how evenly individuals are distributed among those species.

Effective Number of Species
The number of equally abundant species required to give the same value of a diversity index as that observed in the real community.

Coarse-grain scale Diversity
The total species diversity in a large region or across multiple ecosystems.

Hill Numbers
A unified family of diversity indices that incorporate species richness and evenness, parameterized by an order q to emphasize rare or abundant species.

Leinster and Cobbold Index
A similarity-sensitive diversity index that accounts for both species' abundances and pairwise similarities, allowing for a more nuanced assessment of biodiversity.

Macroecology
The study of broad-scale patterns and processes in ecology, typically across large spatial and temporal scales.

Marine Phytoplankton
Microscopic photosynthetic organisms living in the ocean's upper layers, forming the base of the marine food web.

Rarefaction
A technique used to standardise samples to a common size to compare diversity metrics fairly across different sampling efforts.

Remote sensing
The use of satellite or airborne sensor technologies to detect and monitor physical characteristics of an area, such as chlorophyll concentrations in the ocean.

Taxonomic Similarity
A measure of relatedness between species based on their taxonomic classification, often used to weight or refine diversity indices.

Unimodal/Bimodal Latitudinal Diversity Gradient (LDG)
Patterns describing how species richness varies with latitude, with unimodal showing a single peak (often in the tropics) and bimodal showing two peaks (often mid-latitudes).

Modality

A term used to describe the mode of a statistical distribution i.e. unimodal or bimodal.

Table of Contents

List of Figures

List of Tables

Abbreviations

LDG – Latitudinal Diversity Gradient

LCI – Leinster and Cobbold Index

GAM – Generalised additive model

HPC – High performance computing

HSI – Habitat suitability index

EEZ – Economic Exclusion Zone

SDM – Species distribution model

PAR – Photosynthetically active radiation

AUC – Area under the curve

1 Beyond Richness: Exploring the Macroecology of Marine Phytoplankton and Latitudinal Diversity Gradients

1.1 Introduction

The broad aim of this review is to examine how marine phytoplankton may align within the expected Latitudinal Diversity Gradient (LDG) patterns and underlying mechanisms, while highlighting how traditional diversity assessment approaches overlook community evenness and similarity, thereby identifying key gaps this thesis seeks to address. This objective is addressed by examining the ecological and environmental role of marine phytoplankton in key biosphere processes such as carbon cycling, food web structuring, and ecosystem stability, as they relate to the mechanisms and hypotheses underlying the LDG. The study of macroecological patterns such as the LDG and their emergent structures often relies on the assessment of biological diversity, a concept that is frequently conflated with species richness. This review seeks to clarify the distinction between diversity and richness, exploring the crucial role diversity measurement has played in ecology and other analytical pursuits in other disciplines. It includes a detailed discussion of the general definition of diversity and an evaluation of diversity indices that capture its complexities beyond richness alone. Finally, the review examines existing studies of latitudinal diversity gradients (LDGs) and introduces potential explanatory mechanisms across taxa and biomes, while maintaining focus on reviewing our understanding of the marine phytoplankton LDG to date.

1.1.1 Co-evolution of ecological understanding and diversity quantification

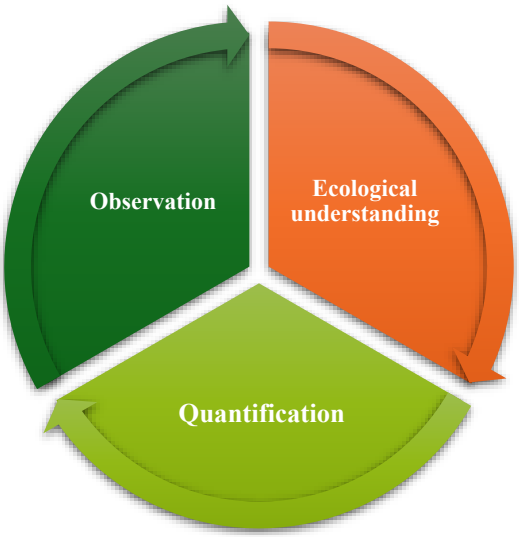


Figure 1.1 - Graphic detailing the cyclic relationship of measuring diversity, developing greater ecological understanding which requires complex quantification of ecological properties.

The observation of species and resulting biodiversity has been a fundamental aspect of human knowledge. Our understanding stretches back to the founders of biological study, who were early naturalists such as Aristotle in the early to mid fourth century BC, who examined the natural world with a systematic state of mind, grouping organisms based on method of motility for example, bipedal or quadrupedal. Aristotle noted the similarity of humans and apes and used broader physical qualities such as animalian jaw shape to suggest relatedness (Aristotle, 1887).

Later, Alexander von Humboldt noted the increase in diversity of life from northern temperate region to the tropics when on expedition to Venezuela, an early description of the latitudinal diversity gradient (Humboldt, Alexander von et al., 1850). Charles Darwin during their survey expeditions on the HMS Beagle and from reviewing scientific thought at the time formularised the theory of evolution through natural selection leading to species diversity (Darwin, 1859). Alongside western scientific thought on the matter, Indigenous knowledge systems (Brooks et al., 2024) have long documented the interactions between organisms and their environments. The origin of the taxonomic classification system still used today dates back to the 18th century, pioneered by Carl Linnaeus, who began formally quantifying and classifying species within a structured taxonomic framework (Linné, 1789), linking them to an evolutionary context.

In the early days of biodiversity studies, a major challenge was the lack of a formally described species unit. Researchers recognised the need not only to classify species but also to count individuals within species and within specific areas (Gleason, 1922; Whittaker, 1960). This led to the development of rank abundance plots (Preston, 1948), which revealed a fundamental ecological pattern: species abundance distributions tend to follow general statistical models, where species and individual counts can be described mathematically (Magurran, 2004). This species-area mathematical relationship was detailed shortly after by MacArthur and Wilson, stating that the number of species (S) scaled by area (A) with a power (z) of 0.1 to 0.4 (Hill, 1973a; MacArthur & Wilson, 1967). From these early relative abundance analyses, species accumulation curves emerged. These curves reflect how species richness increases with sampled area, eventually approaching an asymptote where extensive effort is required to detect

rare species (McGill et al., 2007). The effort required to move from a general diversity estimate to a specific, exhaustive inventory of biodiversity is enormous (Mora et al., 2011). Since complete species inventories are rarely feasible, ecologists apply species accumulation models to estimate total biodiversity. This highlights the role of statistical diversity estimation required to capture patterns of diversity globally (Elith & Leathwick, 2009). As a result, any biodiversity analysis and the diversity indices applied provide only an approximation of true diversity within a taxon, community, or ecosystem. However, the choice of index shapes which components of diversity are emphasised and considered most important in a given study.

The progression of biodiversity indices followed a logical path: starting with species richness (the number of species in an area), then incorporating abundance (counts of individuals per species), and then later including measures of similarity and disparity (Daly et al., 2018). Species richness while being technically simple, counts of different species requires area standardisation to ensure comparability between studies and time points (Whittaker, 1960). As the geographic area assessed is increased the likelihood of discovering new species also increases (Macarthur, 1965; Rosenzweig, 1995). While community evenness, a key component of diversity, emerged as a function of both species' richness and relative number of occurrences within a community (Hill, 1973b). Today, species richness and evenness have moved towards incorporation of quantification methods of community similarity. While we have recognised the related nature of life on Earth for over a thousand years, at present species richness and evenness diversity metrics do not capture how related species are to one another in a community, let alone quantify it.

The quantification of similarity is important; we are balancing the needs of ourselves within the diverse ecosystem of life on Earth. To be pragmatic, if we aim to conserve and protect maximally, the most diverse ecosystems and the area they cover, we need to know where those are (Díaz et al., 2019; Myers et al., 2000). Species richness and community evenness assessments are indicators of these regions, but other diversity metrics now exist with greater capacity for distinctiveness between areas (Daly et al., 2018).

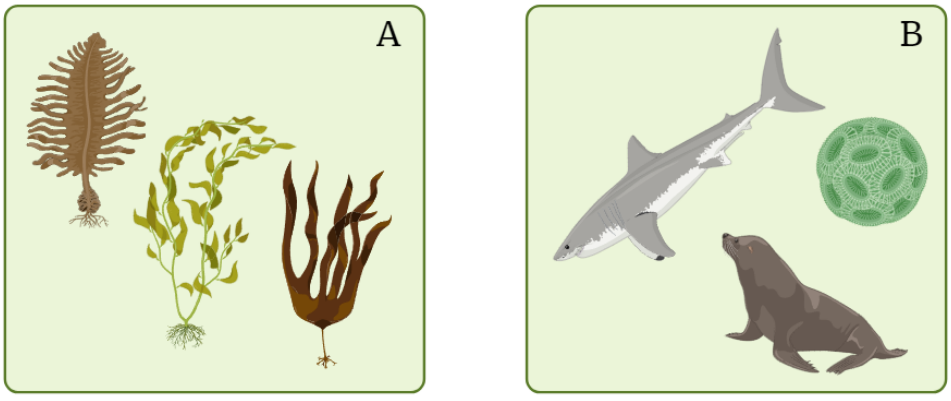


Figure 1.2 - Conceptual diagram of the merits of the Leinster and Cobbold index where Fig. 1A is a community of three seaweeds and Fig. 1B is a community composed of phytoplankton, a great white shark and a sea lion. Figure Made in BioRender.com

Shown in Figure 1.2 are two communities, each with three species and each with a single individual of each species. Traditional diversity metrics, which measure richness or a mix of richness and evenness measurements, are unable to distinguish these two communities (Fig 1.2) as they share the same richness and evenness. This is despite the visual observation that they differ substantially ecologically. Here, community A contains three species of seaweeds while community B contains a great white shark, californian sea lion and a coccolithophore

species. The diversity of these communities differ substantially when we consider the relatedness of the species assessed.

Communities with more similarity, be that genetic, functional or taxonomic, are less diverse (*Community A,* Figure 1.2) and it follows then that communities with less similarity are more diverse (*Community B,* Figure 1.2) (Purvis & Hector, 2000; Stirling & Wilsey, 2001). Community similarity is the quantification of the relatedness of species within a community (Leinster & Cobbold, 2012a). Similar vocabulary is applied by users of the Bray-Curtis Similarity Index (BCI). The difference being that similarity in BCI application, the compositional distance-based similarity between communities is being compared using species abundances. This is also true of the Jaccard dissimilarity measure (Bray & Curtis, 1957; Ricotta & Podani, 2017). On the other hand, the LCI index assesses similarity of a community within the area or community assessed using a user-defined community similarity, such as taxonomy. Therefore, the distinctiveness capacity of the LCI, which is sensitive to similarity, when similarity is quantified is immediately apparent. In addition, when the LCI is calculated with what the authors call a 'naïve' model, i.e. each species is treated as fully distinct objects the results of the index match the Hill number diversity index measure. This unifies a suite of diversity measures using a single methodology.

These metrics capture the evolutionary and ecological relationships between species, considering both their functional roles and phylogenetic distances (Chao, Chiu, et al., 2014; Leinster & Cobbold, 2012a; Ricotta & Podani, 2017). Ultimately, species distributions and diversity patterns are shaped by a complex interplay of area, dispersal, competition, and environmental constraints, highlighting the ever-evolving nature of ecological research and complexity therein. At present, biodiversity assessment on macroecological scale e.g. global has not moved beyond species richness (Righetti et al., 2019; Tittensor et al., 2010a).

Defining the geographic boundaries of a species range is complex, as species distributions are influenced by degree of mobility potential, habitat availability, abiotic barriers to movement or competitive exclusion (Macarthur, 1965). For mobile species, defining strict spatial boundaries is particularly challenging. The observed habitat occupation of a species and the resulting geographic range of said species can differ from the potential geographic range of a species (Guisan et al., 2017). The fundamental niche of a species is anywhere that species can survive, grow and reproduce in absence of all abiotic and biotic constraints for resources or barriers to access. The realised niche of a species is a subset of the fundamental niche where that species actually resides with abiotic and biotic interactions present.

Quantifying diversity inevitably reflects the influence of geographic area, as larger or more heterogeneous regions sampled tend to support greater biodiversity (Kallimanis et al., 2008; Macarthur, 1965). In marine phytoplankton, the relationship between diversity, geographic area, and niche occupation remains poorly characterised due to data availability and sampling biases (Brun et al., 2015; Irwin et al., 2012), complicating efforts to fully capture their role in key ecological processes and patterns such as the Latitudinal Diversity Gradient.

While diversity quantification has evolved over time and our ecological understanding has grown, there is a lag in the implementation of complex diversity indices in macroecological studies owing to poor data resolution to date. Global assessments of diversity typically assess a snapshot in time (Malviya et al., 2016; Tittensor et al., 2010b). As observed for bacterial communities, the LDG may also change seasonally (Ladau et al., 2013) presenting challenges in performing static pattern assessments. Localised research settings typically lend themselves to richer data availability in space and time (Anderson, 2018; Lu & Jetz, 2025). Whereas regional studies is a mix of constraints but typically allows for higher data availability than that

of global macroecological studies. The issue lies in the reality that as diversity quantification complexity increases, more data is required. As shown with the founders of biological study, ecological understanding originates from observation and these observations aid in the development of theories. To test such theories methods of diversity quantification were developed, raising more questions to pursue, starting the cycle all over again (Figure 1.2.1). The time from one part of this cycle to the next is usually not linear or separate, often occurring simultaneously.

1.1.2 Technical introduction of diversity indices

Mathematically speaking, diversity indices are equations used to measure and compare the properties of elements within a set, across sets, and their relationships as a subset and or superset. The delineation of what elements belong to each set and which sets belong to subsets or supersets is based on rules determined by the user. This is the basis of any diversity analysis in any discipline. The broad scope of this statement holds as diversity indices are applied in a variety of subject areas such as economics to quantify the diversity of a stock portfolio (Carmichael et al., 2023; Woerheide & Persson, 1992). They are also used in geology to describe the diversity of geomorphology in an area (Ferrer-Valero, 2018). Irrespective of application, they provide insight to the distribution of the number of elements and or sets measured.

Diversity is not comprised of a single component but three core components at present. Richness, the number of species in a community or sample, is typically considered the primary diversity component and as a result, diversity as a term is often conflated as richness in many papers (Righetti et al., 2019; R. J. Whittaker et al., 2001; Zacaï et al., 2021a). The second component is evenness of the community which is calculated as a relative measure. Each individual of each species in a community or sample is counted, and the sum of all individuals in the community is the total number of individuals. Relative abundance is the relative proportion of each species in the community, this is usually normalised to a range between one and zero using the total number of individuals in a sample. The third component, similarity as it will be defined here, is often not included in diversity measurement. It is commonly used as a measure of beta diversity to quantify the overlap in a communities species representation through time or sample areas (Bray & Curtis, 1957; McCormick et al., 1992).

In this thesis, similarity is defined as the similarity of species attributes as quantified by the Leinster and Cobbold diversity index (LCI). In this case it is taxonomy as it was the most readily tractable, but genetic or functional attributes can be supplanted for taxonomy. Similarity is quantified by using discretised partitions. Taxonomy naturally lends itself to this as it is hierarchal and uniform to each species in the community or sample. Each taxonomic rank e.g. genus, class, phylum, is compared for each species to every other species. The total number of discretised levels in similarity quantification is then used to normalise the similarity between two species between a value of zero (no similarity) and one meaning the two species are the same species (Leinster & Cobbold, 2012a).

1.1.2.1 Richness

Species richness measurement could be thought of as the number of elements within a set e.g. an area or sample. Species richness represents the species count within a given area (R. H. Whittaker, 1972). It is denoted by *S*:

$$Number\ of\ species\ =\ S \tag{1}$$

The counting of number of species can be performed is methodologically agnostic, depending on the biota being observed and recorded, for example whales are counted from prop planes to

cover large areas while phytoplankton are counted by microscopy. Species richness serves as the foundation for all other diversity indices and is a core measurement of difference. This difference is dependent on the Linnean classification of species. What defines a species as a species is generally the ability of two individuals of a population to produce fertile offspring. We can also review the DNA of individuals to confirm taxonomic relationships to determine relatedness (Avise, 2000; Zuckerkandl & Pauling, 1965) or we can also assess similarity of ecological niche space usage and relative conservatism (Darwin, 1859; Losos, 2008; Malfertheiner et al., 2026). Reproductive isolation also occurs evolutionarily, often different species methods of reproduction are not compatible, inhibiting reproduction (Dobzhansky, 1937; Mayr, 1996). Where inter-species reproduction can occur is with closely related species, such as lion and tigers, or donkeys and horses. Species which have high morphological similarity due to close relatedness are more likely to hybridise in this manner. Hybrid individuals may also not be able to reproduce with either of their parental species but with other hybrids of the same parental mix or be able to reproduce at all (Rieseberg, 1997).

Another way in which the species concept loses its rigid definition is in small micro-species such as phytoplankton. As noted, reproduction, morphological similarity and ecological niche space can be used to characterise a species unit. Phytoplankton can reproduce asexually and in addition do not require reproduction to transfer genetic material. By using a process called horizontal gene transfer, phytoplankton can gain beneficial adaptive capabilities, i.e. being able to metabolise a nutrient in greater environmental availability due to the transfer of genes to do so (Keeling & Palmer, 2008). In addition, cyanobacteria genera such as *Prochlorococcus* and *Synechococcus* contain species but without genetic analysis are morphologically indistinguishable and could considered a species complex instead. These exceptions to the rule of species definition are important for context, we are applying a generalised system of classification to a naturally complex system. It is also of importance to note that for most assessments describing a species, the Linnean approach is applicable and accepted.

1.1.2.2 Shannon-Wiener Index

Evenness measures are widely used, though their sensitivity can vary depending on how the relationship between abundances and richness is weighted. It is mathematically challenging to make a diversity measure of evenness without influence of species richness because relative abundance distributions require at least two species to form a distribution. The number of species in a community and the number of those individuals within each species population is required for community evenness estimation. For a community with only one species, the relative abundance will always be maximal ($p = 1$), a scenario rarely encountered in practice.

The Shannon-Wiener Index (H) quantifies the diversity of a community by considering both species richness (S) and the evenness of species abundances. It is calculated using the formula:

$$H = -\sum_{i=1}^{S} p_i \ln(p_i) \tag{2}$$

where p_i is the proportional abundance of the i-th species.

The maximum possible value of H occurs when all species are equally abundant, meaning p_i = 1/S for each species. In this case:

$$H_{max} = -\sum_{i=1}^{S}\left(\frac{1}{S} ln\left(\frac{1}{S}\right)\right) \quad (3)$$

1.1.2.3 Pielou's Evenness

Pielou's Evenness (J) is calculated using the Shannon-Wiener index (H) and the maximum possible Shannon-Wiener index (H_{MAX}), which depends logarithmically on the species richness (S):

$$J = \frac{H}{H_{max}} = \frac{H}{\ln(S)} \quad (4)$$

1.1.2.4 Simpson's Index

Simpson's Index (D) quantifies diversity by incorporating both richness and evenness. It is calculated as:

$$D = \sum_{i=1}^{S} p_i^2 \quad (5)$$

Where:

- S: species richness
- p_i: relative abundance of species i

Simpson's Index combines richness and evenness, but it is commonly expressed as its reciprocal (1-D) to ensure intuitive scaling with diversity. Without taking the reciprocal, D ranges from 0 (maximal diversity) to 1 (minimal diversity), which inversely scales with diversity. By using 1-D, higher values directly correspond to greater diversity.

1.1.2.5 Hill Numbers

Hill numbers provide a flexible framework for combining richness and evenness. This framework is comprised of major diversity indices incorporated into one equation, species richness, exponential of Shannon entropy and the reciprocal of Simpon's index. The core axiom of the Hill number framework is the calculation of the effective number of species. This effective number is the number of equally abundant species required to equate to the diversity composition of community assessed. The benefit of effective number creation is the cross-comparative, sampling effort not withstanding, ability this number provides as the diversity value produces is always this effective number.

This effective number is calculated on a sensitivity scale, by allowing weighted adjustment of common to rare species through a sensitivity parameter q:

$$ {}^{q}D = \left(\sum_{i=1}^{S} {p_i}^{q} \right)^{\frac{1}{(1-q)}} \tag{6} $$

Where:

- q : sensitivity parameter
- S : species richness
- p_i : relative abundance of species i

Specific cases of Hill numbers correspond to well-known diversity metrics:

- ${}^{0}\mathrm{D} = S$: Species richness

$$ {}^{0}D = \left(\sum_{i=1}^{S} {p_i}^{0} \right)^{\frac{1}{(1-0)}} = \left(\sum_{i=1}^{S} 1 \right)^{1} = S \tag{7} $$

When $q = 0$, each relative abundance value (p_i) is equal to 1, as such when the sum is taken of these values the result is the number of species assessed using this sensitivity of q parameter. This is the maximally sensitive parameter where each species is considered equally irrespective of the evenness, and the rarity or commonality of the species therein.

- ${}^{1}\mathrm{D} = \mathrm{e}^{\mathrm{H}}$: Exponential of Shannon entropy

$$ {}^{1}D = exp\left(-\sum_{i=1}^{S} p_i \ln p_i \right) = e^{H} \tag{8} $$

When $q = 1$, this is a special case in the Hill number series as the power scaling component of the equation where 1 is divided by 1-q becomes 1 divided by 0. This indeterminate special case results in the equation not being able to be calculated when $q = 1$, breaking the series. As the resulting exponent becomes $1 \pm \infty$ and the base becomes 1, now the equation is $1^{\pm\infty}$. The solution to this is to calculate the diversity value as it approaches the limit of the Hill number equation as $q = 1$. This is performed using the L'Hôpital's rule to calculate the quotient.

$$ln\left({}^{q}D\right) = \frac{1}{1-q}\, ln\left(\sum_{i=1}^{S} {p_i}^{q}\right) \tag{9}$$

Let $f(q) = \ln\left(\sum_{i=1}^{S} {p_i}^{q}\right)$ and $g(q) = 1 - q$.

As $q \to 1$:

- $f(q) \to ln(1) = 0$
- $g(q) \to 0$

Which makes the indeterminate form of 0 divided by 0 where L'Hôpital's rule can be applied. Where the denominator is;

$$g'(q) = \frac{d}{dq}(1-q) = -1 \tag{10}$$

And the numerator is;

$$f'(q) = \frac{1}{\sum_i \ p_i^q} \times \frac{d}{dq} \sum_i p_i^q \tag{11}$$

As $\frac{d}{dq} p_i^q = \ p_i^q \ln p_i$:

Then;

$$f'(q) = \frac{\sum_i \ p_i^q \ln p_i}{\sum_i \ p_i^q} \tag{12}$$

Here we apply the L'Hôpital's rule:

$$\lim_{q \to 1} \frac{f(q)}{g(q)} = \lim_{q \to 1} \frac{f'(q)}{g'(q)} = \frac{\frac{\sum_i \ p_i^q \ln p_i}{\sum_i \ p_i^q}}{-1} \tag{13}$$

There by evaluation at $q = 1$, where $\sum_i p_i^1 = \ \sum_i p_i \ = 1$:

$$= \frac{\sum_i \ p_i \ln p_i}{1 \times (-1)} = \ - \sum_{i=1}^{S} p_i \ln p_i \tag{14}$$

Which as seen initially is the calculation of the Hill number when q = 1:

$$\lim_{q \to 1} ln\left({}^{q}D\right) = -\sum_{i=1}^{S} p_i \ln p_i = H_{Shannon} \tag{15}$$

And by exponentiating both sides, which in turn is the exponential of Shannon entropy:

$${}^{1}D = exp\left(-\sum_{i=1}^{S} p_i \ln p_i\right) = e^{H} \tag{16}$$

- ${}^{2}\mathrm{D} = 1/D$: Reciprocal of Simpson's index

$${}^{2}D = \left(\sum_{i=1}^{S} {p_i}^{q}\right)^{-1} = \frac{1}{\lambda} \tag{17}$$

Where $\lambda = \sum p_i^2$ is the Simpson index detailed earlier. In this case, as the sensitivity parameter q increases to the value of 2, the relative abundances are squared resulting in larger abundances being weighted more heavily (abundance value is kept closer to measured) while rare species with small relative abundance are reduced more heavily. Therefore, the effective number of species is weighted to be the effective number of species which are dominant in the community.

Hill numbers provide a unified and intuitive approach to diversity, with sensitivity (to relative abundances, rare species) parameter q controlling the emphasis on common versus rare species. As q is adjusted, Hill numbers produce a suite of diversity indices.

1.1.2.6 Leinster and Cobbold index (LCI)

This approach retains the flexibility of Hill numbers, allowing sensitivity adjustments to rare species while integrating an additional layer of ecological information. Notably, the LCI achieves this with minimal modifications to the standard Hill number calculation. The following equations detail the mathematical formulation of the index. Programmatically, equation 2 is calculated first, then equation 1 for all cases of q (rarity sensitivity parameter) except for the case of $q = 1$, where equation 3 is applied.

The Leinster and Cobbold index equation is as follows:

$${}^{q}D^{Z}(p) = \left(\sum p_i (\boldsymbol{Zp})_i^{q-1}\right)^{\frac{1}{1-q}} \tag{18}$$

Where,

$$(\boldsymbol{Zp})_i = \sum_{j=1}^{S} Z_{i_j} p_j. \tag{19}$$

When $q = 1$:

$$^{1}D^{Z}(p) = \frac{1}{(\boldsymbol{Zp})_1^{p_1}}, \frac{1}{(\boldsymbol{Zp})_2^{p_2}}, \ldots, \frac{1}{(\boldsymbol{Zp})_S^{p_S}} \quad (20)$$

Detailed in the equations above is the similarity matrix (Z_{ij}), where i is the i^{th} row and j is the j^{th} row in matrix Z in equation 2 and the relative abundances of each species in the community (p). Following equation 2, the product is then used in equation 1 (Zp) which then utilises the community relative abundances a second time and is weighted by the sensitivity parameter (q). The mathematical details of how this index works will first be described mathematically and then programmatically for the LDG LCI assessments (section 3.2.2).

As Leinster and Cobbold outline in their paper, for a community of species (*number of species, S*), the individual relative abundance *p1...pS* of each species is calculated. Following this, I assured that $p_i \geq 0$ and the sum of $p_i = 1$. The $S \times S$ similarity matrix $Z = (Z_{ij})$ is composed of measured similarities between the i^{th} and j^{th} species. This ensures every species of the dataset is compared to every other species of the dataset. Each taxonomic level (genus, family, order, class, phylum) match results in an insertion of a value of 1, if there is no match a 0 is added. When all taxonomic levels equal a match, the total is equal to the number of taxonomic levels included in the similarity assessment. This value is then divided by the number of comparisons made per species. This results in a value of 1 for species identical and a value between 0 and 1 for species with a relative level of similarity. Similarity is defined by the user e.g. taxonomic, functional or genetic, for each species through the construction of a similarity matrix. The integration of the sensitivity parameter q such that $0 \geq q \geq \infty$ provides a weighted change to the output of the index $^{q}D^{Z}(p)$. When q is small, the output is affected in almost equal proportion by both rare and common species. In the instances when q is large, the output is largely unaffected by rare species. The inclusion of this q parameter satisfies the mathematical statements outlined in (Chao, Chiu, et al., 2014; Hill, 1973b; Jost, 2010).

The similarity matrix (Z_{ij}) is then used to weight the relative abundances of the community of the area of assessment (AOA). The dot product, where Z_{ij} and p_j are multiplied which then becomes Zp as noted in part two of equation 1. This vector product of length *Nsp* is a measure of species abundances, weighted by the relative similarity of the community. The maximal similarity sensitive abundances are met when all species are fully dissimilar as seen in the naïve case, resulting in species richness. The vector Zp is then scaled by the *q-1,* where q is the sensitivity parameter which adjusts the diversity measurements sensitivity to rare species. When $q = 0$, the index is equally sensitive to rare and common species. As q is increased the diversity measurement is more representative of the relatively more common species in the dataset. The reason for the deviation from the standard Hill number framework equation is the incorporation of similarity *Z.* Due to this weighting procedure and to maintain the Hill number functionality within the LCI when naïve similarity matrix is used, the q is reduced by 1. After the Zp is scaled by the sensitivity value, the vector is then rescaled by the relative abundances to ensure alignment with Hill numbers weighting.

This vector is then summed, and the resulting value is then scaled by 1/1-q in line with the Hill number framework. In the naïve case, i.e. Hill numbers, when q = 0, this value is richness, the similarity case is a similarity-sensitive version of richness. As q is increased the diversity value closer represents the value of the more common species though the generalised weighted mean

procedure used by both the Hill number and the LCI procedure. In all assessments, per AOA, I assess the diversity of the community at *q0 – q5* in steps of 1.

1.1.2.7 Chen and Grinfeld decomposition of Leinster and Cobbold Index

Chen and Grinfeld (2024) have devised a decomposition which supersedes the existing work of Alje van Dam who performed a similar piece of work decomposing Hill numbers into three diversity components (Van Dam, 2019). Their approach advocates that the Van Dam decomposition presents an asymmetrically biased version of balance (i.e. evenness) as it does not include the similarity component. Chen and Grinfeld incorporate similarity into all decomposed diversity components. They highlight, when similarity is included the axiom which states that diversity is maximised when evenness (i.e. balance) is fully even when each species is equal is not being met in Van Dam's case.

To maximise evenness when similarity is considered, Chen and Grinfeld include the vector p^*. Whereby solving a system of linear equations, you can produce a vector which maximises relative abundance per species while considering the similarity of the community. Therefore this vector, p^*, does not necessarily follow the maximally even case, p_h, but can do depending on the similarity of community, i.e. if the similarity matrix, Z, is equal to the full dissimilar case or assumption in the naïve LCI approach (i.e. Hill numbers) which would be an identity matrix (I_n). Chen and Grinfeld show this p^* leads to greater values than the traditional evenness (balance) vector of relative abundances(Chen & Grinfeld, 2024).

The mechanics of this decomposition and how I implemented it into R for the purpose of the global assessment of marine phytoplankton diversity follow the same order of operations of the Leinster and Cobbold analysis. The key difference in calculation of the decomposed LCI is the calculation of p^* and p_h. Therefore, the decomposed values are produced per latitude. Future work will establish the fine-scale grain decomposed diversity assessment.

The following equations are sourced from Chen and Grinfeld paper. For consistency with their work, I will use balance to describe the evenness of the community. Balance is preferred in this case as balance is more representative of finding the maximal balance over evenness which only refers to full equal abundances (Chen & Grinfeld, 2024).

Full community decomposed LCI per AOA:

$$F(p,Z,q) = \sqrt{\frac{F(p,Z,q)}{F(p^*,Z,q)}} \cdot \sqrt{\frac{F(p,Z,q)}{F(p,I_n,q)}} \cdot \sqrt{F(p^*,Z,q) \cdot F(p,I_n,q)}, \tag{21}$$

Decomposed community balance per AOA:

$$B(p,Z,q) \coloneqq \sqrt{\frac{F(p,Z,q)}{F(p^*,Z,q)}}, \tag{22}$$

Decomposed community dissimilarity per AOA:

$$D(p, Z, q) \coloneqq \sqrt{\frac{F(p, Z, q)}{F(p, I_n, q)}}, \tag{23}$$

Decomposed community richness per AOA:

$$R(p, Z, q) \coloneqq \sqrt{F(p^*, Z, q) \cdot F(p, I_n, q)}, \tag{24}$$

This describes the decomposition mechanics for the LCI. To calculate the LCI by latitude, the observational data is assessed for number of unique species and relative abundances using the previously outlined LCI approach. The difference here, is that each LC function is calculated independently with required combinations of the relative abundance vector (p, p^*, p_h), similarity matrix ($Zsim$ or I_n which is the identity matrix). As shown in the above equation, the original LCI is retrievable by multiplicative action of each component, increasing interoperability with the existing approach.

1.1.3 Qualitative assessment of a similarity sensitive diversity index

1.1.3.1 Introduction

The choice of diversity index, the locations assessed for diversity, and the effect of sampling effort are important facets of diversity research studies (Morris et al., 2014) to be considered when drawing conclusions and developing theories pertaining to the ontology of diversity. In this introductory section, I will explain the rationale for our choice of index and detail simple quantitative results obtained when assessing the behaviour of the similarity-sensitive index with a series of hypothetical communities. This will provide the necessary understanding from which to interpret the results of the following research chapters.

The aim of this hypothetical community work is to give the reader an intuitive understanding of the effect of integrating similarity into diversity approaches which typically treat diversity as species richness or species richness and evenness naively without considering species similarity. My approach is focused on the application of the Leinster and Cobbold diversity index and the diversity components therein, for a wider scope of the assessment of the Leinster and Cobbold parametric index in the context of more traditional univariate diversity measures, Daly et al. 2018 examine this aspect in greater detail. The research in this section is to specifically visually examine the effect of adjusting the three components of diversity: richness, evenness and species similarity.

1.1.3.2 Methods

To address the aim of the hypothetical community work, I created various communities with varied diversity components properties. I did this by creating three spectra of relative abundances to explore the effect of relative abundance on a similarity sensitive diversity index.

I also created three scenarios with varying degrees of taxonomic similarity to examine the effect of this integrated similarity approach in different communities with varying relative abundances. The benefit of this approach instead of examining real world biological community data is the experimental control of the community sample, where the majority of the hypothetical communities have a fixed richness of 5 species. This is with the exception of the community scenarios with a 'missing' rare fifth species, termed the 'Missed species scenario'. This provides a comparative interpretation of the similarity-sensitive diversity index when richness is adjusted.

For the following hypothesised communities, the first diversity component, richness, was fixed at 5 theoretical species in all cases except for the missed species scenario. This scenario represents scenarios where rare species are 'missed' i.e. not sampled in a sampling regime. The inclusion of a missed species scenario was motivated by the desire to assess how the output of the Leinster and Cobbold index behaves comparatively to other communities where richness differs.

The second diversity component, evenness, was represented as a scale of uneven to maximally even i.e. equally balanced species abundances. An uneven community is characterised by the dominance i.e. high abundance, of one or two species with other species of the community being rare with low abundances comparatively (Fisher et al., 1943).

Table 1 - Theoretical community set-up for LC index behavioural analysis.

Theoretical Community	Scenario	Relative abundance distribution for 5/4 species communities	Number of Individuals
Moderately similar taxonomy	Dominant abundance	6,1,1,1,1	10
	Even abundance	2,2,2,2,2	10
	Intermediate abundance	3,2,2,2,1	10
	Missed species scenario	3,2,2,2	9
Equivalent taxonomy	Dominant abundance	6,1,1,1,1	10
	Even abundance	2,2,2,2,2	10
	Intermediate abundance	3,2,2,2,1	10
	Missed species scenario	3,2,2,2	9
Dissimilar Taxonomy	Dominant abundance	6,1,1,1,1	10
	Even abundance	2,2,2,2,2	10
	Intermediate abundance	3,2,2,2,1	10
	Missed species scenario	3,2,2,2	9

I represented this through the creation of four sets of relative abundances (Table 1). The dominant scenario is representative of the ecological scenario where one species is much more prevalent than all other species. This can be analogous to phytoplankton blooming conditions where one or two species bloom, leading to high occurrences of these species, out competing other species and creating very steeply declining relative abundance curves. The second scenario is the intermediate scenario where one species is not especially dominant but does have the highest abundance and there are some even abundances in the middle of the distribution, followed by one rare species.

The third relative abundance scenario is the ecologically missed scenario which adopts the same abundance distribution of the realistic scenario, but the fifth rare species is not 'observed' in the community. The last scenario is the case, although highly improbable, where all species observed are equally abundant. This provides a comparative scenario to measure the behaviour of the LC index when abundances are equivalent thus nullifying the effect of abundance in the output of the index.

Table 2 - Synthetic taxonomic trees for three theoretical communities. Taxonomic similarity distribution for each taxonomic case, highlighted cells indicate matches in taxonomy. Colours; blue, pink and orange, indicate matching taxonomy of species within each scenario.

Theoretical Community	Species Name	Species	Genus	Family	Order	Class	Phylum	Kingdom	Domain
Moderately similar taxonomy	Species 1	40	30	20	15	8	4	7	1
	Species 2	41	30	20	15	8	4	7	1
	Species 3	42	32	21	16	10	3	7	1
	Species 4	43	33	22	17	10	3	7	1
	Species 5	44	34	23	19	12	5	7	1
Equivalent taxonomy	Species 1	40	30	20	15	8	4	2	1
	Species 2	40	30	20	15	8	4	2	1
	Species 3	40	30	20	15	8	4	2	1
	Species 4	40	30	20	15	8	4	2	1
	Species 5	40	30	20	15	8	4	2	1
Dissimilar Taxonomy	Species 1	40	30	20	15	8	4	2	1
	Species 2	41	31	21	16	9	5	3	2
	Species 3	42	32	22	17	10	6	4	3
	Species 4	43	33	23	18	11	7	5	4
	Species 5	44	34	24	19	12	8	6	5

The third diversity component is the taxonomic diversity of the communities to explore the effect of taxonomic similarity three scenarios were developed to explore interactions varying richness and evenness on the output of the similarity-sensitive diversity index. For this hypothetical experiment I conceptualised similarity to have a similar behaviour to that of relative abundance where similarity is a continuous spectrum within the community assessed. The limits of this spectrum are fully similar or fully dissimilar, the prior describes where all taxonomy is equivalent which describes one species as present. The other end of the spectrum is maximally different at all taxonomic levels which describes 5 or 4, scenario dependant, unique species.

The three taxonomic scenarios are as follows: fully dissimilar, equivalent and moderate (Table 2). The fully dissimilar scenario (Table 2, Dissimilar Taxonomy) represents a collection of species with no taxonomic similarity, i.e. no taxonomic relatedness. The moderately similar scenario (Table 2, Moderately similar Taxonomy) showcases a sampling scenario where species 1 and 2 are of the same genus and as a result have high taxonomic similarity. Species 3 and 4 are from the same class but diverge in order to genus, these species are related to species 1 and 2 and to each other. Species 5 is taxonomically distinct from species 1 to 4. This sampling scenario is moderate and acts as an intermediary taxonomic similarity case compared to equivalent taxonomy scenario. In the equivalent taxonomy scenario (Table 2, Equivalent Taxonomy) species 1 to 5 have equal similarity for the taxonomic ranks assessed.

Here I have chosen the Leinster and Cobbold index to assess the diversity of the hypothetical communities. This index presents multiple advantages with the main one being the inclusion of similarity attribute weighting in the non-naïve diversity calculation. In addition, the index is an effective number in both the naïve and non-naïve cases for diversity of order q. This means that 18 totally disparate species will have an effective species number of 18. Furthermore, modulation of the weighting parameter q adjusts the relative importance of common and rare species via the abundance measure.

1.1.3.2.1 Fully dissimilar taxonomy hypothetical community scenario

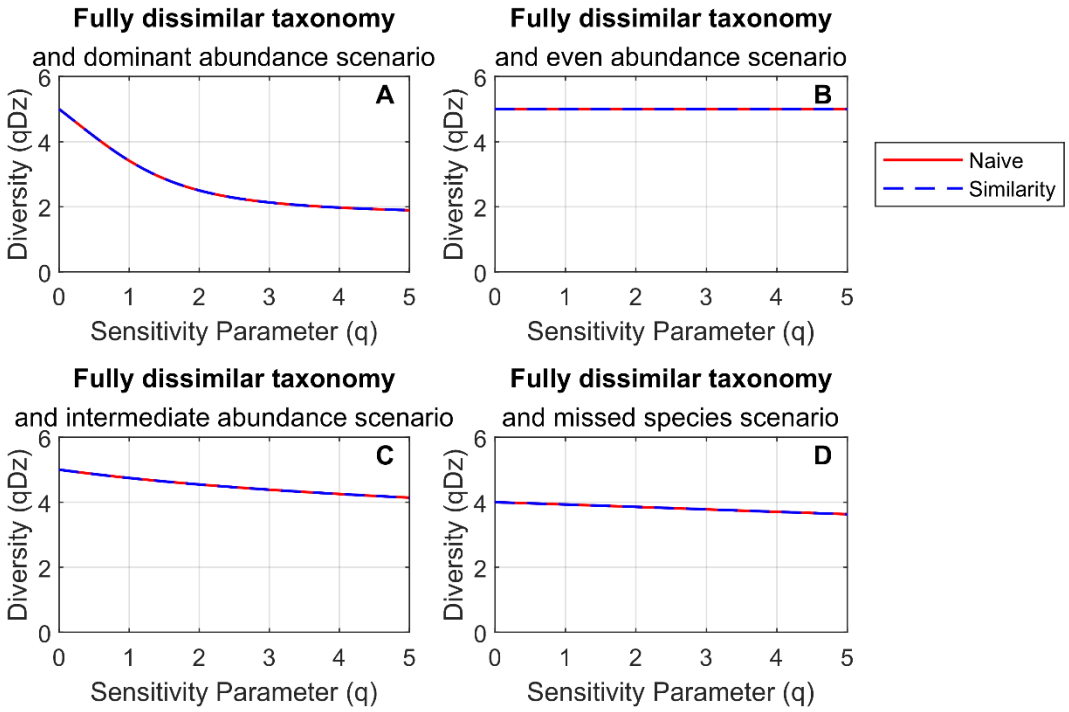


Figure 1.3 - Fully dissimilar taxonomy scenario. For all fully dissimilar taxonomic scenarios, the red line is the naïve case (richness and evenness) and the dashed blue line is the similarity (richness, evenness and similarity).

In the fully dissimilar scenario, richness was fixed at five species scenarios labelled A to C, with 4 species in the fourth scenario (Figure 1.3D). The evenness of the hypothetical community scenarios was adjusted as detailed in Table 1. The third component, taxonomic similarity, was fully dissimilar for each species, representing a community of species with no taxonomic relatedness. This approach tells us that that the similarity-sensitive case of the Leinster and Cobbold index is equivalent to the naïve case when each species is distinct as expected (Leinster & Cobbold, 2012a). The diversity profile of this parametric index (both naïve and similarity-sensitive) in each community changes depending on the relative abundances of the species therein. In the dominant abundance scenario, the diversity profile rapidly declines compared to the other communities as a result of the species abundance distribution (SAD). In this case (Figure 1.3A) the majority of individuals are of species 1 (6 individuals) and the remaining four species each have an abundance of (1 individual per species). The diversity of the community declines as the sensitivity parameter (q) increases, as the index is gradually weighted towards the most abundant species, the diversity of the community is 1.89 species effectively at $q = 5$.

When the SAD is even (Figure 1.3B), with each of the five species having two individuals, the q parameter has no weighted effect on the diversity of the community. As the taxonomy is also equally distinct in this scenario (Figure 1.3B) the effective number of species for all cases of q is 5 species. This is the case for both the naïve case and the similarity-sensitive case of the index. Whereas when the SAD is less characterised by a dominant species in the intermediate case (Table 1), the profile does not decline as rapidly as observed in the dominant scenario (Figure 1.3C). The diversity profile shows that the diversity of this community is higher than that of the dominant hypothetical community (Figure 1.3A). The missed species scenario has a SAD of greater evenness than the intermediate case due to the missing rare species with an individual count of one. As a result, the overall diversity is lower due to the richness of this

hypothetical community being 4 species rather than 5 at $q = 0$, compared to the intermediate case. Yet, the decline in diversity as q is increased is less than that of the intermediate case.

1.1.3.2.2 Fully equivalent taxonomy hypothetical community scenario

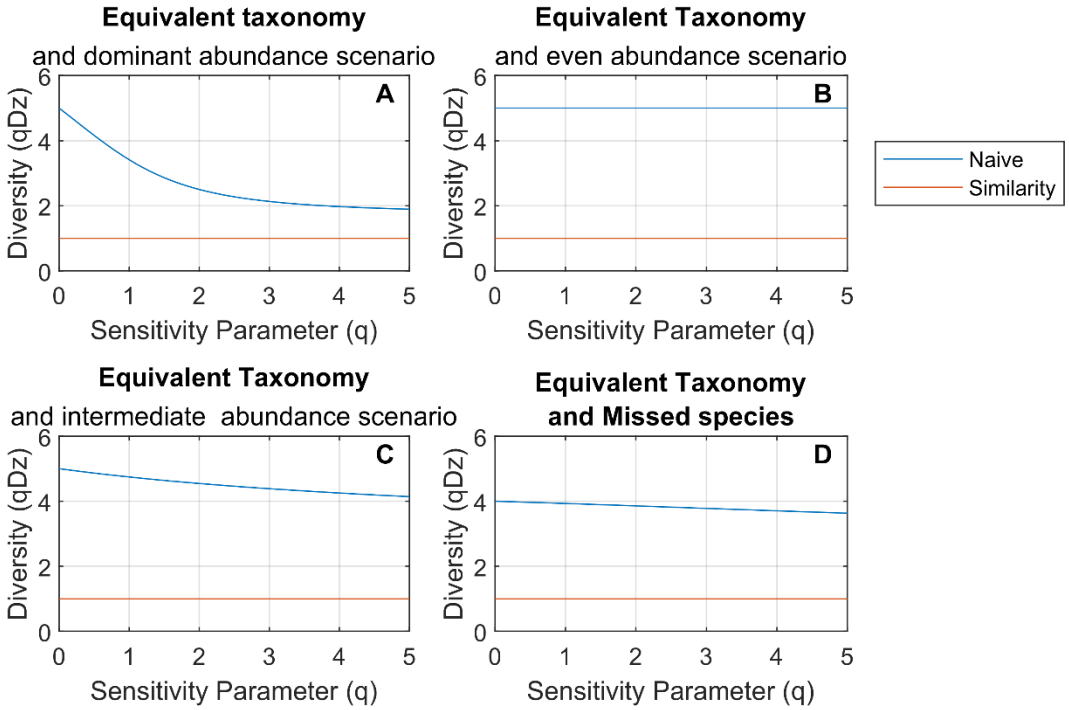


Figure 1.4 - Equivalent taxonomy scenario. For all equivalent taxonomic scenarios, the blue line is the naïve case (richness and evenness) and the orange line is the similarity-sensitive case (richness, evenness and similarity).

For the following scenarios presented in Figure 1.5 and Figure 1.5, the naïve case of the index has the exact same properties and is equivalent in quantity as the scenarios in Figure 1.4. The results from here on will be concerned with the properties of the similarity-sensitive case of the Leinster and Cobbold index across the different hypothetical scenarios. When the taxonomic linage of each sampled species is the same (Table 2, Equivalent taxonomy), the resulting similarity-sensitive index output for all communities (Figure 1.5A-D) irrespective of richness and abundance distribution correctly details there is effectively one species.

1.1.3.2.3 Moderately similar taxonomy hypothetical community scenario

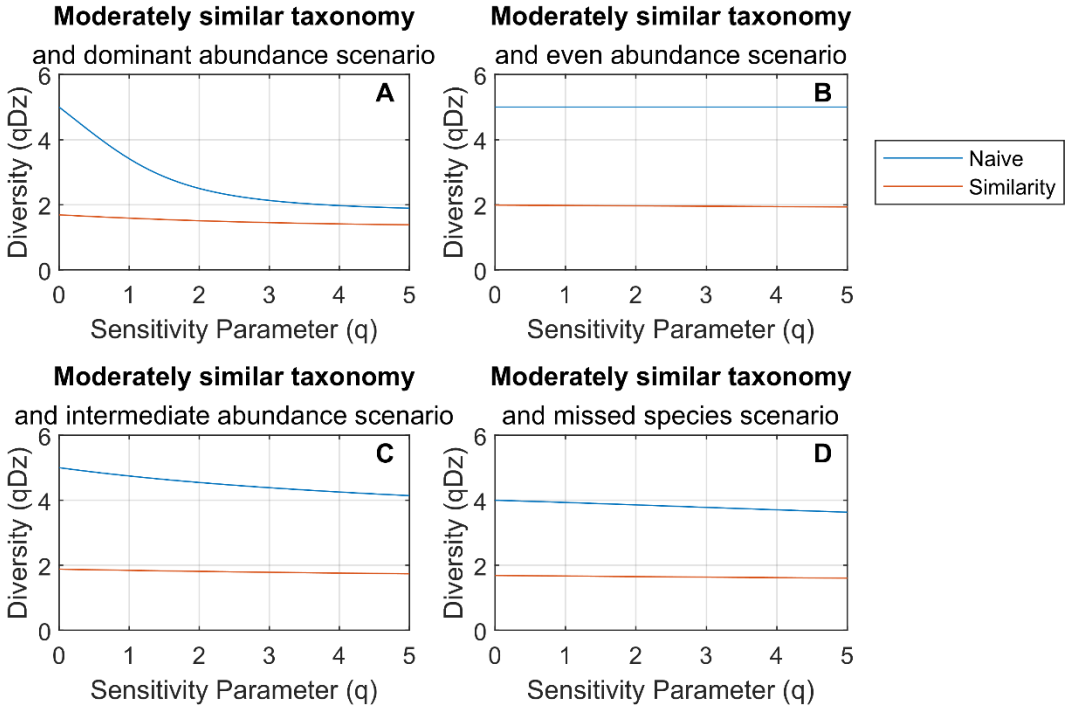


Figure 1.5 - Moderately similar taxonomy scenario. For all Moderately similar taxonomic scenarios, the blue line is the naïve case (richness and evenness) and the orange line is the similarity (richness, evenness and similarity).

In the moderately similar case (Figure 1.5A-D) the similarity-sensitive case of the Leinster and Cobbold index shows greater community diversity in all communities than that observed in the equivalent taxonomic case. Here, the taxonomic similarity is less than that of the equivalent scenario but is greater than the fully dissimilar scenario. This result highlights that the inclusion of taxonomic similarity results in a scaled diversity profile which has an absolute minimum of 1 species (Figure 1.5A-D) to the 5 species maxima of the naïve case (Figure 1.4A-D) depending on the similarity of the community. This inference is true for the taxonomic case, but the inference is mutable by the researcher definition of similarity included in the index.

1.1.4 Explanatory theories for macroecological patterns

Macroecology is a subfield of ecology that examines large-scale patterns and processes governing biodiversity, species distributions, and ecosystem structure (Worm & Tittensor, 2018). It assumes that fundamental ecological principles, such as metabolic scaling and body-size relationships, apply broadly across species and serve as generalisable mechanisms for explaining diversity and productivity across multiple ecosystems (Keith et al., 2012). Unlike localised ecology, which often focuses on localised species interactions, macroecology seeks to develop predictive and explanatory diversity frameworks that extend across spatial and temporal scales, offering insights into global biodiversity trends and mechanisms pertaining to the existence of large-scale patterns of diversity.

A major focus of macroecological research is the latitudinal diversity gradient (LDG). The key feature of the LDG is an observed pattern across fauna and flora where species richness is greatest at or near the Earths equatorial region with a declining gradient in observed species richness towards the polar regions. Research focused on the LDG is concerned with mechanisms by which this pattern arises and why it appears to be taxonomically ignorant. The strength of this pattern has increased and decreased over geological time periods. In cooler

periods of climate, the LDG is represented by a unimodal i.e. singular peak of diversity on or near the equator line, in warmer periods of climate the peak weakens and a bimodal i.e. two peak pattern of relatively high species richness is observed on either side (north and south) of the equator line in the sub-tropics (Beaugrand et al., 2015; Yasuhara et al., 2020). The modality of the LDG pattern i.e. unimodal or bimodal can differ across different groups.

While the general pattern is a widely observed feature of the diversity of life on Earth, there are biological groups which are exceptions to this general rule, such as groups where evolutionary adaption has led to greater diversity in latitudes closer to the polar regions, e.g. pinnipeds (Tittensor et al., 2010a). The existence of inter-group differences in the relationship to this pattern exemplifies the complexity of LDG research. Evolution is universal to all life and the LDG pattern has a long geological history of existence while biological evolutionary adaption has been ongoing during the same time. It then raises the question of, if some groups are evolutionary adapted to not have high species richness around the equator, why have other groups or taxa not followed suit.

Therefore, the above being true, diversity of life on Earth must have a general spatial mechanism(s) on a macroecological scale which generally leads to this macroecological pattern of diversity of life on Earth arising (Rohde, 1992; Velasco & Pinto-Ledezma, 2022). The four core mechanistic components which quantify diversity; birth, death, speciation and mutation must be generally spatially influenced on Earth (Saupe, 2023). The mechanisms by which this spatially occurs are still debated and discussed at present due to the complexity of interacting factors which influence diversity of life on Earth. Evolutionary adaption i.e. where through birth and random mutations lead to selection of traits which increase likelihood of reproduction and survival in a given location does not occur in a biological vacuum. This adaptive trait selection is driven by the ability of a trait to either have no effect on or increase the access to resources required for growth and survival (Wadgymar et al., 2022). For example, in no hierarchal order, available habitat space for reduced inter- and intra- species competition, which when high leads to competitive exclusion, food availability required for growth be that prey, nutrients and sunlight, access to a tolerable physical environment which does not result in exhaustive internal biological resource usage.

While biological adaption does not occur in a biological vacuum i.e. invulnerable to physical and chemical requirements for life, biological competition and survival are core to the reproductive succession of new species via the selection of mutated traits. This selection process occurs within populations of species and is influenced by the network of species which co-inhabit the same habitat as others. Resources for growth are typically limited; therefore, the maximum size of any species population is a function of access to resources for growth (Rosenzweig, 1995). If no resources remain, a population cannot sustain new births to maintain the population. In addition, access to resources is mediated by other populations of species also vying for the same resources. What has not been discussed when speaking of resources, is what these resources are and how they are dependent on the trophic level of the species being observed.

Primary producers typically require sunlight for photosynthesis i.e. phytoplankton at the sea surface, or biotically available molecules for chemosynthesis i.e. giant tube worms at hydrothermal vents. Herbivores are consumers of primary producer species such as rabbits and grasses on land, or zooplankton predation of phytoplankton in the ocean. The next trophic levels are dependant on carnivorous predation of other species to survive and grow, such as the bald eagle and fish predation. The requirement of resources is also dependant on the body size and metabolism of a species, larger species with high metabolism require a greater quantity of resource to maintain body growth and sustained function (Kaiser & Williams, 2011).

Exceptions to this rule are found in species occurring in low temperature regimes of the deep sea, where higher oxygen availability due to the low ambient sea temperature (around 2 deg C) and slower metabolisms because of this reduced temperature combine to result in long living species with reduced resource requirements to sustain their slow metabolisms such as the Greenland shark (Nielsen et al., 2016).

As noted above, resource availability, size of an individual, where that individual inhabits and the traits evolved over time influence the survival of a species. Species survival is ultimately what is represented by diversity assessments of extant species today. Another key feature of Earth, our blue planet, is the two major biomes represented by terrestrial i.e. land above sea level and the marine, the global ocean of Earth. These two biomes cover the planet entirely in different spatial quantities and have different constraining factors on the survival of a species.

These biomes do differ in constraints which inevitability lead to different suites of survival strategies. In no hierarchal order of importance, I will now detail some of these key differences. First is that of spatial area, the ocean covers around 70% of the Earths surface, while land covers the remaining 30%, therefore species residing in the ocean have a greater quantity of spatial area as a resource. While on land, species are constrained by land area coverage and requirement for air to breathe. Reduction in available area typically leads to higher competitive capacity for acquisition of habitat for species using the same habitat in similar behavioural or morphological ways. In additions, species capable of flight or capable of climbing have a reduced vertical constraint, most other terrestrial species are vertically challenged due to being unable to fly. While comparatively, in the ocean, its vertical depth ois traversable given the fluid nature of water and swimming ability. Although there are constraints to vertical movement in the ocean due to the increase of pressure at greater depths, which without adaption prevent high vertical movement capacity of species. In addition, the variability of temperature is greater on land than in the ocean comparatively, where sea temperature varies by 30 deg C at the equator to -2C at the poles, land temperatures can range from -50C to 40C typically (Deser et al., 2010; Y.-R. Wang et al., 2022). Despite these key differences, the LDG pattern is observed in both systems (Hillebrand, 2004). As noted, the role of temperature and metabolism is consistent for all life on Earth, for endothermic and exothermic species.

The Metabolic Theory of Ecology (MTE) offers a physiological perspective on macroecological patterns, proposing that metabolic rates, which scale predictably with body size and temperature (Gillooly et al., 2001), are the primary drivers of diversification (Brown, 2014; Brown et al., 2004). As temperature increases, the kinetic energy required to enable metabolic chemical reactions is increased, as such the speed of these reactions increase to a point of optimum (Chen, 2015). This connects the LDG via the gradient of temperature peaking at the equator with minimum temperatures in polar regions. It is theorised that the greater occurrence of optimal temperatures near the equator leads to faster metabolic rates, increasing probability of mutation (Brown et al., 2004). This leads to greater probability of mutation and thus diversification, it has been observed this differs between exothermic (increased effect) and endothermic species (reduced effect) (Machac et al., 2012). Although MTE has been widely applied in macroecology, its universal applicability remains a topic of discussion (Price et al., 2012), as ecological structure and neutral processes can also shape biodiversity in tandem with metabolic constraints (Tittensor & Worm, 2016).

Several key macroecological mechanisms have been identified to explain biodiversity distribution across space and time (Pianka, 1966; Rohde, 1992). These include Rapoport's Rule (Rapoport, 1982), the Mid-Domain Effect (MDE)(Colwell et al., 2004), and the Metabolic Theory of Ecology (MTE)(Brown et al., 2004), each of which provides a different perspective on species diversity and related distributions.

Rapoport's Rule suggests that species at higher latitudes tend to have larger geographic ranges than those in tropical regions. Proposed by Eduardo Rapoport in 1982, this rule is based on the idea that species in variable and extreme environments, such as temperate and polar regions, must adapt to a broader range of conditions, leading to wider distributions (Rapoport, 1982). In contrast, species in stable tropical environments tend to have smaller, more specialised ranges. While Rapoport's Rule has been widely studied in biogeography, it is not universally observed across all taxa and regions, as factors such as historical climate fluctuations, dispersal limitations, and species-specific traits influence its applicability (Gaston & Chown, 1999).

The Mid-Domain Effect (MDE) is a statistical model that explains species richness patterns through spatial constraints rather than environmental gradients. It suggests that when species' geographic ranges are randomly placed within a bounded domain, more overlap occurs in the center, leading to a peak in species richness (Colwell et al., 2004). This effect has been used to explain latitudinal diversity gradients, where species richness is highest at low-latitudes. While MDE provides a neutral, geometric explanation for biodiversity patterns, its significance relative to ecological and evolutionary drivers, such as climate and historical processes, remains debated (Currie & Kerr, 2008).

The diversity-productivity hypothesis proposes that greater energy availability, primarily in the form of solar radiation and ambient temperature is required to drive increased production of organic matter, sustaining larger population sizes (Wright, 1983). This elevated abundance of individuals is, in turn, hypothesised to increase the probability of in situ speciation and immigration, ultimately generating greater species diversity in high-energy environments. Marine phytoplankton, alongside other small planktonic organisms such as marine bacteria and zooplankton, are ecologically distinct in being free-floating and globally dispersed (Finlay, 2002). This cosmopolitan distribution means that phytoplankton draw upon a global pool of resources, of which solar irradiance is the primary driver of growth. Bacteria and zooplankton represent the complementary lower and upper trophic levels relative to phytoplankton, respectively, and each maintains a functionally distinct dependency on phytoplankton productivity (Kaiser & Williams, 2011).

Zooplankton graze directly on phytoplankton, and the organic matter released through this predation, including cell lysis products and sloppy feeding which provides substrate for bacterial growth. Phytoplankton are not, entirely independent of these trophic partners. Bacterial remineralisation of phytoplankton-derived organic matter regenerates limiting nutrients such as nitrogen and phosphorus, returning them to the dissolved pool where they become available to support further phytoplankton growth. Zooplankton predation accelerates the rate at which phytoplankton biomass enters this remineralisation pathway. The productivity of marine phytoplankton is therefore partially dependent upon, these trophic feedback loops operating at a global scale (Kaiser & Williams, 2011).

Yet a fundamental question persists, why do marine phytoplankton maintain such high species diversity and does this diversity follow the LDG pattern. According to the competitive exclusion principle, species competing for the same limiting resource cannot stably coexist; the superior competitor will exclude all others (Hardin, 1960). Applied to phytoplankton, which appear to compete for a relatively small set of common resources such as light, nitrogen, phosphorus, and trace micronutrients, the prediction would be a community dominated by few competitive specialists. Yet empirical observation reveals the opposite: hundreds of phytoplankton species co-occur within the same parcel of water, drawing on the same resource base (Hutchinson, 1961). This is what Hutchinson termed the "paradox of the plankton," and it remains one of the central unresolved problems in marine ecology. The resolution of this paradox is complicated by the considerable variability in phytoplankton cell size, resource

uptake kinetics, and growth strategies across species, all of which influence competitive outcomes under different resource regimes. Understanding the mechanisms by which such diversity is generated and maintained in the marine phytoplankton therefore requires an examination of how diversity-productivity relationships manifest across spatial gradients in resource availability, and how species-level traits mediate coexistence within those gradients.

The unified neutral theory of biodiversity and biogeography proposed by Hubbell, 2001 is a stark contrast to the deterministic and mechanical nature of ecological theory of the 20th century. The theory of neutrality in ecology is formed from the idea that the core components of diversity are random. Speciation, extinction, immigration and emigration are stochastic within communities and when modelled as such, population dynamics such as species turnover can be replicated (Hubbell, 2001). As previously reviewed here, macroecological theories are focused on isolating common mechanisms tied to biological and environmental covariates of diversity. It is theorised that due to the commonness of the latitudinal diversity gradient, there should be a common mechanism (Rohde, 1992).

Macroecology faces limitations, particularly in the availability and resolution of global biodiversity data. While local and regional studies often provide detailed species-level observations, macroecological approaches must generalise across vast spatial scales, sometimes relying on incomplete or biased datasets (Lawton, 1999). The challenge remains to integrate increasing levels of diversity complexity into macroecological studies while acknowledging these data limitations (Gaston & Blackburn, 1999). Advancing macroecological research requires bridging the gap between broad-scale biodiversity patterns and local ecological processes. By incorporating more refined diversity measures, accounting for species interactions, and improving data accessibility, macroecology can continue to refine its predictive capacity and contribute to our understanding of global biodiversity change.

1.1.5 Marine phytoplankton biogeography, what is known and what remains to be learned?

Despite their microscopic size, marine phytoplankton exhibit macroecological patterns, occupying nearly all aquatic domains from pole to pole. Their vast distribution, high diversity, and fundamental role in global biogeochemical cycles make them an ideal group for macroecological studies (Falkowski et al., 2004; Field et al., 1998). Phytoplankton drive primary production, form the base of marine food webs, and contribute to key ecosystem processes such as carbon sequestration through the biological pump. Understanding their macroecological patterns is essential for predicting ecosystem responses to environmental change.

While individual phytoplankton are small, ranging from less than a micrometer (e.g., *Prochlorococcus*) to over a millimeter (e.g., *Coscinodiscus* diatoms), their relative size variation spans several orders of magnitude, comparable to the difference between a dandelion and a redwood tree. This size diversity is crucial in determining trophic interactions, nutrient cycling, and energy transfer within marine ecosystems. Numerous studies have examined phytoplankton body size relationships to improve our understanding of carbon fluxes, particularly in relation to the biological pump and energy transfer from primary producers to higher trophic levels (Acevedo-Trejos et al., 2015a).

However, macroecological studies of marine phytoplankton have historically been limited by sparse data availability (Righetti et al., 2020). Advances in remote sensing, autonomous sampling, and genetic tools are now expanding research opportunities, allowing for broader-scale assessments of phytoplankton diversity, distribution, and ecosystem function (de Vargas et al., 2015; Pierella Karlusich et al., 2020).

Phytoplankton blooms are episodic events characterised by rapid population increases, in response to changes in light or nutrient availability driven by increased surface layer mixing (Kaiser & Williams, 2011). These blooms can occur naturally due to physical oceanographic processes such as increased stratification and light availability in the spring and upwelling, stratification breakdown, and winter mixing, which bring nutrients from deep waters to the surface. However, anthropogenic influences, such as agricultural runoff and coastal eutrophication, can also trigger blooms, sometimes leading to harmful algal blooms (HABs) with significant ecological and economic impacts (Sellner et al., 2003). Understanding the macroecological drivers of phytoplankton bloom dynamics and subsequent increases in productivity is critical for predicting ecosystem responses to climate change and human activities.

Despite their ecological importance, macroecological studies of marine phytoplankton remain fragmented, with limited global coverage (Righetti et al., 2020). Many datasets originate from oceanographic cruises (Brotas et al., 2023), which, while valuable, only capture a subset of phytoplankton distributions across latitudinal gradients. Satellite remote sensing provides broader-scale observations (Savtchenko et al., 2004), but it primarily detects surface chlorophyll concentrations rather than species composition or functional diversity.

Another challenge in phytoplankton macroecology arises from data acquisition disparity across different size scales (Richardson et al., 2006). The smallest phytoplankton, such as *Prochlorococcus* and *Synechococcus*, are nearly indistinguishable at the species level without genetic analysis. Current diversity assessments rely on species-level classification, but integrating alternative approaches, such as trait-based metrics, functional diversity indices, and evenness measures could provide a more comprehensive understanding of phytoplankton community structure and dynamics. Given the growing uncertainty posed by climate change, there is a high demand for predictive techniques that can assess phytoplankton responses to shifting environmental conditions (Tittensor et al., 2021). Advancing macroecological research on marine phytoplankton will require integrating novel modeling techniques, expanding high-resolution observational datasets, and bridging the gap between genetic, taxonomic, and functional diversity assessments.

1.1.6 Aim of thesis

The aim of this thesis is to assess the global diversity of marine phytoplankton and the effect of how diversity is quantified on macroecological assessments through index choice. This assessment is done in recognition of a limited observational dataset. Specifically, I examine the alignment of marine phytoplankton diversity to the well known macroecological pattern: the latitudinal diversity gradient. Traditionally, macroecological patterns are assessed using species richness, here we continue and expand on this traditional approach to assess global community evenness and similarity diversity components. Additionally, I evaluate the findings of the LDG assessment with respect to sampling biases in latitudinal data availability. This expansion of the traditional approach to macroecological research provides novel insights and informs future directions in the field of macroecology. Following research pertaining to the LDG of marine phytoplankton, the aim was to assess the capacity of an ensemble SDM to reconstruct the LDG of marine phytoplankton to resolve sampling effort bias and generate global remotely-sensed chlorophyll-a measurements. This aim was motivated by early results of the SDM research. These findings serve as a potential method for developing proxies of chlorophyll-a globally outwith the temporally limited operational satellite data collection.

1.1.6.1 Thesis synopsis

To achieve this aim, I use a global observational dataset and a suite of methodological approaches. The approaches range from observational data analysis using a suite of diversity indices (Chapter 2 – Hill numbers and Chapter 3 – Leinster and Cobbold diversity index) to statistical species distribution modelling (Chapter 4). Chapter 2 explores the LDG of marine phytoplankton by focussing on species richness and the effect of community evenness on marine macroecology assessments with sampling effort bias assessed with rarefaction. Chapter 3 expands on this work and integrates the community taxonomic similarity, exploring the effect of diversity quantification on macroecological assessments of diversity. In addition, chapter 3 applies a decomposition of the Leinster and Cobbold index to examine how individual components of diversity; richness, evenness and similarity change over latitude. Chapter 4 applies a novel ensemble modelling framework to assess model the LDG of marine phytoplankton, aiming to account for spatial sampling biases. In addition, the modelled outputs are then correlated to satellite chlorophyll-a measurements and in-situ measurements of the Atlantic Meridional Transect of the Atlantic Ocean.

I find that the role of sampling effort on the LDG of marine phytoplankton is more prominent in the Hill number analysis than that of the LCI approach, influencing the diversity pattern more in the Hill number assessment. The global diversity of marine phytoplankton will require further observational data collection to reach consensus on the pattern, particularly in the Pacific Ocean. In the LCI chapter, the inclusion of taxonomic information results in a shift of diversity peak to the southern hemisphere due higher taxonomic evenness and taxonomic dissimilarity. In the third research chapter, SDMs provide a potential solution to sampling effort bias through distributional modelling. This approach could not model the entire suite of the diversity contained in PhytoBase, an already limited collection of observations. The lack of ability to model rare species of marine phytoplankton presents an issue in this approach. Although through modelling the most abundant species, there was a correlative relationship found between these modelled outputs and mean chlorophyll-a of the global ocean. This presents a step forward to resolving community composition driving primary productivity using SDMs, which to date has been limited to generalised modelling approaches without modelling specific species.

2 Global diversity of marine phytoplankton using Hill numbers

2.1 Introduction

Marine phytoplankton play a foundational role in ocean ecosystems, contributing significantly to global carbon cycling and supporting marine food webs (Henson et al., 2021). Despite their ecological importance, their global diversity patterns remain understudied (Irigoien et al., 2004; Rodríguez-Ramos et al., 2015). Community structuring and subsequent diversity arising from immigration and emigration processes are not well characterised for marine phytoplankton (Chust et al., 2013; Clayton et al., 2013). While phytoplankton may exhibit exceptional dispersal capabilities, this in combination with high rates of turnover and potential for adaption leads to an unclear picture of their observed macroecological patterns (Dutkiewicz et al., 2020; Ward et al., 2021). This research approach also lays the foundation for future comparative analyses of mechanisms driving diversity by exploring how community organisation such as relative abundances shape diversity.

This chapter provides the first global assessment of marine phytoplankton diversity using the Hill number diversity index framework. Unlike traditional species richness metrics, the Hill number index incorporates evenness into diversity estimation, offering a more comprehensive measure that reflects both species richness and the relative abundances of individuals within communities (Hill, 1973b). This approach allows for a more nuanced exploration of the LDG across different spatial scales, revealing detailed insights into phytoplankton macroecology.

By addressing sampling biases through rarefaction of sampling effort. The aim of this chapter is to investigate how the combined effects of evenness and richness contribute to shaping macroecological patterns (Stirling & Wilsey, 2001; J. Wang et al., 2017). Additionally, examining the global distribution of marine phytoplankton diversity by exploring the impact of sampling effort on observed patterns (Chao, Gotelli, et al., 2014). By achieving these objectives, this chapter contributes to long-standing discussions about the modality of LDG patterns, particularly for small-sized species with potentially global dispersal capabilities (Chaudhary et al., 2016; Falkowski et al., 2004; Menegotto & Rangel, 2018). Additionally, this work highlights the importance of considering both abundance data and environmental gradients to construct a clearer picture of global marine phytoplankton diversity.

2.1.1 Global data availability for this study

This study focuses on the global diversity distribution of marine phytoplankton using the PhytoBase dataset (Righetti et al., 2020). The study area covers ocean regions spanning latitudes between 88.92°N and 78.04°S and encompasses diverse marine environments, ranging from coastal regions to pelagic systems. It also includes a variety of ecological settings, such as nutrient-rich upwelling regions, oligotrophic gyres, and polar seas (Reygondeau et al., 2013). The global scale of this work aims to provide a comprehensive understanding of phytoplankton diversity across different biogeographic zones and to capture the full spectrum of environmental gradients that influence phytoplankton communities (Grattepanche et al., 2022; Strauss et al., 2023; Xu et al., 2022).

The temporal scope of this study spans nearly two centuries, from July 2, 1817, to May 26, 2014, with data gathered by researchers worldwide. However, this time span is not consistent across all sampled locations, which highlights a key challenge in studies reliant on data compilations: inconsistencies in sampling periods can introduce biases, particularly in groups like marine phytoplankton, where populations are strongly influenced by seasonal dynamics

such as phytoplankton blooms influencing community structure (Gonçalves-Araujo et al., 2016; Sukhanova et al., 2009). To address this challenge, this study assumes that the variability in diversity potentially arising from these temporal sampling inconsistencies is minimised when all samples are treated as a single time point and aggregated spatially (Chaudhary et al., 2021). The justification for this approach comes from the high sampling bias of this dataset, as there is high sampling in some regions at not others, and the time of sampling is biased also where research cruises are biased to April – October for suitable weather conditions. Therefore, to maximise the spatial coverage of the dataset I felt it was beneficial to approach the analysis aseasonally and apply a temporal aggregation to increase the spatial coverage. This was informed by the long-standing existence of the LDG pattern in geological time even though the strength of this pattern has been observed to strengthen and weaken over long time periods. The temporal range of this dataset was treated as a snapshot of the LDG pattern (150 years or so) is very short compared to the thousands of years required for significant recorded change of the LDG spatially.

A limitation of this assumption is that I only examine the spatial range of environmental covariates where individuals were observed, while a seasonally sensitive approach could examine a wider range of environmental tolerances. The latter could be true in the seasonal approach as a species may be observed in a wider range of environmental covariates in one location over time, that it is observed and related to spatial covariates. In addition, this assumption assumes that the LDG is generally static in the order of hundred years temporal range, for phytoplankton this is unknown to be the case or not. As there as been no consensus on the general LDG pattern for phytoplankton, it is not known how static the LDG is for this group of species. As has been evidenced, there can be large seasonal modifications to the location of maximum diversity of marine bacterial groups, it is not known if this is the case for marine phytoplankton. The analytical approach taken here will not resolve this but it is an area for further research provided a solution to the seasonal disparity of sampling can be addressed. This approach instead aggregates the variability to the area assessed in a non-seasonal manner, explicitly assuming that each species is observed within its realised niche rather than as a measure of its fundamental niche (Guisan et al., 2017), which a seasonal approach be closer to approaching dependant on the variable range of environmental covariates experienced by a species seasonally in said location(s). Therefore, this study does not attempt to assess marine phytoplankton diversity through seasonal periods so the reduction of observations to an aseasonal aggregate was treated as appropriate.

2.1.2 Global data bias

The PhytoBase dataset exhibits two main biases that need to be considered for this study: spatial coverage of samples and sampling intensity. The coverage of samples is biased to the North Atlantic region with 44.7% (608,197.587 samples) of the total samples (1,267,490 post data cleaning processing) occurring here while only 1.4% of samples originate in the South Atlantic (19,048.694 samples)(Righetti et al., 2020). These values also represent the disparity in sampling intensity between northern and southern hemispheres. The Atlantic Ocean region has the highest sampling coverage and intensity as a result (Table 3, *NTemperate*).

Table 3 - General statistics describing the latitudinal sampling coverage of the PhytoBase dataset.

General Statistics per Region					
Region	**Observations**	**Richness**	**Rate of detection (Observations per species)**	**Number of latitudes with data**	**Latitudinal range**
Global	1267490	1591	796.66	167	88.5 to -78.5
Finer latitudinal scale breakdown					
NPolar	148073	401	369.26	28	88.5 to 60
NTemp	528595	951	555.83	26	60 to 34
NSubT	86841	826	105.13	11	34 to 23.27
Tropical	461788	1065	433.60	48	23.27 to -23.27
SSubT	11147	601	18.55	11	-23.27 to -34
STemp	21999	616	35.71	26	-34 to -60
SPolar	9493	248	38.28	20	-60 to -78.5
Coarser latitudinal scale breakdown					
NTemperate	580315	759	764.58	40	80 to 40
NTropical	303082	1276	237.53	40	40 to 0
STropical	351861	846	415.91	40	0 to -40
STemperate	27246	542	50.27	39	-40 to -80

From calculation of the rate of detection i.e. the number of observations per species within the area delineation, interesting insights can be discerned. The rate highlights that areas with larger observation counts require more observations on average to detect new species in those areas such as the northern temperate and polar zone (Coarser area - NTemperate) requiring 765 samples on average per species while the southern temperate and polar zone (Coarser area - STemperate) requires 50 samples on average to detect a new species.

Table 4 - The origin of observational record in PhytoBase dataset.

Basis of Record	**Number of observations**	**Number of unique species**	**Percentage (%)**	**Rate of detection (Observations per species)**
Observation	752519	1201	59.37	626.57702
NA	368017	966	29.04	380.96998
Preserved specimen	94446	834	7.45	113.2446
Human observation	49446	629	3.90	78.61049
Occurrence	3056	76	0.24	40.21053
Machine observation	2	2	0.00	1
Living specimen	2	2	0.00	1
Literature	2	2	0.00	1
Total	1267490	1591	100.00	796.66248

The basis of record in the PhytoBase dataset (Table 4) highlights the gaps of available information within the dataset. While most records (59%) are classified as ‘Observation’, this terminology gives no indication of the method by which this observation was collected. In addition, 29% of the records have no information at all (‘NA’). Such information to the method

of observation is useful when sampling effort is biased to apply methods of sampling effort control.

Table 5 - Method of quantification in sampling methodologies supplied in the PhytoBase dataset.

Organism quantification type	Number of observations	Number of unique species	Percentage (%)	Rate of detection (Observations per species)
NA	1235352	1582	97.46	780.88
Number of cells per Litre	31775	315	2.51	100.87
Number of colonial form cells per Litre	363	2	0.03	181.5
Total	1267490	1591	100	796.66

Specifically, to marine phytoplankton, performing cell counts per litre allows cross comparison across different time and sampling locations. Unfortunately records with this quantification time accounts for less than 3% of the dataset (Table *5*). The majority (97%) of the data, as seen in Table 4 also has no information which could give rise to stricter methods of sampling standardisation.

2.2 Methodology

2.2.1 Introduction

The purpose of this section is to detail the general method I employed in the assessment of the global diversity of marine phytoplankton. This section will first outline the data pre-processing steps. The subsequent sections will detail the computation method, analytical methods and statistical approaches used to analyse and interpret the data. Additionally, this section will address the methodological challenges encountered, the strategies employed to mitigate potential biases, and the assumptions stated clearly. This section aims to provide a transparent and replicable framework for the assessment of marine phytoplankton diversity and be readily applicable to other fauna and flora using Hill numbers at a global scale.

The majority of the analyses conducted in this study were performed using R (version 4.2.2) and RStudio (version 2023.9.1.494), which provided a flexible and reproducible framework for data manipulation, statistical analyses, and visualisations. For regional ocean assessments, ArcGIS Pro (version 3.4.0) was used to attribute ocean region information, enabling spatial analysis and integration of geospatial data. This combination of tools ensured efficient handling of large datasets while leveraging specialised software for geospatial operations.

2.2.2 Pre-processing data prior to analysis

Before conducting the LDG analysis, several preparatory steps were undertaken to ensure the consistency, completeness, and accuracy of the data. These steps are outlined below. In addition to PhytoBase (Righetti et al., 2020), a supplementary taxonomic dataset (Table 3, supplementary material section 6.1) was compiled using records from AlgaeBase, an authoritative source of taxonomic and nomenclatural information for algae species worldwide (Guiry, M.D. & Guiry, G.M., 2023). This dataset was used to refine and validate species identifications within PhytoBase as PhytoBase contained scientific name, class and phylum information only. For later diversity assessments in subsequent chapters using similarity, complete taxonomic lineages were required. In cases of unclear or incomplete classifications

in AlgaeBase, WoRMS was consulted to resolve discrepancies and provide additional taxonomic clarity (Bernot et al., 2024).

2.2.2.1 Retrieving Taxonomic Information

To ensure accurate and complete taxonomic data, the PhytoBase dataset, containing 376 unique marine phytoplankton genera, was used as the basis for querying authoritative databases:

- **Querying AlgaeBase:** Genus names were extracted from the *scientificName* field using a string separator process in R and then used as search queries in AlgaeBase to retrieve full taxonomic lineages (Guiry, M.D. & Guiry, G.M., 2023).
    - Taxonomic levels sourced.
        - Species, Genus, Family, Order, Class, Phylum, Kingdom, Domain
- **Consulting WoRMS for Missing Data:** For genera not matched in AlgaeBase, the World Register of Marine Species (WoRMS) was used to fill in missing taxonomic information (Bernot et al., 2024).

2.2.2.2 Data Cleaning in R

The data cleaning process was carried out in RStudio (see Appendix: *PhytoBase_Cleaning.Rmd*). The primary goal was to standardise taxonomic names across the dataset to prevent mismatches when comparing the taxonomic lineage of each species. Key cleaning steps included:

- **Standardising Scientific Names:** Hyphens were removed from species names (e.g., "Pseudo-nitzschia seriata" was changed to "Pseudonitzschia seriata").
- **Excluding Incomplete Entries:** Records lacking genus-level information (e.g., entries listed as "Prochlorococcus" without further species identification) were removed. The total number of observations in the PhytoBase dataset is 1,360,621. When the data is filtered to remove incomplete taxonomic information, 1,271,863 observations remained. Of the 88,764 removed, 30,048 were Synechococcus and 19,921 were Prochlorococcus. In total 119 ‘species’ were removed during this processing step. The loss of data was 6.52% of the total number of original observations in the PhytoBase dataset. In particular for the small cyanobacteria, Synechococcus and Prochlorococcus which comprise 56.29% of the removed observations, the impact of this removal difficult to ascertain. These two ‘species’ are known to have extensive range in the world ocean and contribute to the global carbon cycle significantly. In this methodological approach the lack of full taxonomic resolution prevents their inclusion in subsequent analysis.

2.2.2.3 Taxonomic Data Integration and Validation

After retrieving and cleaning the data, full taxonomic lineages were successfully obtained for 326 of the original 376 genera, which were then matched to their corresponding entries in PhytoBase. Observations lacking complete taxonomic information were excluded, as similarity analysis requires full taxonomic lineages to avoid bias. To further ensure consistency, all taxonomic information was de-capitalized to prevent mismatches during subsequent analyses.

2.2.2.4 Integration of Data Sources

Combining the historical coverage of PhytoBase with the detailed taxonomic resolution provided by AlgaeBase and WoRMS, ensured the most accurate and comprehensive representation of phytoplankton diversity. This integrated approach facilitated a robust

foundation for analyzing global marine phytoplankton diversity while addressing potential inconsistencies in taxonomic classifications. These preparatory steps were crucial for maintaining the integrity and reliability of the LDG analysis, ensuring that the data used were both standardized and comprehensive.

2.2.2.5 Spatial delineation for analysis

Diversity assessments require partitioning areas into subsets to evaluate how diversity changes with location. The primary aim is area regularisation, ensuring that the area of assessment does not bias results. As previous research has shown, increasing the area also increases the number of species observed (Macarthur, 1965; Rosenzweig, 1995). Consequently, diversity measurements across areas of different sizes are not directly comparable. To address this, traditional approaches for assessing LDG often use consistent spatial units. Here, the terms coarse-grain (1° resolution) and fine-grain (55km^2 resolution) refer to the different area aggregations of the observational data.

2.2.2.5.1 Approaches to Area Selection

Two common methods for selecting areas include:

1. **Coarse-grain scale =** Fixed Latitude Bands: A 1° latitude band is widely used in LDG studies as a standard spatial unit.
2. **Fine-grain scale=** Equal-Area Grid: Subsetting into smaller, uniform grid cells (e.g., 55 km²) reduces inherent biases related to varying area sizes, particularly at higher latitudes where spatial distortion occurs.

2.2.2.5.2 Coarse-grain Diversity assessment

To assess coarse-grain diversity, 1° latitudinal bands spanning the Earth were used, irrespective of land boundaries. This was implemented through a coordinate-centering process in R using the PhytoBase dataset.

The steps involved were as follows:

1. Longitude (x) and latitude (y) data from PhytoBase were extracted.
2. A function was applied to calculate the grid cell centers:
    - Each observation's coordinates (e.g., 23.89° longitude, 35.68° latitude) were floored to the nearest whole integers (e.g., 23 and 35).
    - Half the grid size (0.5° for 1° x 1° cells) was added to the floored values, centering the observation in the middle of the grid cell (e.g., 23.5 and 35.5).
3. Three new columns were generated:
    - X for centered longitude,
    - Y for centered latitude, and
    - Cell for the combined grid cell identifier.

Using the Y-coordinates, coarse-grain diversity for each 1° latitudinal band was calculated with diversity indices.

2.2.2.5.3 Fine-grain scale Diversity assessment

Fine-grain diversity was assessed using an equal-area grid dataset (EAG) file with a cell size of 55 km². This grid ensures that the reduced area near the poles does not bias comparisons with the larger equatorial regions.

The procedure included the following steps:

1. The PhytoBase dataset was converted into a shapefile using raw longitude and latitude data.
2. A spatial join operation matched observation coordinates to the grid cells in the EAG file.
3. The merged dataset contained a unique grid cell identifier for each observation.

This combined dataset facilitated the calculation of fine-scale diversity using indices while accounting for the uniform area delineation.

2.2.3 Hill number assessment method

To assess the marine phytoplankton LDG, the study utilises the taxonomically enhanced PhytoBase dataset previously outlined, which integrates data from PhytoBase, AlgaeBase, and the World Register of Marine Species (Bernot et al., 2024; Guiry, M.D. & Guiry, G.M., 2023; Righetti et al., 2020). This dataset encompasses global observational records of phytoplankton species along with their taxonomic classifications. Diversity assessment focuses solely on richness and evenness, employing the naïve Leinster and Cobbold (LC) index (Leinster & Cobbold, 2012a), which is mathematically equivalent to the Hill number framework (Hill, 1973b) and is important for the context of subsequent work. Consequently, the results and methodology are referred to as Hill numbers throughout this chapter.

Following the latitudinal grouping of the observations for the coarse-grain diversity assessment, the relative abundances of each species within the 1° latitude bins were calculated. Each row of the taxonomic PhytoBase dataset represents one observation of a species of marine phytoplankton. To calculate the relative abundances, a new data column (“distinct_values”) was added to represent one individual recorded per row. This column contains an integer value of 1 for each observational data row in the full dataset. The total number of observations per 1° latitude was then summed from the 'distinct_values' column (new column called “cellobs”). The species richness per 1° latitude was calculated for each latitude (new column called “specount”) using a distinct function. The total number of recorded presences per species per 1° latitude was determined by grouping the dataset by latitude and scientific name and summing the “distinct_values” column (new column called “TotalpLat”).

The total number of recorded presences per species per 1° latitude (“TotalpLat”) was divided by the total number of observations per 1° latitude (“cellobs”), providing the relative abundances of each species within each latitudinal area of assessment (coarse-grain diversity assessment). Notably, the relative abundances here are a modified statistical interpretation of this common terminology. Traditionally, relative abundances are calculated per sampling regime and/or per methodological approach by using known sampling effort one can then calculate the relative abundance of individuals of species under a replicable method of detection. Here I deviate from this, owing to the methodological assumption of the removal of time in assessment of marine phytoplankton diversity. Where I group observations by spatial location and do not consider time due to the sampling effort variability. As highlighted previously the sampling methodology of the observations in this dataset is not consistent across all samples used here. Therefore, relative abundance in this dataset is more akin to relative

probability of persistence through time. Though this probability of persistence is likely biased by effort in the same way as relative abundance, but it provides a methodological approach by which this global assessment can be performed. This preparation of relative abundances, taxonomic classification information, and latitudinal grouping provides the foundation for the Hill number diversity assessment which requires area delineation, species richness and species abundance distribution to assess the diversity of a location.

The diversity assessment was performed for each area, at the coarse and the fine-grain scale. The fine-grain method followed the same general approach except instead of latitude, the assessment was performed per grid cell of the EAG. Four diversity indices were calculated: Hill numbers (*q0,q1,q2,q3,q4,q5*), species richness (Hill number at *q0*), the Shannon index, and Pielou's evenness index. Where *q* values for the Hill Number index are the sensitivity parameter to species rarity, decreasing in sensitivity as the value of *q* increases. The Shannon index was calculated independently of the Hill number framework using the vegan package in R (Oksanen et al., 2024), which served as the basis for calculating Pielou's evenness index (Pielou, 1996).

2.2.4 Generalised additive modelling method for trend fitting

The tidyverse geom_smooth function is used in RStudio when plotting the marine phytoplankton data (RStudio, 2020; Wickham et al., 2019). The geom_smooth function fits a smoothing function for the independent variable, latitude. The smooth function allows for a flexible, non-linear relationship between variable *x* and variable *y* using these parameters. When used in this form, the geom_smooth function uses a LOESS (Locally estimated scatterplot smoothing) method for data with less than 1000 data points, as is the case in the latitudinal diversity gradient analysis. The local estimation occurs at each focal point (a data point, in this case the diversity value) and uses a 75% data window to fit a unimodal curve for that data window around the focal point. The data points closest to the focal point are weighted higher than those further away using a tricubic kernel function. This procedure is repeated for each data point and a 95% confidence interval is constructed for each of the localised fits per focal data point. The smoothed trend line is then constructed by joining all of the focal unimodal line functions to generate a smooth trend line of the data. As this function applies a series of functions and joins them, there is not a single function governing the trendline and as such is data visualisation approach rather than a formal statistical model.

2.2.5 Rarefaction method for sample bias correction

Sampling effort is a critical consideration in large-scale diversity analyses. To address potential biases introduced by uneven sampling effort, rarefaction analysis was conducted for four large global regions to generate rarefied Hill number values at different sensitivity levels (represented by the parameter *q*).

2.2.5.1 Data Preparation

Following the methodology of splitting the PhytoBase dataset into four regions were selected to assess the impact of sampling effort on diversity quantification: NTemp, NTrop, STrop, and Stemp. These regions were chosen based on observed differences in sampling effort between the northern and southern hemispheres.

- Latitudinal ranges:
    - NTemp: [40°N, 80°N]
    - NTrop: [0°N, 40°N]
    - STrop: [-40°S, 0°N/S]
    - STemp: [-78.04°S, -40°S]

2.2.5.2 Rarefaction Procedure

Two rarefaction approaches were applied to the different latitudinal regions: continuous and non-continuous subsampling. This procedure is useful to standardise the sampling effort at a continuous and non-continuous level to capture the greatest range of subsampling while minimising the assessment processing time.

Firstly, as the continuous subsampling performs 999 assessments per area, it is replicated 10 times to generate the rarified Hill number diversity to optimise computing resource demand. Each region is randomly subsampled without replacement 2-1000 times. During this process, the number of observations, n, sampled by n +1 [1:1000]. At each sample point a diversity assessment is performed using the Hill number framework applied latitudinally at coarse and fine-grain spatial scales. This gives insight into the accumulation rate of diversity in these regions irrespective of sampling effort disparity.

Secondly, the non-continuous approach subsamples the diversity in chunks: 1, 5, 10, 50, 100, 500, 1000, 5000, 10000, 15000, 20000. The choice of chunks aimed to give a wide even spread of samples up to the limit of 20000 samples. This limit is due to the max number of observations in the Southern Temperate region (Temperate and Polar) reaching a maximum of 27246 compared to the > 300,000 observed in the other three regions. As there are 11 subsample events per region the resources needed are reduced. This analysis was replicated 50 times. The sensitivity parameter q was varied from 0 to 5 to account for different aspects of diversity, richness and evenness. When the rarefaction procedure is performed in Chapter 3, by using the LCI, similarity is also included.

2.2.5.3 Applied Rarefaction Method

The rarefaction analysis was performed for each marine region using a nested loop structure. The outer loop controlled the number of replicates, while the inner loop iteratively increased sampling effort from 2 to 1000 observations.

For each iteration:

1. Sampling: Rows were randomly sampled (with replacement) from the regional dataset to match the current sampling effort.
2. Richness Calculation: The number of unique species in the sampled rows was calculated.
3. Hill Number Calculation: Diversity metrics (relative abundances, observations per species, and total observations) were computed, and Hill numbers for $q=0,1,2,3,4,5$ were calculated using a modified diversity function developed in prior analyses.

For each region, the mean, standard deviation, and confidence intervals were calculated for richness and each Hill number across replicates.

2.2.5.4 Evenness Assessment

Pielou's index was used to evaluate the effect of sampling effort on community evenness. In cases where samples consisted of a single species, the calculation of Pielou's index produced *NaN* values. To address this, these instances were assigned a value of 1, reflecting the theoretical assumption that a single-species sample represents complete evenness.

This comprehensive analysis framework ensured robust assessment of sampling effort effects on diversity quantification, providing insights into the reliability and variability of regional diversity estimates.

2.3 Results

2.3.1 Coarse-grain scale latitudinal diversity assessment

Examining richness, the latitudinal-richness relationship for marine phytoplankton prior to sampling correction reveals a bimodal species richness LDG (Figure 2.1). The maximum richness, encompassing 586 species, is located at 17.5°N latitude within the northern subtropics. However, the greatest peak of diversity, as identified by the GAM, occurs further north at approximately 30°N. A slightly smaller secondary peak, also indicated by the GAM, is centered around -19°S in the southern subtropics. Beyond these peaks, richness declines toward the polar regions, where the minimum levels of species richness are observed. This bimodal pattern deviates from the previously observed global LDG observed in marine phytoplankton (Righetti et al., 2019).

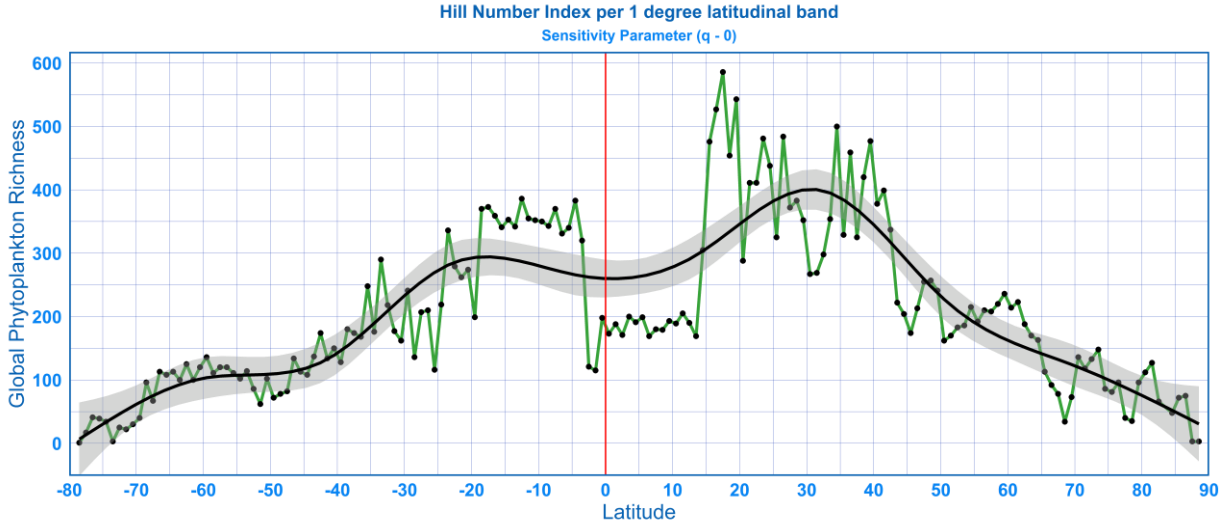


Figure 2.1 - Global marine phytoplankton richness across latitudes derived from a Generalised Additive Model (GAM) prior to sampling correction. Red vertical line denotes equator, black dots symbolise raw richness and grey shading is 95% confidence interval from GAM.

2.3.2 Coarse-grain scale observational effort latitudinal assessment

Marine phytoplankton sampling effort is not equal across latitude: observational data is focused in two latitudinal peaks, firstly in the northern hemisphere (40° to 60°) and a smaller peak in the southern hemisphere (-5° to -25°) (Figure 2.2). The remaining coverage is characterised by relatively low observation counts (<5000 observations).

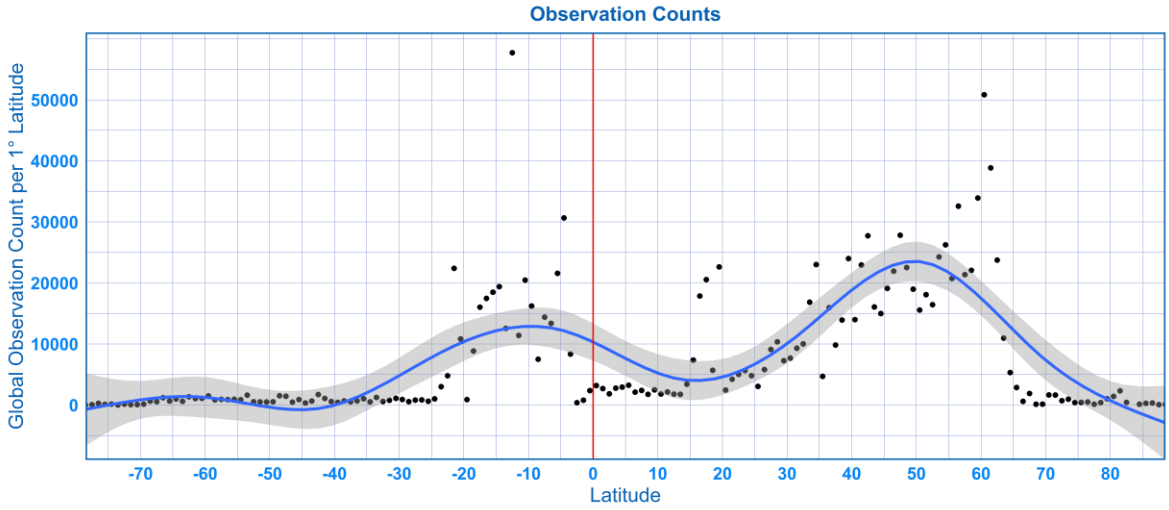


Figure 2.2 - Coarse-grain scale aggregated number of global marine phytoplankton observations. The blue line is a generalised additive model applied to the observation counts with grey shading detailing the 95% confidence interval.

The two peaks in observational effort (Figure 2.2) do not latitudinally fully align with the peaks in species richness diversity (Figure 2.1). While the southern hemispheric peak in richness does show alignment, the high sampling effort at 60°N doesn't match a peak in diversity. Consequently, when assessed for effort to richness saturation, Figure 2.3 shows that species richness saturates with further sampling.

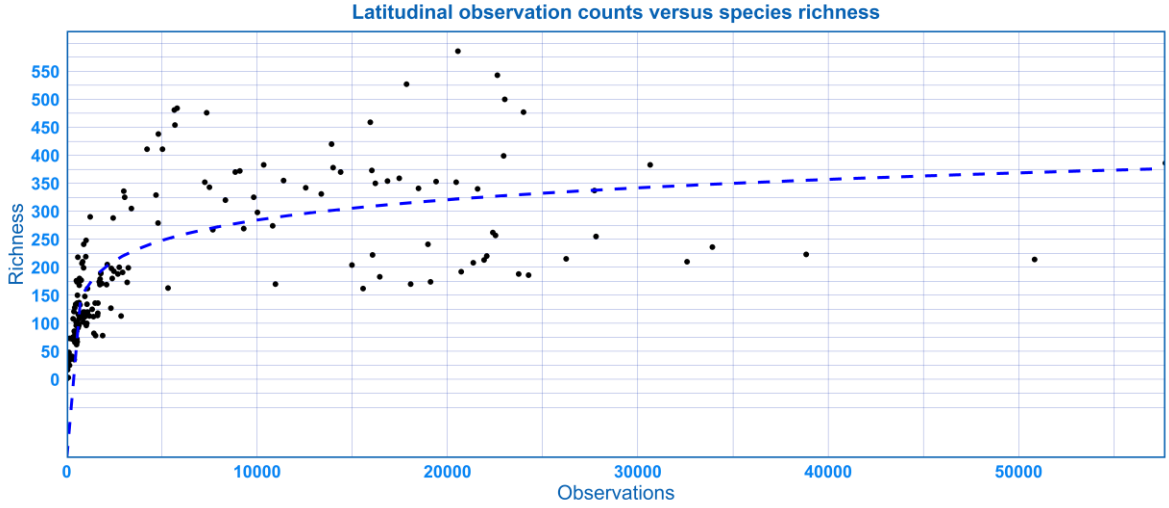


Figure 2.3 – Coarse-grain scale observations and species richness correlation. Blue dashed line is linear relationship of variables with an R^2 value of 0.28.

2.3.3 Fine-grain scale latitudinal diversity assessment

Following the fine-grain scale Hill number *q0,* the greatest richness is located in coastal regions such as Peru, the pacific coastline of Mexico, Atlantic France and the east China sea (Figure 2.4A and 2.4B). When log scaling regional differences in species becomes apparent, varying by two orders of magnitude from 1 species to a maximum of 527 species at 16.69°N in the northern sub-tropics (Figure 2.4B). The fine-grain LDG pattern (Figure 2.4B) mirrors that of the coarse-grain scale LDG species richness assessment (Figure 2.2) with a bimodal (two peaks of richness) distribution and a dip in diversity at the equatorial region. The fine-grain assessment differs from the coarse-grain assessment by detailing the spatial distribution of the species richness (Figure 2.2A and 2.2C).

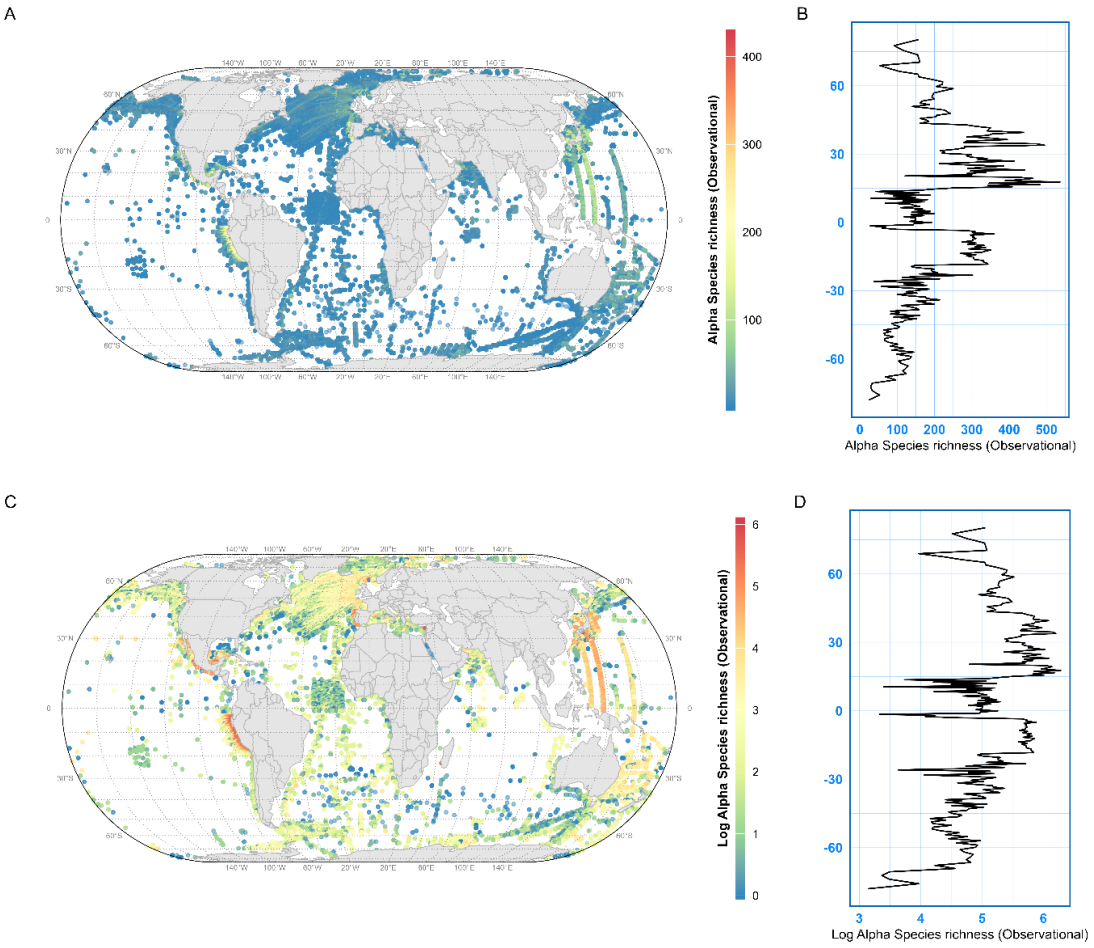


Figure 2.4: Phytoplankton diversity on fine-grain scale. (A) Marine phytoplankton species richness (Hill number – q0), (B) Latitudinal richness using fine-grain scale, (C) Logarithmic (log) marine phytoplankton species richness, (D) Log latitudinal species richness using fine-grain scale.

2.3.4 Fine-grain scale observational effort latitudinal assessment

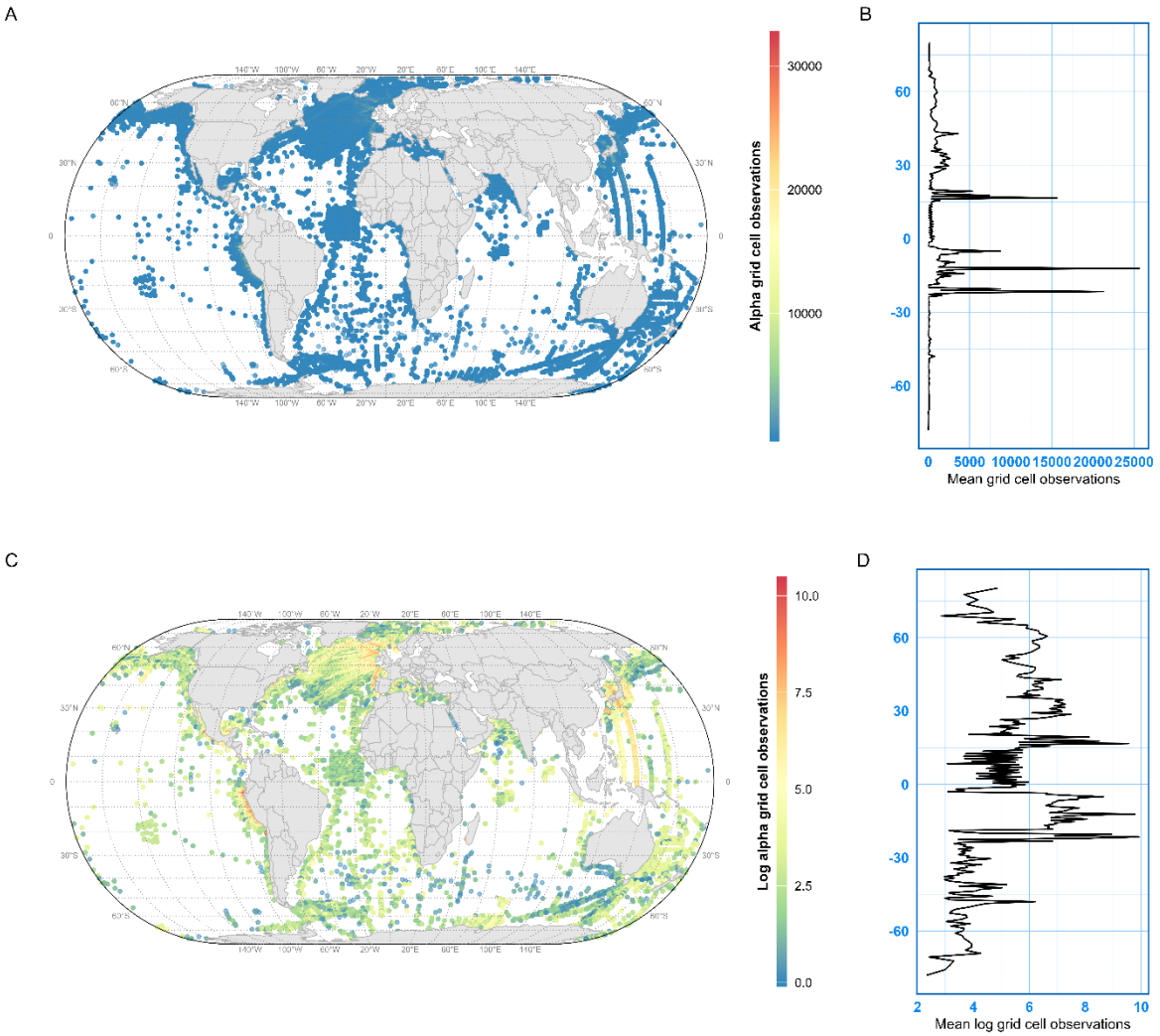


Figure 2.5 - Phytoplankton observational sampling effort on fine-grain scale. (A) Marine phytoplankton aggregated sampling effort, (B) Latitudinal aggregated observational effort, (C) Logarithmic (log) aggregated sampling effort, (D) Log latitudinal aggregated sampling effort in fine-grain scale.

There is a high disparity in sampling effort (observations of marine phytoplankton) as the majority of locations assessed have less than 5000 observations (shown by blue markers)(Figure 2.5). As can be observed in Figure 2.5A, most fine-grain grid cells exhibit lower observational counts. Regionally, the northeastern Atlantic around the UK coastal shelf system, the coastal upwelling system of Peru and the coastal region around Japan exhibit high sampling intensity of marine phytoplankton (Figure 2.5C). The LDG of mean observations per 55km^2 equal area grid latitudinal coordinates highlights this disparity with the maximum mean number of observations located on the coastal upwelling system of Peru in the southern hemisphere (25588.76 at -12.14°S). The logarithmic scaling of observations (Figure 2.5C) underlines the regions highlighted in the non-logarithmically adjusted observation plot (Figure 2.5A). The logarithmic mean observational LDG in Figure 2.5D provides further insight to the relative observational intensity latitudinally, where overall the northern hemisphere has a high sampling effort compared to that of the southern hemisphere with the exception of the southern sub-tropical latitudes which contain the highest sampling effort regime for marine phytoplankton. The high sampling disparity globally presents challenges to ecological interpretations.

2.3.5 Coarse-grain Hill number latitudinal diversity assessment

Hill numbers, through the adjustment of the sensitivity parameter (q), offer an alternative perspective on the LDG of marine phytoplankton. As shown in the figure, richness is represented by $q = 0$, consistent with prior observations. From $q = 1$ to $q = 5$, the key distinction in the LDG lies in the reduction of measured diversity across all latitudes as q increases, reflecting decreased sensitivity to rare species. Notably, the general bimodality of the LDG persists across all values of q, suggesting that even the most commonly observed species, those least affected by sampling effort, follow this bimodal pattern.

The Hill number diversity index uniquely allows for the depiction of a diversity profile by varying q. As q increases and rarity is down-weighted, the relationship between diversity and q becomes evident. This profile is curvilinear, and the rate at which diversity decreases with increasing q reflects the unevenness of the community. This feature is particularly useful for comparing communities; two communities with similar richness can differ markedly in evenness, which would be evident as crossing lines in their diversity profiles.

The diversity profile also reveals patterns of unevenness by latitude (Figure 2.6). The change in diversity between successive values of q from $q = 0$ to $q = 5$ is not uniform across latitudes. The relative percentage decline in diversity from $q = 0$ to $q = 5$ serves as a measure of community unevenness at a given latitude and shows that the northern temperate region exhibits higher unevenness compared to other latitudes, a pattern further substantiated when the percentage decline is calculated and plotted against latitude.

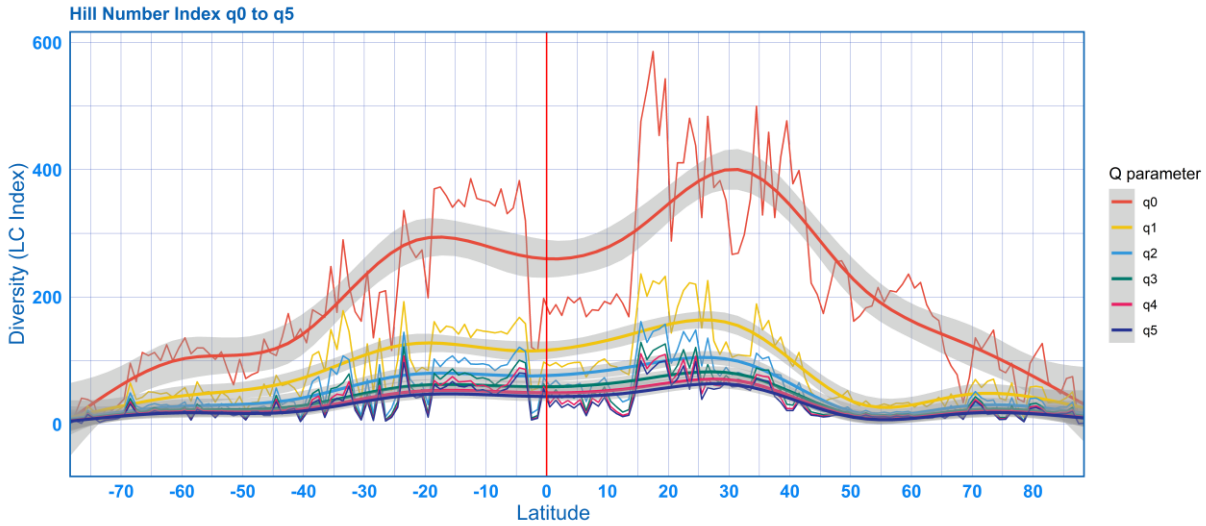


Figure 2.6 - Coarse-grain scale global marine phytoplankton diversity across Hill numbers (Q parameter: q0–q5). The plot illustrates coarse-grain scale diversity per 1° latitude with a generalised additive model (GAM) smoothing function applied to each Hill number line, shown as associated colour lines with grey shading representing 95% confidence intervals. The horizontal red line indicates the equator at 0° latitude.

The northern temperate region exhibits a greater percentage decline per latitude in diversity from Hill number *q0 - q5* (Figure 2.7), indicating higher community unevenness in this region. In contrast, the polar regions show a smaller percentage decline, suggesting that these communities are more even. However, it is important to note that these regions correspond to areas of lower sampling effort, resulting in reduced confidence in these findings.

Broadly, latitudes outside the northern temperate and polar regions display a similar rate of decline in Hill number diversity as q increases, which suggests comparable evenness profiles across these communities despite variations in richness. The observed deviations, particularly

in the northern temperate region, are of particular interest and warrant further exploration to better understand the factors driving these patterns.

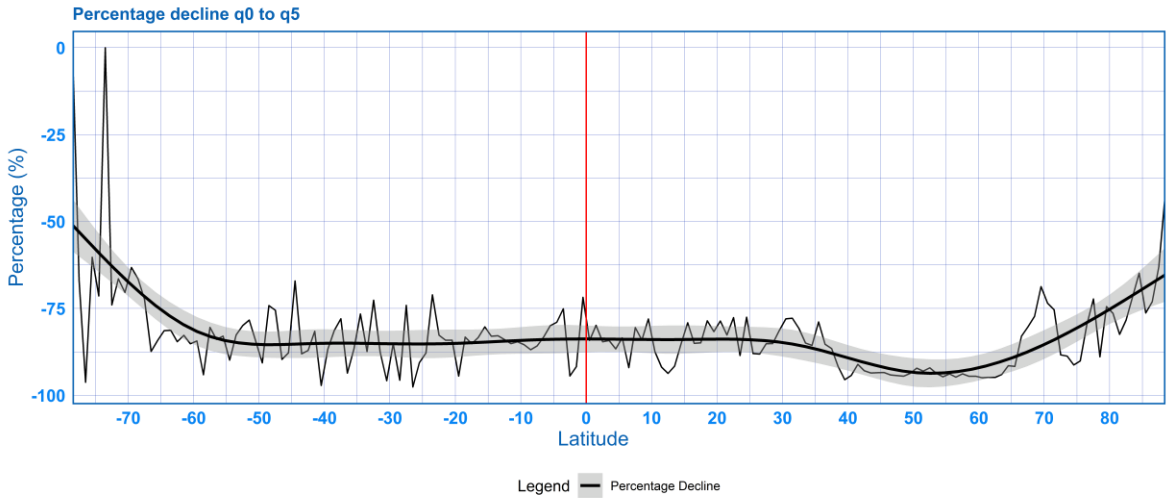


Figure 2.7 - Percentage decline in diversity from q0 to q5 per latitude. The plot illustrates coarse-grain scale percentage decline from q0 to q5 per 1° latitude (thin black line) with a generalised additive model (GAM) smoothing function applied, shown as associated black line with grey shading representing 95% confidence intervals. The horizontal red line indicates the equator at 0° latitude.

To further confirm this finding, we assessed the evenness of the latitudinal communities using Pielou's evenness index (Figure 2.8). Here a value of 1 represents a fully even community of species with equal relative abundances, whereas as the index tends towards zero, indicates an uneven community. Specifically, the evenness declines in the northern hemisphere to around 0.625 between the latitudes of 30° and 70°.

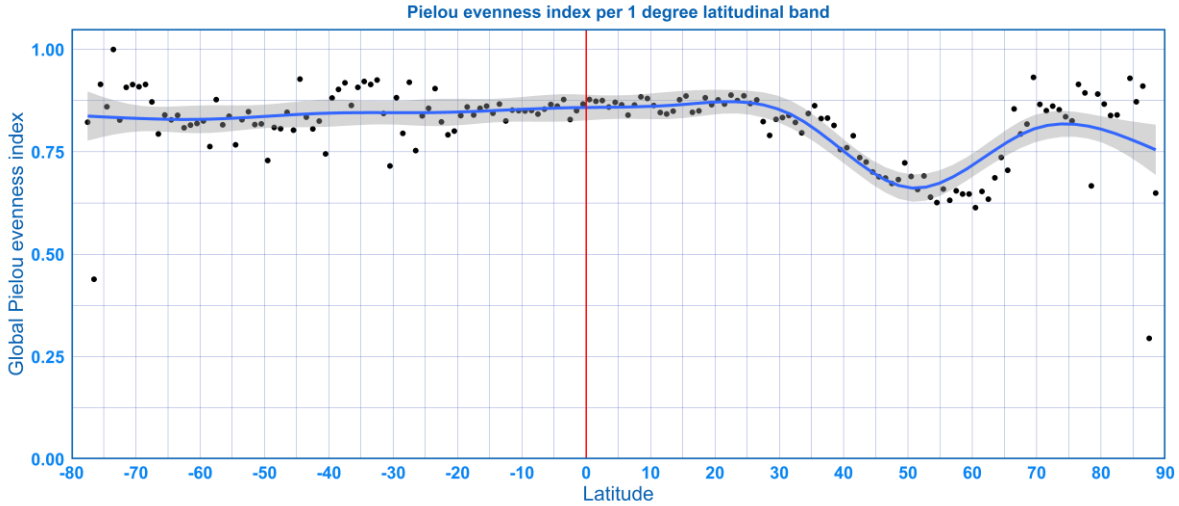


Figure 2.8 - Coarse-grain scale global marine phytoplankton Pielou evenness index (J) assessment. This graph details the coarse-grain scale Pielou evenness index (J) assessment of marine phytoplankton per 1° latitude (black circle points. The horizontal red line shows the equatorial line at 0° latitude. The blue line is a generalised additive model smoothing function applied to the results of the Pielou index diversity assessment with grey shading detailing the 95% confidence interval.

The correlation plot between the percentage decline in diversity (from *q0* to *q5*) and the Pielou index reveals a positive relationship (Figure 2.9), indicating that communities with higher evenness, as measured by the Pielou index (Figure 2.8), tend to exhibit smaller declines in diversity as rarity is downweighted. This relationship highlights the intrinsic link between

evenness and the sensitivity of diversity indices to rare species within a community. Higher evenness implies a more balanced distribution of species abundances, leading to less pronounced effects of downweighting rare species. Consequently, regions with elevated Pielou index values, which denote more equitable community structures, are associated with reduced variability in diversity across Hill number sensitivities.

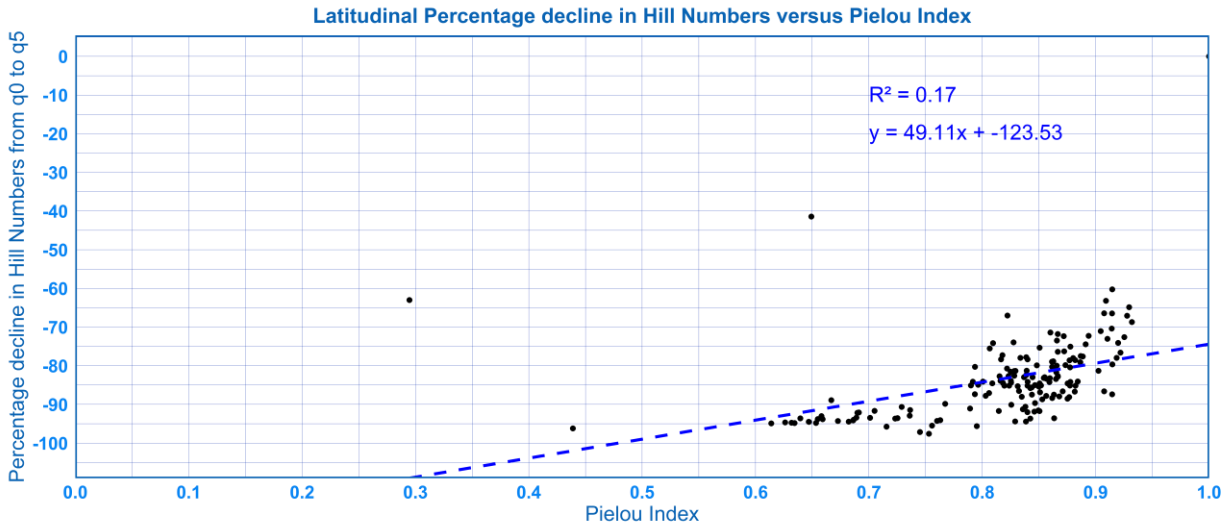


Figure 2.9 - Correlation plot for percentage decline and Pielou-index relationship. Blue dashed line is linear relationship of variables with an R^2 value of 0.17.

The fine-grain scale Pielou evenness index (*J*) assessment of marine phytoplankton (Figure 2.10) deviates from the coarse-grain scale assessment (Figure 2.8) with a substantial decline observed near the equator. Whereas the coarse-grain scale has an observable dip in the northern temperate hemisphere. The fine-grain scale assessment (Figure 2.10B) does indicate a slight dip in the mean evenness observed in the northern temperate latitudes ~60°N. This decline in evenness does not match the extent of the dip near the equator. The spatial map of the Pielou evenness index (*J*) does show that the northern Atlantic and the tropical mid-Atlantic is comparatively more uneven than the global pelagic systems in general.

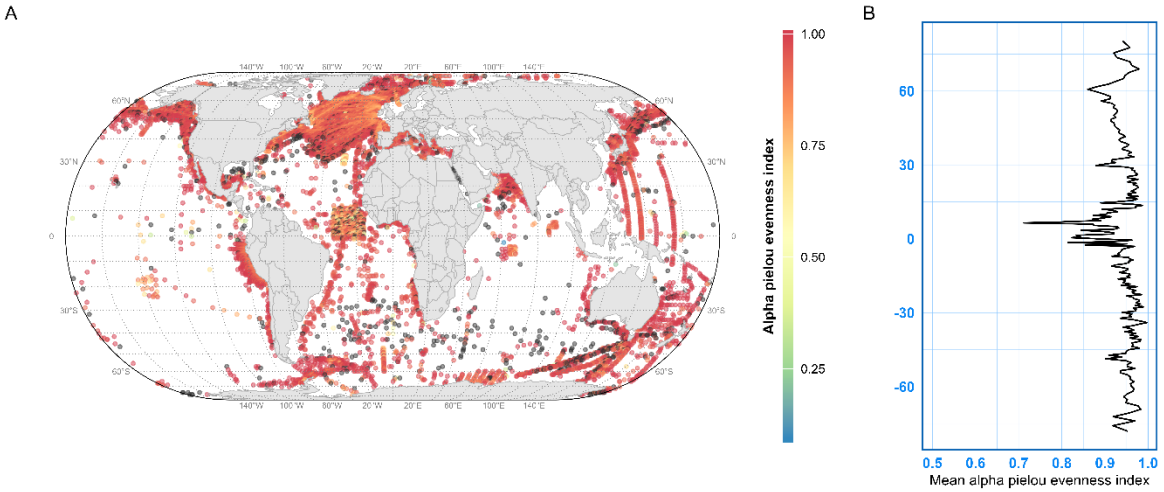


Figure 2.10 – Fine-grain scale aggregated global marine phytoplankton Pielou evenness index (J). (A) the Fine-grain scale Pielou evenness index (J) assessment of marine phytoplankton per 55km2 equal area grid cells (coloured circle points indicating center point of grid cells). Warmer colours (red/orange) indicates higher community evenness and cooler (green/blue) colours indicates lower community evenness (grey circles indicate cells where index did not compute). (B) details the latitudinal (Fine-grain scale) mean Pielou's index.

2.3.6 Hill number rarefaction

The coloured sectioning of the hill number coarse-grain latitudinal assessment (Figure 2.11) shows how the data was split for the rarefaction diversity reassessment. While the data spans from 88.92°N to 78.04°S, I wanted to make an even split as far as possible to reduce bias. Therefore, using the southern data boundary, proceeding northwards each region covered as close to 40° latitude as possible. The southern temperate and polar region (STemp, blue) spans [-78.04°S, -40°S]. The southern tropical and sub-tropical region (STrop) is covered by the next region between latitudes, [-40°S, -0°S]. The northern tropical and sub-tropical region (NTrop) is bounded by [-0°S, 40°N] latitudes. Lastly, the northern temperate and polar region (NTemp) is covered by the truncated latitudes of [40°N, 80°N].

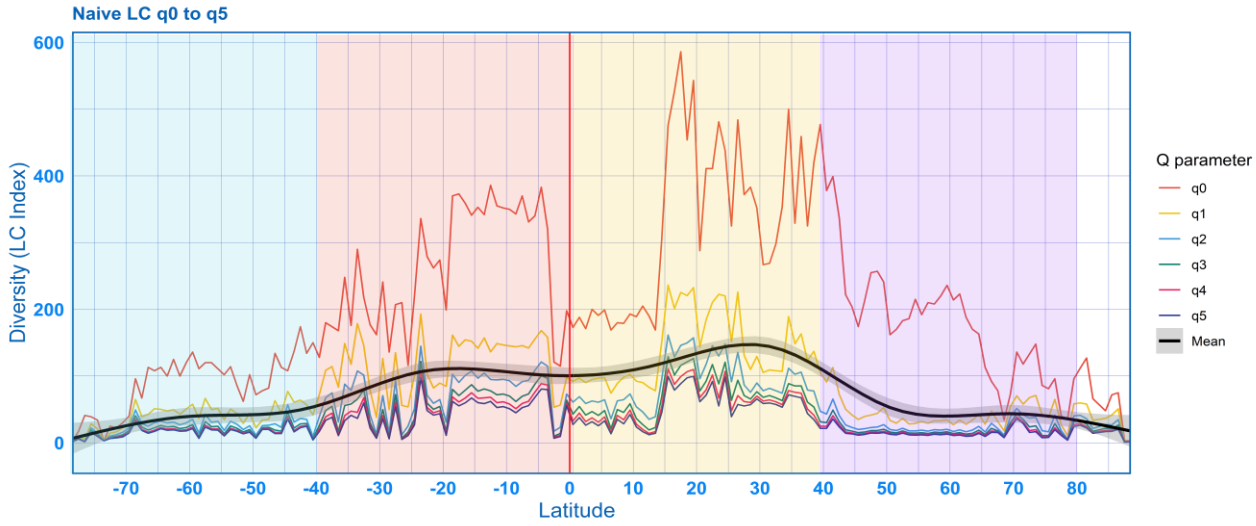


Figure 2.11 - Hill number LDG assessment. Colour coded for the rarefaction assessment regions. Blue block is southern temperate, red block is southern tropical, yellow block is north tropical and the purple block is north temperate.

The 0-1000 continuous rarefaction results for each region and suite of Hill numbers are shown in Figure 2.12. Here it is confirmed that the largest rarified richness is the northern tropical region with a estimation that 280 species are sampled in the first 1000 random samples (replicated 10 times). The lowest richness rarefied is the northern temperate region showing 140 species. Whereas the difference between the southern regions is less pronounced with 240 in the southern tropical region and 200 in the southern temperate region. What the Hill number rarefaction details in the other curves drawn per plot is the evenness relationship as detailed in the coarse-grain Hill number assessment (Figure 2.6). Here we observe that the northern temperate region is the least even with the greatest relative spread as the sensitivity parameter q is increased. Both the tropical regions show greater evenness than their respective temperate counterparts.

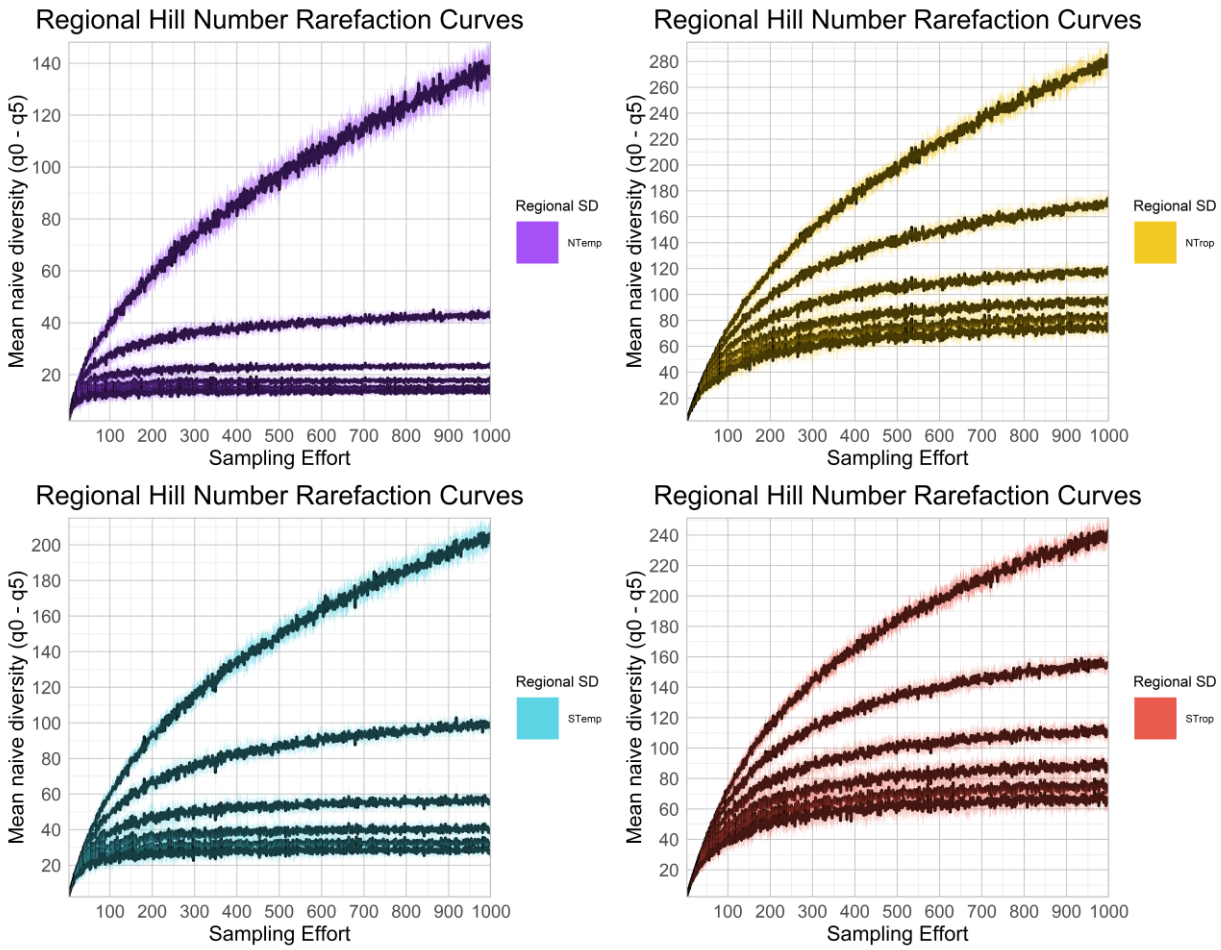


Figure 2.12 - Hill number rarefaction assessment from 0 to 1000 samples repeated 10 times. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

The 0 to 20,000 discontinuous rarefaction results for each region and suite of Hill numbers are shown in Figure 2.13. Here it is further confirmed that the largest rarified richness is the northern tropical region with a estimation that >800 species are sampled in 20,000 random observation samples (replicated 50 times). The lowest richness rarefied is the northern temperate region showing 140 species. Whereas the difference between the southern regions is less pronounced with 240 in the southern tropical region and 200 in the southern temperate region. What the Hill number rarefaction details in the other curves drawn per plot is the evenness relationship as detailed in the coarse-grain Hill number assessment (Figure 2.6). Here we observe that the northern temperate region is the least even with the greatest spread as the sensitivity parameter q is increased. Both the tropical regions show greater evenness than their respective temperate counterparts.

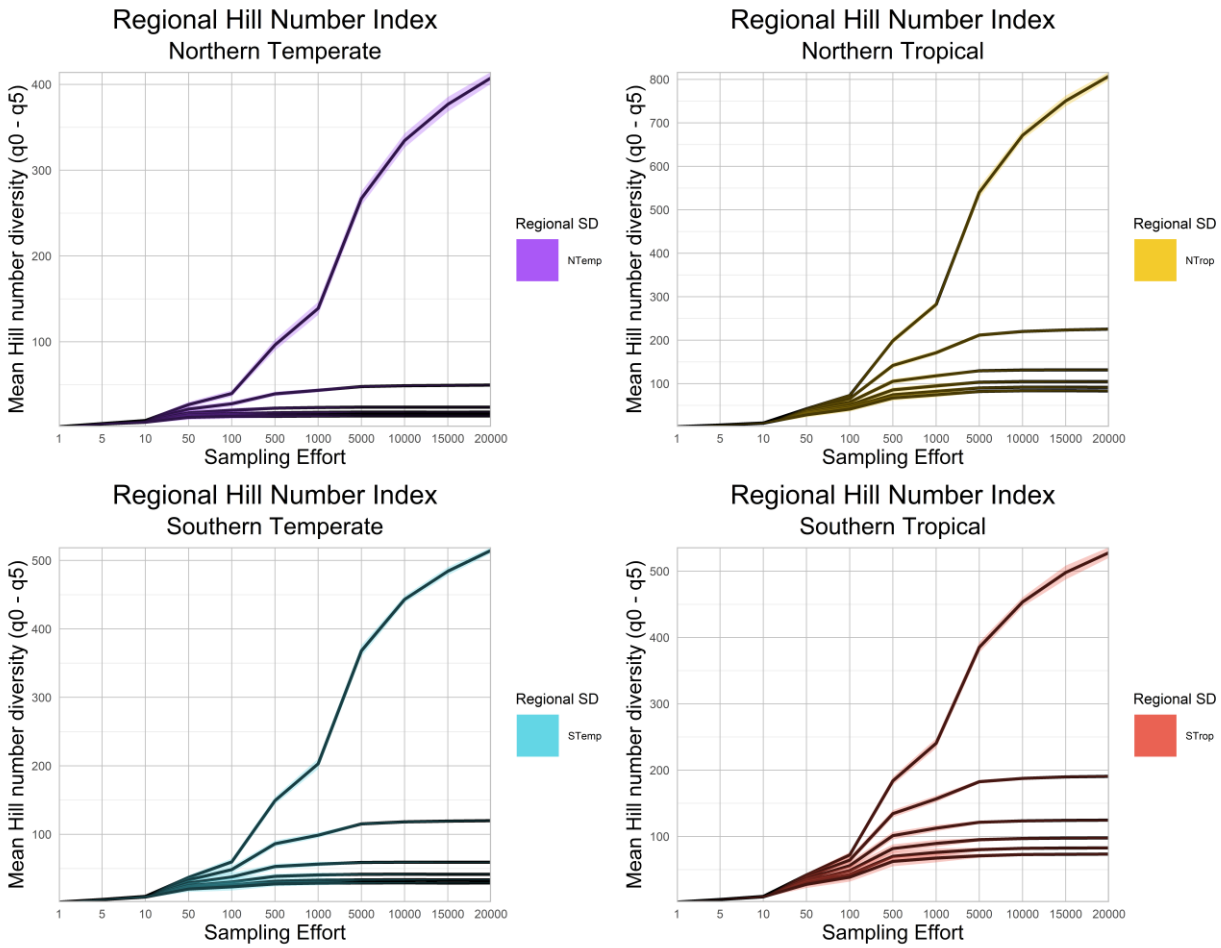


Figure 2.13 - Hill number rarefaction assessment from 0 to 20,000 samples (discontinuous) repeated 50 times. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

The Pielou's index evenness assessment for each region further underlines the disparity in community evenness measures as community resampling increases (Figure 2.14). This further exemplifies the northern temperate region as having low community evenness. The southern temperate region is intermediate while the tropical regions are the comparably even compared to temperate regions.

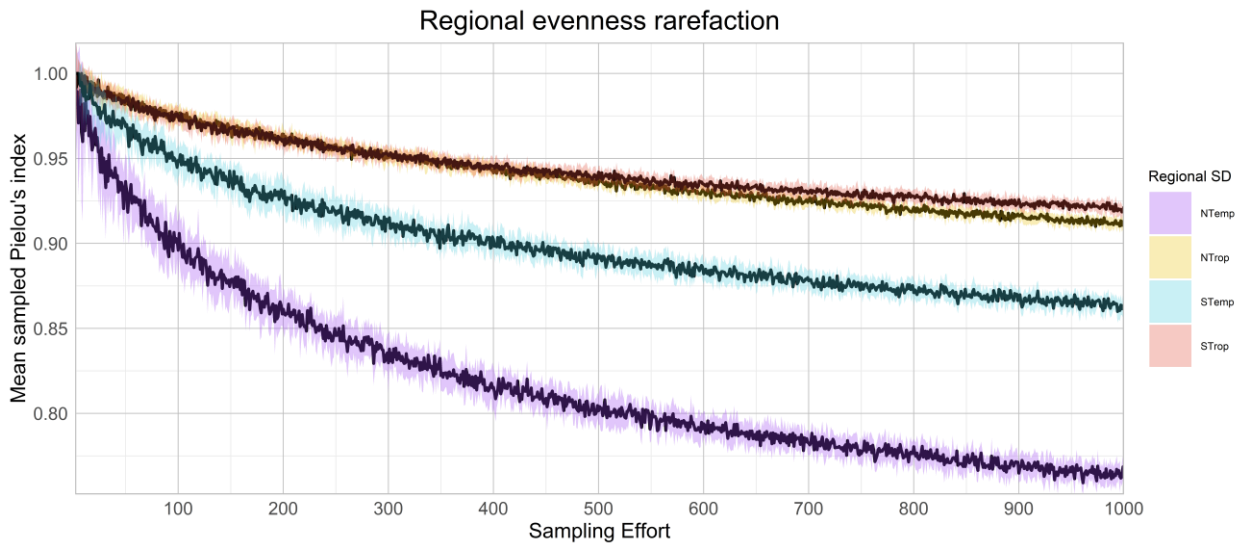


Figure 2.14 – Pielou's evenness rarefaction assessment from 0 to 1000 samples repeated 10 times. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

The species richness rarefaction curves (Figure 2.15) indicates that the southern hemisphere increases in species richness at a faster rate than the northern regions. The northern temperate region shows the lowest richness accumulation rate. The northern tropical region and southern temperate region show similar rates of species accumulation.

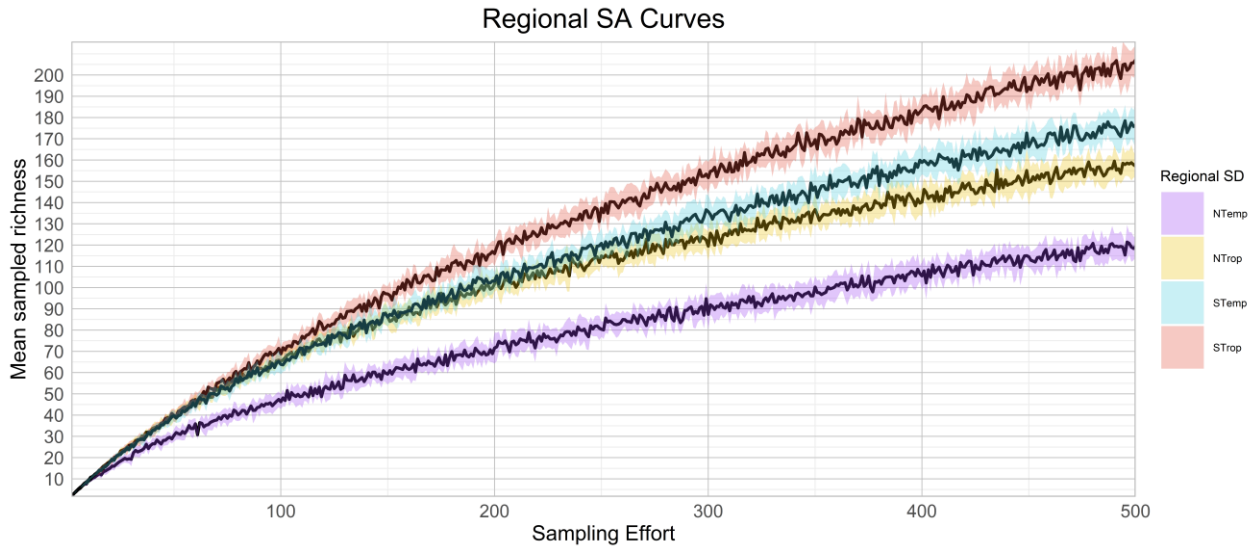


Figure 2.15 – Species richness rarefaction assessment from 0 to 1000 samples repeated 10 times. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

2.4 Discussion

Marine phytoplankton form the foundation of marine food webs and drive global carbon cycling (Litchman et al., 2015), yet their global diversity remains poorly understood beyond regional and localised dynamics (Righetti et al., 2020). This study applied Hill numbers to provide an evenness weighted, global perspective on phytoplankton diversity, not yet performed to date. By addressing key gaps in our macroecological knowledge of marine phytoplankton, this research not only unveils novel diversity patterns but also expands the scope of diversity definition in macroecological research.

After the pre-processing stage, the global diversity analysis and examining coarse and fine-grain diversity at differing spatial scales was designed and coded to be universally applicable to any diversity dataset given available data. This methodological flexibility ensures the framework can be adapted for broader applications in biodiversity science. By leveraging this approach, we move closer to a broader understanding of global marine phytoplankton diversity while simultaneously highlighting existing sampling biases in available data.

The latitudinal diversity gradient (LDG) of marine phytoplankton remains largely undescribed compared to larger, more readily observable fauna and flora (Colwell & Coddington, 1994). While multiple extensive reviews and data syntheses are available for other groups (Chaudhary et al., 2016; Hillebrand, 2004; Tittensor et al., 2010b), marine phytoplankton research has been limited to one recent data synthesis and a single global species distribution model study to date (Righetti et al., 2019, 2020). Traditionally, the research community has relied on mechanistic models to describe the biogeography of marine phytoplankton (Barton et al., 2010; Dutkiewicz et al., 2020). These models offer a bottom-up understanding of potential patterns, providing valuable insights into phytoplankton distributions. The mechanistically modelled diversity with allometric body size functions not explicitly grounded in specific species, making ground-truthing of species distributions difficult. In contrast, the results presented here contribute a novel, descriptive perspective to the field of macroecology and marine phytoplankton biogeography.

It is important to note that marine phytoplankton may not adhere to a single general pattern as expected or shown considering the substantial sampling effort disparity. Deviations from the expected LDG have been observed for other distinct groups, such as pinnipeds and marine mammals (Tittensor et al., 2010b) underscoring the complexity and variability inherent in marine ecosystems.

When focusing solely on species richness, ongoing debates persist regarding the modality of the LDG pattern and the influence of climate change on macroecological trends (Chaudhary et al., 2021; Menegotto & Rangel, 2018). Climate change has already demonstrated significant effects on local and regional scales (Kok et al., 2017), altering community dynamics through shifts in extinction, immigration, and emigration rates as environmental niche space changes (Harley et al., 2006). However, at macroecological scales, assessing the diversity patterns of smaller organisms like phytoplankton remains challenging due to the lack of consensus on what the ‘expected’ pattern should look like (Chaudhary et al., 2016; Visser et al., 2014).

2.4.1 Comparison to other LDGs

Two notable global assessments of small organism marine LDGs conducted in the past 15 years which examined diverse marine groups, include zooplankton, bacteria and foraminifera (Egorova et al., 2025; Ladau et al., 2013; Tittensor et al., 2010b). Foraminifera, unlike many other planktonic species, leave a geological record due to their calcium carbonate tests (Jonkers et al., 2022). This highlights a geological sampling bias for geological reconstructive LDG

work for marine phytoplankton which outside of coccolithophores and diatoms do not leave a substantial geological footprint (Falkowski et al., 2004).

Although our study did not explicitly replicate their methods, an comparison of fine-scale grain diversity patterns suggests potential overlap in richness distributions, particularly with cetaceans in temperate regions. However, differences in grid resolution, our analysis uses finer spatial aggregation, highlight the potential for future research to explore the implications of these methodological choices. Nonetheless, such exploratory analyses provide a foundation for generating hypotheses about shared and unique drivers of biodiversity across marine taxa.

2.4.2 Differences in terrestrial versus marine LDG research

The LDG patterns observed in terrestrial and marine systems highlight key differences and unique limitations, primarily shaped by the sampling accessibility and physical characteristics of each environment (Burrows et al., 2011). Terrestrial systems are more regularly studied due to their greater accessibility, enabling more frequent species observations and data collection (Raffaelli et al., 2005; Webb, 2012). In contrast, marine systems present significant challenges for research, driven by the vast geographical area of the ocean and the logistical barriers associated with oceanic transportation. While terrestrial studies do encounter biases related to fiscal resource limitations, these challenges are amplified in marine systems, contributing to the disparity in our general knowledge of LDGs across these ecosystems.

Another fundamental distinction between terrestrial and marine LDGs lies in the physical medium influencing species distribution. Terrestrial ecosystems are often shaped by physical barriers such as mountain ranges, which restrict the movement of species and contribute to the evolution of new species over time. In marine environments, identifying equivalent barriers is more complex, but ocean circulation patterns and salinity gradients in estuarine systems can serve similar roles by hindering species movement biotically. However, the effects of these marine barriers are nuanced by their size and the nature of the species involved (Wilson et al., 2016).

The global advection of phytoplankton may enable more extensive population dispersal, potentially altering the expected relationship between the LDG and the level of dispersal. This could result in a more diffuse, flatter LDG pattern, where species are more widely distributed, and diversity is more evenly spread across latitudes, but we did observe that in this assessment, a strong bimodal outcome. This outcome is the result of a non-standardised sampling effort distribution, without further sampling to fill data gaps, this pattern is a indication of the potential pattern. However, recent research suggests that not all marine phytoplankton species exhibit the same level of global dispersal, and limits in niche adaptability may limit this potential (Ward et al., 2021). We observe some species in this data are observed at both poles, raising the question of whether these populations represent the same population or separate, isolated groups. If the latter is true, then the distance between populations could function similarly to terrestrial mountain ranges, acting as a barrier to genetic exchange and contributing to the unique shape of the marine LDG.

2.4.3 Novelty of evenness inclusion in LDG assessment

By including evenness as a diversity attribute, we introduce a deeper understanding of global marine phytoplankton community structure (Alatalo & Alatalo, 1977). While richness provides a snapshot of diversity, evenness can reveal the health and stability of ecosystems. Communities dominated by a single species, or a few rare species, are more vulnerable to shifts in productivity, particularly in response to habitat changes driven by climate change (Dutkiewicz et al., 2021; Hillebrand et al., 2022). In contrast, more even communities tend to

have greater functional redundancy, making them more stable and resilient. The Hill number framework, which includes a sensitivity parameter to evenness, allows us to assess the combined effects of richness and evenness across latitudes (Hill, 1973b). The observed decline in evenness in the northern temperate region, when q values exceed zero, suggests that these communities are more dominated by a few species compared to those in the southern hemisphere.

One of the key challenges in operating on the scale of LDG for marine phytoplankton is the limited interpretation of how local or regional effects might influence diversity (Wadgymar et al., 2022). While we can measure richness and evenness, we lack the ability to distinguish the specific impacts of individual species on community structure and subsequent function at a given latitude. To advance our understanding, future research should focus on specific groups of marine phytoplankton and their functional roles within the broader LDG context like has been done for size traits and copepods (Acevedo-Trejos et al., 2018a; Benedetti et al., 2025). Incorporating taxonomic and functional trait data could provide valuable insights into range shifts and their contribution to local diversity functioning. Without these finer resolution assessments, relying solely on richness metrics would only reveal shifts due to climate change but fail to capture more subtle changes in community dynamics. Which could cascade tropically through community productivity changes i.e. shifts from larger species of phytoplankton to smaller species as the upper latitudes warm due to climate change (Atkinson et al., 2024; Gardner et al., 2011).

However, caution is needed when interpreting these findings, as sampling effort in the northern latitudes is disproportionately high, potentially skewing results. The seasonal nature of marine phytoplankton blooms, coupled with the sampling biases (concentrated effort in the North Atlantic and lower winter sampling) inherent in the CPR system (Richardson et al., 2006), may be hypothesised to contribute to the observed dip in evenness. This may be due to the northern temperate region's highly productive environment, characterised by frequent phytoplankton blooms, resulting in CPR samples being dominated by a small number of individual species at different times. Investigating this phenomenon further would involve assessing diversity using the Hill number framework at various time points, while considering the functional ecological biology of marine phytoplankton. A more detailed understanding of how biological processes shape community structure over time, particularly in high-productivity regions like the North Atlantic, could provide new insights into the relationship between diversity and productivity on a global scale. Further research is needed to capture these dynamics more comprehensively and understand their role in shaping the LDG of marine phytoplankton.

2.4.4 Temporal and spatial bias

To address temporal biases in our dataset, we chose not to include time in our diversity assessments, focusing solely on observation location. This decision was based on the likelihood that standardising sampling times would result in a highly fragmented dataset, as it is improbable to have substantial global-scale observations on any given day. Even with seasonal binning, significant differences could arise, and while the role of seasonality in the LDG of marine phytoplankton is a critical area for future research, it was not a focus of this thesis. To first assess the LDG of marine phytoplankton, a general assessment was required before considering the effects of large-scale temporal processes, such as seasonality or climate change. To mitigate spatial biases, we used rarefaction to assess the effect of sampling effort across four broad regions. The rarefaction results confirm the LDG assessment of marine phytoplankton for both species richness and Hill numbers at these broad scales. This consistency across diversity measures strengthens the reliability of our conclusions regarding the spatial distribution of marine phytoplankton diversity. This provided a general understanding of the

impact of effort at this scale. A higher-resolution correction, such as one per latitude, was not feasible due to computational and time constraints.

2.4.5 Size and sampling methods

Marine phytoplankton present unique challenges in the biological sciences due to their broad size range. Typically, organisms can be observed using consistent methods, such as microscopy, but marine phytoplankton vary widely in size, from diatoms, readily classifiable under a microscope, to the smallest genera like *Prochlorococcus*, which require genetic methods for differentiation. This methodological inconsistency makes traditional species richness metrics incompatible for assessing phytoplankton diversity. Small phytoplankton species, despite being less observable, are vast in numbers and contribute significantly to biomass (Goericke & Welschmeyer, 1993; Visintini et al., 2021).

Diversity assessments, therefore, require methods that bridge these methodological divides. Accounting for this 'hidden' diversity, which extends to bacteria and viruses of macroecological interest, calls for a unified approach, potentially a genetic method. However, this raises challenges regarding the definition of a species, particularly given processes like horizontal gene transfer, which allows rapid adaptability and shifts in niche space and resource use. At a global scale, this complexity is difficult to account for, compounded by sampling disparities and methodological inconsistencies.

This need for universal equations or rules governing diversity highlights the importance of generalisable frameworks. Such frameworks could play a role in biology analogous to the Navier-Stokes equations in physics, enabling broad insights without requiring resolution of specific edge cases. However, without real-world validation and predictive capabilities, the general applicability of such insights remains unrealized, limiting their practical utility.

2.4.6 Functional Diversity and LDG Coupling Across Trophic Levels

The role of phytoplankton functional separation in the LDG is likely tied to the functional differences among diatoms, dinoflagellates, and cyanobacteria, which are primarily nutrient and size driven differences (Banas, 2011; Dutkiewicz et al., 2020). However, the size variability within these groups, their niche differentiation in nutrient requirements, and subsequent spatial diversification and subsequent broad group level specialisation might weaken any strong peaked latitudinal relationship. Instead, these functional differences may manifest more prominently at longitudinal scales, with coastal, pelagic, and deep-ocean environments serving as major contributors to spatial differences among these groups and species (Acevedo-Trejos et al., 2015b; Malviya et al., 2016).

While size diversity relationships at a global scale provide additional insights, these analyses are currently constrained by existing spatial sampling biases (Acevedo-Trejos et al., 2018b). The relationship between LDGs of marine phytoplankton and their consumers, such as zooplankton and higher trophic levels, is not well characterised but is a topic of significant interest. This underscores the need to specify the type of diversity being examined. For instance, exploring functional diversity, such as the size-to-energy relationship between producer and consumer, is distinct from examining species richness (Cermeño et al., 2013). These two types of diversity are not necessarily coupled, and hypotheses about their ecological interactions must be grounded in clarity regarding the diversity metric in question.

To date, a global zooplankton diversity assessment within the euphotic zone has not been undertaken in a replicated approach to what has been done here for phytoplankton. Developing hypotheses about LDG couplings or decoupling's between phytoplankton and zooplankton requires such foundational mapping. For example, if phytoplankton exhibit maximum diversity

in subtropical regions while zooplankton peak in temperate regions, this could indicate an interaction between the functional LDG of these groups and factors such as predation mechanisms, temperature, dispersal, and evolutionary history could lead to depressions in species richness at higher latitudes. These potential decoupling's offer intriguing opportunities to explore how LDGs interact with functional diversity patterns.

Functional diversity is challenging to quantify because it involves both continuous traits (e.g., size, biomass) and categorical traits (e.g., feeding strategies, color, mating behavior). Developing unbiased measures that capture this complexity equally is no small task. Achieving equality in diversity measures is essential unless equity is explicitly the goal, as in restoration projects where targeted measures are needed to assess the recovery of specific species, such as seagrass. Traditional measures like richness and evenness may not be sufficient to capture these nuanced dynamics. Tailored functional diversity metrics, developed with practical applications in mind, are critical for addressing these challenges. This complexity in measuring and quantifying functional diversity introduces a high degree of nuance that can be initially daunting for mathematicians or statisticians. However, it also highlights the importance of interdisciplinary approaches to create frameworks that are both theoretically sound and practically applicable.

2.4.7 Limitation and assumptions

The limitations of this study highlight several areas for future research. Firstly, the relationship between sampling effort and diversity measures, beyond richness, was not explored in depth. While richness is relatively straightforward to assess, understanding how effort influences evenness and similarity remains an open question. Future work should consider alternative methods, such as Chao2 or Absolute Effective Diversity Estimators, to better characterise the effects of increased effort and spatial heterogeneity on these metrics (Cazzolla Gatti et al., 2020; Chao, Gotelli, et al., 2014). Additionally, this study assessed LDG without incorporating temporal variability. Marine phytoplankton diversity and hemispheric patterns are likely to change seasonally, as suggested by (Ladau et al., 2013), but these dynamics were beyond the scope of this analysis.

A further limitation involves the handling of undescribed species, particularly *Prochlorococcus* and *Synechococcus*. These genera, though globally widespread (Hirata et al., 2011) were excluded from statistical assessments due to lack of available taxonomic information in PhytoBase which may have implications for evenness profiles. While their minimal impact on richness metrics was noted, their potential influence on evenness at specific latitudes highlights a methodological challenge when integrating such species into diversity analyses. Similarly, this study did not investigate the community structure of marine phytoplankton or its seasonal variability. Evenness plays a critical role in understanding community structure, but its implications at latitudinal and global scales remain uncertain. Yet methods such as the approach taken in this thesis give insights to community evenness distributions globally even if further work remains to resolving sampling effort effects in these results.

The interplay between sampling effort and community characterisation is another area requiring further exploration. High sampling effort is expected to capture more rare species, providing a more complete picture of the community. However, biases introduced by mesh filters, particularly in excluding small phytoplankton species, must be addressed to ensure accurate representation (Richardson et al., 2006).

Finally, the study did not examine the size diversity of marine phytoplankton, which is closely linked to productivity (Acevedo-Trejos et al., 2018a). While studies such by Beardall et al. (2009) highlights the importance of size diversity, this aspect was outside this works' primary

scope. Additionally, the potential effects of differing sampling methods on LDG estimates were not addressed, leaving a question of how methodological choices may influence results. These limitations underscore the complexity of diversity assessments and the need for further research to refine methodologies and expand our understanding of marine phytoplankton diversity.

In this study, it is assumed that the LDG of marine phytoplankton has remained static over the period of data collection or that the temporal coverage in our dataset represents an average state of the LDG (Fenton et al., 2023). This approach was taken by assessing the LDG based on location only, without accounting for temporal variation. However, it is important to note that the samples in our dataset were collected at different times, seasons, years, and effort differing across latitudes. The assumption of a static LDG was considered acceptable because the gradient has been observed over geological timescales, >200 years of marine phytoplankton present day data collection (Mannion, 2020; Zacaï et al., 2021b). Making our 200-year 'snapshot' a reasonable representation of a long-standing pattern. Nonetheless, this assumption is likely flawed if the impacts of recent climate change are as significant as current evidence suggests, potentially altering the spatial and temporal dynamics of marine phytoplankton diversity (Chaudhary et al., 2021). The results of this thesis indicate the LDG of marine phytoplankton is bimodal, with two distinct peaks in the subtropical regions. Barring the potential influence of sampling effort (Menegotto & Rangel, 2018), there is a decline in diversity at the equator. A possible reason for this is that the equator has risen in SST to a sufficient degree to be thermally limiting for larger species and thus more selective, reducing overall diversity (Chen, 2015; Thomas et al., 2012). The observed bimodal pattern could be a recent development rather than being representative of the LDG of marine phytoplankton in deep time which may in fact have been unimodal with a peak in the tropics.

A key assumption of Hill numbers, as with all diversity measures, is the assumption of full detectability, meaning that all species present are assumed to have been observed. This assumption can only be addressed through estimation methods, by fitting a function to the abundance distribution and extending it to account for missed species (Chao, Gotelli, et al., 2014). However, this would only be an approximation, as there are likely more species present than those observed, particularly for small organisms like marine phytoplankton, where diversity maxima are uncertain and difficult to quantify accurately (Mora et al., 2011). As diversity quantification is dependent on classification, estimation measures do not detail what species are missing, only quantifying how many may have been missed. Therefore, that is outside of the scope of this research currently. Mechanistic models can resolved this by creation of realistic theoretical size spectra and empirical nutrient demand relationships (Dutkiewicz et al., 2020).

Another assumption in Hill number calculations is that the measured diversity accurately represents the phenomena of the area under study. This is often not the case, as many species may remain undetected, especially in environments with high species turnover or in areas with limited sampling (Rodríguez-Ramos et al., 2014). The diversity captured in these measurements may therefore be incomplete or biased, further complicating our understanding of global or regional patterns.

Additionally, Hill numbers assume that species are correctly identified, and there are no misidentifications. Misidentification is a practical issue that diversity indices cannot account for. This is a limitation that could only be addressed through reanalysis using updated and corrected species identification data.

3 Global diversity of marine phytoplankton using a similarity-sensitive diversity index

3.1 Introduction

This chapter is an extension of the research on the Hill number diversity gradient of marine phytoplankton conducted in Chapter 2. Here, I expand the definition of diversity to include similarity using the Leinster and Cobbold Index (LCI) (Leinster & Cobbold, 2012a). Measuring community similarity is beneficial as it allows for deeper differentiation of the diversity between communities (Chen & Grinfeld, 2024; Daly et al., 2018). The creation of the LCI was motivated by the observation that species are assumed to be distinct entities. Yet the ignorance of this biological reality results in indices unable to distinguish between equally rich and balanced communities.

In marine phytoplankton studies, diversity analysis is generally restricted to species richness and a mix of richness and evenness diversity assessment metrics (Ibarbalz et al., 2019; O'Brien et al., 2016; Righetti et al., 2019). Globally, marine taxa groups are also studied using the same two forms of diversity quantification, richness and evenness (Blowes et al., 2019; Tittensor et al., 2010b). This is also follows for terrestrial macroecological studies i.e. large-scale diversity studies (Gagné et al., 2020; Jenkins et al., 2013). These indices, however, consider species as unrelated entities, failing to account for the relatedness among species in their diversity calculations (Chen & Grinfeld, 2024; Daly et al., 2018). This omission limits the understanding of community structure by not capturing the inherent evolutionary similarities between species within the communities. A comprehensive assessment of diversity should therefore include not only richness and evenness, but also the similarity or relatedness of species within a community. This aspect of diversity is often overlooked due to the complexities and challenges of its interpretation (Daly et al., 2018). The LCI offers a solution by incorporating not only richness and evenness, but also the relatedness of species, addressing this critical aspect of diversity that traditional indices often miss.

This chapter presents the first global assessment of marine phytoplankton diversity that incorporates species taxonomic similarity, addressing a significant gap in current knowledge. Not only to obtain a deeper understanding of marine phytoplankton macroecology but also the application of the LCI to a macroecological question. This study applies the LCI to assess the impact of incorporating similarity into global marine phytoplankton diversity estimates. Traditional diversity metrics rely on a single information channel, species richness, to quantify community structure. The Hill number framework adds evenness, capturing the influence of dominant species (Hill, 1973b). However, by integrating similarity, the LCI offers a new perspective beyond the richness-evenness paradigm, providing deeper insights into community composition. The definition of similarity is defined by the user i.e. taxonomic, functional or genetic. Similarity in this case is defined as the taxonomic similarity between species, this was chosen as it was the simplest form of similarity to calculate in this novel application of the index.

In this study, I measure similarity of marine phytoplankton using taxonomic classification of species, the more taxonomic information two species have in common, the more similar they are. Each species within a community is compared, taxonomically, with all other species of a community creating a measurement of the similarity/dissimilarity within the community. To achieve this, I developed novel methodological procedures, specifically integrating the LCI and its decomposition for macroecological assessments at varying spatial scales.

3.2 Methodology

As noted in the introduction, the methods of this chapter overlap, specifically the data used, and the method of spatial assessment are the same as the previous chapter. This chapter is distinguished by the usage of the taxonomic data in the diversity assessment for similarity quantification. Additionally, a new element of the methodology is the decomposition of the LCI by Chen and Grinfeld (2024). The method for doing so is described in detail after a general overview of the previous methods.

3.2.1 Overview of previous methods

The methodology of this chapter closely follows the previous chapter. The pre-processing (Section 2.2.2), spatial delineation, how diversity is quantitatively assessed (Section 2.2.3) is applied and the rarefaction method (Section 2.2.5) are the same. This chapter methodology deviates from the previous chapter by quantification of similarity using the Leinster and Cobbold Index. In addition, the decomposition of the LCI is performed at coarse-grain spatial scale to review the decomposed richness, evenness and similarity latitudinal patterns. Therefore, I will briefly give an overview of the methods for those detailed in the previous chapter and detail explicitly the role of similarity in this chapter and its application in this research chapter.

A second important note for clarification is that the index used in this chapter, the Leinster and Cobbold diversity index, is an extension of the Hill number framework. The LCI builds upon the existing Hill number framework by integrating and weighting the resulting diversity measure by the relative similarity of the community measured for diversity. As such in the previous chapter, the Hill number framework is referred to in totality due to the unique nature of the LCI formulation. The formulation and nature of the LCI means that when the similarity is assumed to be fully distinct i.e. all species are assumed unrelated, the ‘naïve’ similarity matrix integration results in the same results as would be obtained using the Hill number framework. This is outlined and shown in the work of Leinster and Cobbold (2012).

3.2.1.1 Phytoplankton diversity data

The diversity data used in this chapter is the same as described in Chapter 2 methods section (Section 2.2.2). The taxonomic data gathered from AlgaeBase and WoRMS is used in this chapter to define and quantify the relatedness of species within the diversity assessment (Section 2.2.2.3). All the species taxonomic data was checked for any missing gaps in taxonomy, and if found, was removed from the dataset.

3.2.1.2 Diversity indices used

It should be noted in cases where previous results are used, the previously diversity analysis may be referred to as Hill numbers which is presented as the naïve case of the LCI in Leinster and Cobbold’s work, these are equivalent terms. This was previously noted in Section 1.1.2. The mathematical, qualitive and practical application of the LCI are outlined in sections 1.1.2.6, 1.1.3 and 3.2.2 respectively.

3.2.1.3 Area partitioning

The area partitioning for this chapter diversity analysis follows the previous method of Section 2.2.2.5 for consistency and comparison in the following results section and discussion. The largest spatial scale used is the coarse-grain scale, which corresponds to a 1° latitude grid. This scale is commonly employed in Latitudinal Diversity Gradient (LDG) studies (Barton et al., 2010; Cermeño et al., 2008; Righetti et al., 2019); however, its use is sometimes contested due to potential biases arising from the unequal area coverage by latitude, a consequence of the

Earth's spherical shape. Therefore, we also assessed the LDG of marine phytoplankton using an equal area grid (55km^2) at a finer resolution than that of the coarse-grain assessment.

3.2.1.4 Rarefaction and the Leinster and Cobbold index

The rarefaction method used in this chapter follows the previous chapter also with the addition of the similarity being included in the diversity assessment. The details of how this is calculated is outlined in the technical section of the previous chapter (section 2.2.5).

3.2.2 Leinster and Cobbold applied calculation

The diversity assessment is performed on the previously outlined areas of assessment (AOA), coarse-scale (1° latitude) and fine-scale grain (55km^2 equal area grid). The marine phytoplankton data are partitioned into these different spatial areas prior to initialising the diversity assessment procedure, a detailed description of this method is in section 2.2.2.5. The calculation of marine phytoplankton diversity per AOA is performed sequentially per each area of partitioned phytoplankton observation data, using a for-loop process. The length of the looping sequence is determined by the number of areas being assessed for diversity. In coarse-grain we assess the diversity of marine phytoplankton in 167 latitudes, less than the global 180 latitudes due to the lack of data in latitudes where no observations are made. In the fine-scale grain diversity assessment we assess (9246 grid cells), which is less than the 164,120 total grid cells due to lack of spatial heterogeneity in effort.

To calculate the LCI per AOA, the first step is to subset the global dataset to a unit of area, either a latitude band or a grid cell of the fine-scale grain area partitioning. Following this, a dataset of all observations for all species per unit area is created. Here and in the following text, I will refer to number of species as *NSp.* This differs from the Hill number application of this index which uses an identity matrix of size of *NSp* x *NSp*.

Following this, a dataset of all observations for all species per unit area is created. The second step in the assessment is calculating the relative abundances of each species in the area. This is performed by tallying the number of observations of each species in the area and dividing these subsequent species totals by the total number of observations in the area in a row-wise fashion. This results in a vector of relative abundances mutated across the dataset which at this stage contains all observations. The following step is keeping only the distinct values in the dataset, which results in a dataset containing one row per species record in the area and a value of relative abundance for that species. Finally, the row-length of the distinct values dataset is the number of species in this AOA. The *NSp* and the resulting column of relative abundances is p and all values of p_i is equal to 1.

The next stage in the assessment is the creation of the similarity matrix (Z). Firstly, a matrix of zeros of size *NSp x NSp* is created for the AOA for diversity measurement. To calculate similarity, I cycle loop-wise in i and j which are coordinates of the similarity matrix. These coordinates represent the rows and columns of dataset in the previously subsetted dataset. This subsetted dataset only the distinct values, i.e. number of species in that area. As the loop operates, a value of 1 is inserted into the Z_{ij} coordinates when there is a taxonomic match between two species. A match check is performed for each rank of taxonomy from species to phylum. Kingdom and domain were not included as these are highly similar due to being the largest and most basic of classifications. Therefore they do not result in a distinctive contribution to the value of similarity measured globally. Each species is compared taxonomically to each other species. The diagonal of this matrix will be the maximum number of matches as i and j will be equivalent at the diagonal, meaning each species is compared to itself at least once. Each element of Z_{ij} is then divided by the number of ranks of taxonomy

used in the similarity assessment. This results in a matrix where the diagonal is 1 representing full similarity and all other values within a range of 0-1 describing the relative similarity of each species to each other. Once the relative abundances of the community and the similarity matrix are calculated, the LCI is then calculated and the value is stored for graphical analysis.

3.3 Results

Here I present the LDG patterns of marine phytoplankton with the inclusion of taxonomic similarity. This is then interpreted in greater detail using the Chen and Grinfeld decomposition (Chen & Grinfeld, 2024). As done previously, the results are then sampling effort standardised using the rarefaction approach outlined in the previous chapter. The decomposition at this stage is performed for the coarse-grain assessment only.

3.3.1 Leinster and Cobbold index coarse-grain assessment

The LDG of marine phytoplankton with similarity inclusion (Figure 3.1B) differs when compared the previous assessment (Figure 3.1A). Here, a single peak in the modality of the LDG occurs but does not occur at the expected equatorial regions. The southern hemisphere peak (between -45°S and -30°S) indicates that richness and the dissimilarity in the communities here are greater than that of other latitudinal regions. Here, the maximum number of effective species at *q0* with taxonomic information in ~10. The actual number of species at this latitude is 175 if no similarity is considered. Therefore, when taxonomic similarity is considered, there is ~10 effective species in this location. What this means in practice is that the taxonomic information is mostly not different, in terms of effective diversity this community of 175 species would be represented by the diversity information contained in ~10 maximally distinct species of equal abundance. By comparison to this, if all species were maximally similar and misclassified as multiple species the value of the similarity-sensitive LCI would be 1. As these similarity-sensitive LDG diversity values are close to an effective species number of 1, the general degree of taxonomic similarity in phytoplankton communities globally is high. In line with the LDG pattern, the expected decline in diversity towards the poles does appear as expected. The driving force for the increase of relative dissimilarity in the southern hemisphere is not readily apparent from single LCI diversity value as it contains information from species richness, evenness and similarity.

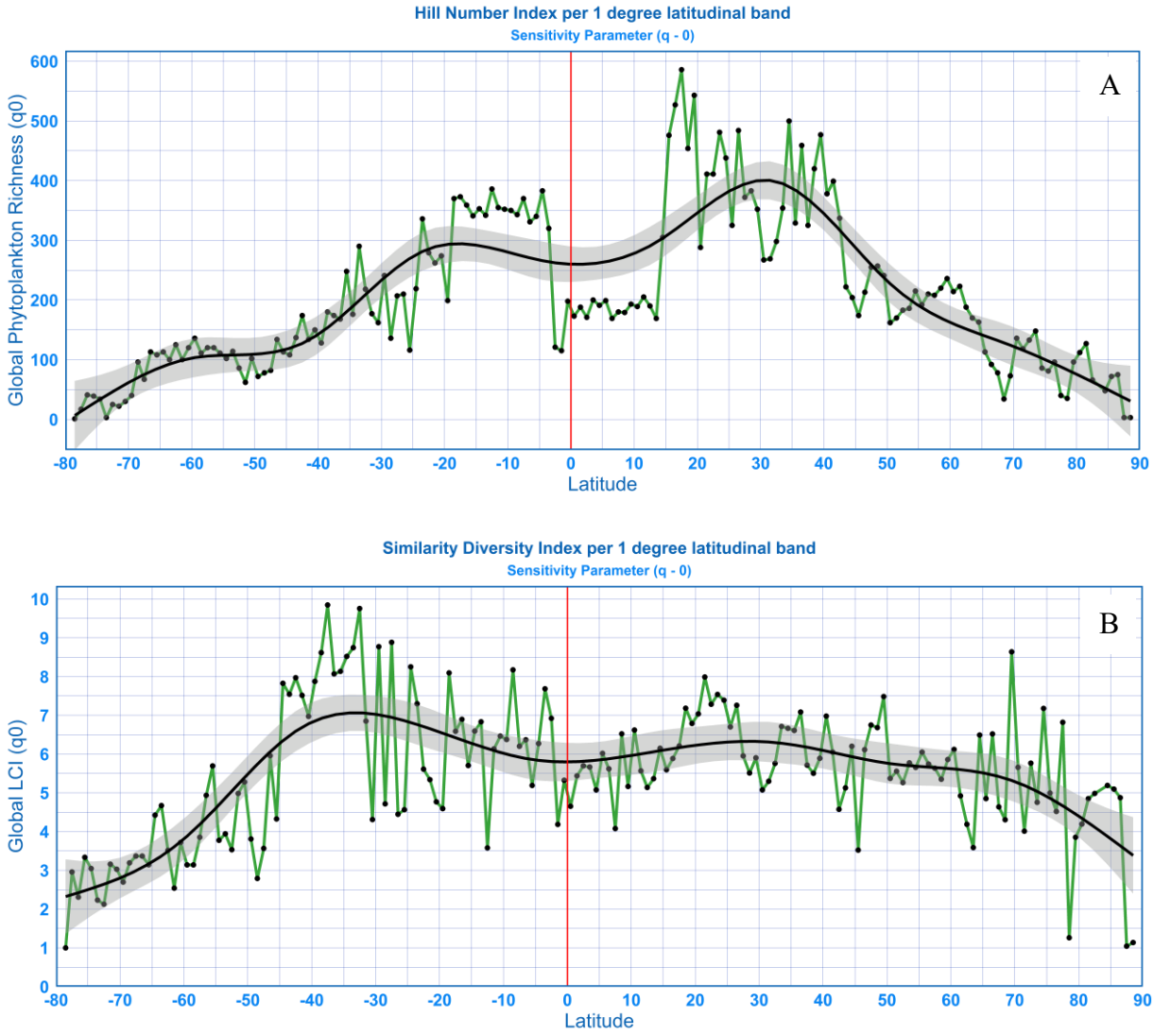


Figure 3.1 - Global marine phytoplankton similarity-sensitive richness across latitudes derived from a Generalised Additive Model (GAM) prior to sampling correction (B). Global marine phytoplankton richness across latitudes derived from a Generalised Additive Model (GAM) prior to sampling correction (A). Red vertical line denotes equator, black dots symbolise raw richness and grey shading is 95% confidence interval from GAM .

The LDG of marine phytoplankton that has similarity incorporated results in the reduction in diversity. Figure 3.2B shows the *q0,* similarity-sensitive species richness, along with the other *q* values used. This reduction in diversity is greater in the northern hemisphere. Whereas the southern hemisphere shows greater variability in response to the inclusion of similarity and evenness parameters in diversity calculation.

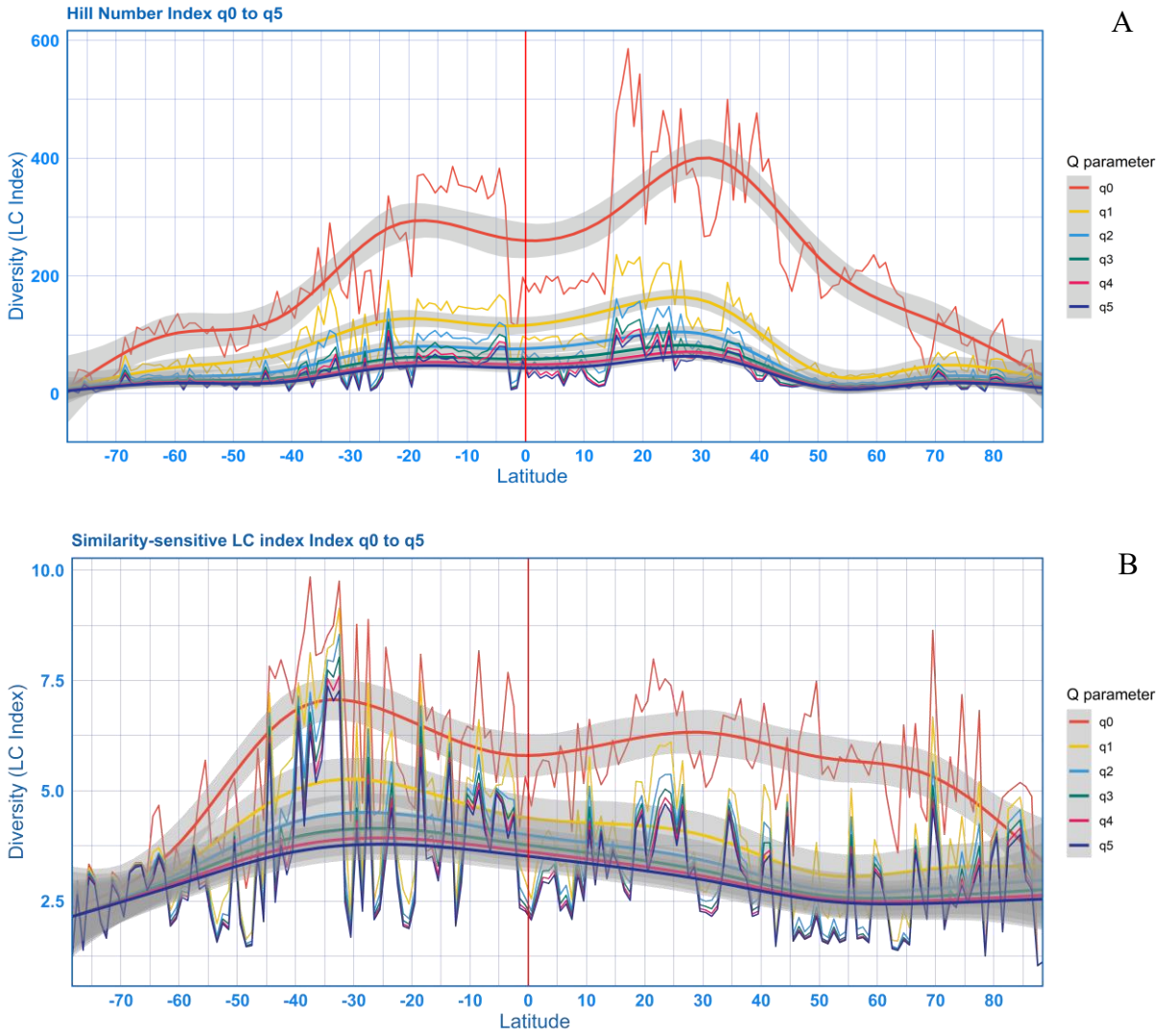


Figure 3.2 - Coarse-grain scale global marine phytoplankton diversity across LCI (Q parameter: q0–q5) (B) and the Hill number Index (A). The plot illustrates coarse-grain scale diversity per 1° latitude with a generalised additive model (GAM) smoothing function.

The comparison between the Hill number approach and the similarity-sensitive LCI (Figure 3.3) cannot be made directly due to the different diversity component weighting. The Hill numbers contain richness weighted by the relative abundance whereas the similarity-sensitive LCI is weighted by both relative abundances and the community similarity. Figure 3.3 presents an insight into the general differences depending on the number of diversity components considered. In the hill number approach (Figure 3.3A and 3.3B) as previously examined in the previous chapter, there is a bimodality of distribution with sub-tropical peaks. It is also observed that the decrease in effective number of species q increases from 0 to 1 there is a decrease due to relative weighting of rare species, which are weighted less, and the weighting of the common species are increased, this leads to a northern temperate [40°N, 60°N] decline in diversity while the bimodality remains. In the similarity-sensitive LCI, the bimodality exhibited in the Hill number approach has changed to a unimodal southern peak. The increase of the abundance sensitivity parameter, q, from 0 to 1 shows a similar general decline in diversity in the northern hemisphere, as seen in Figure 3.2.

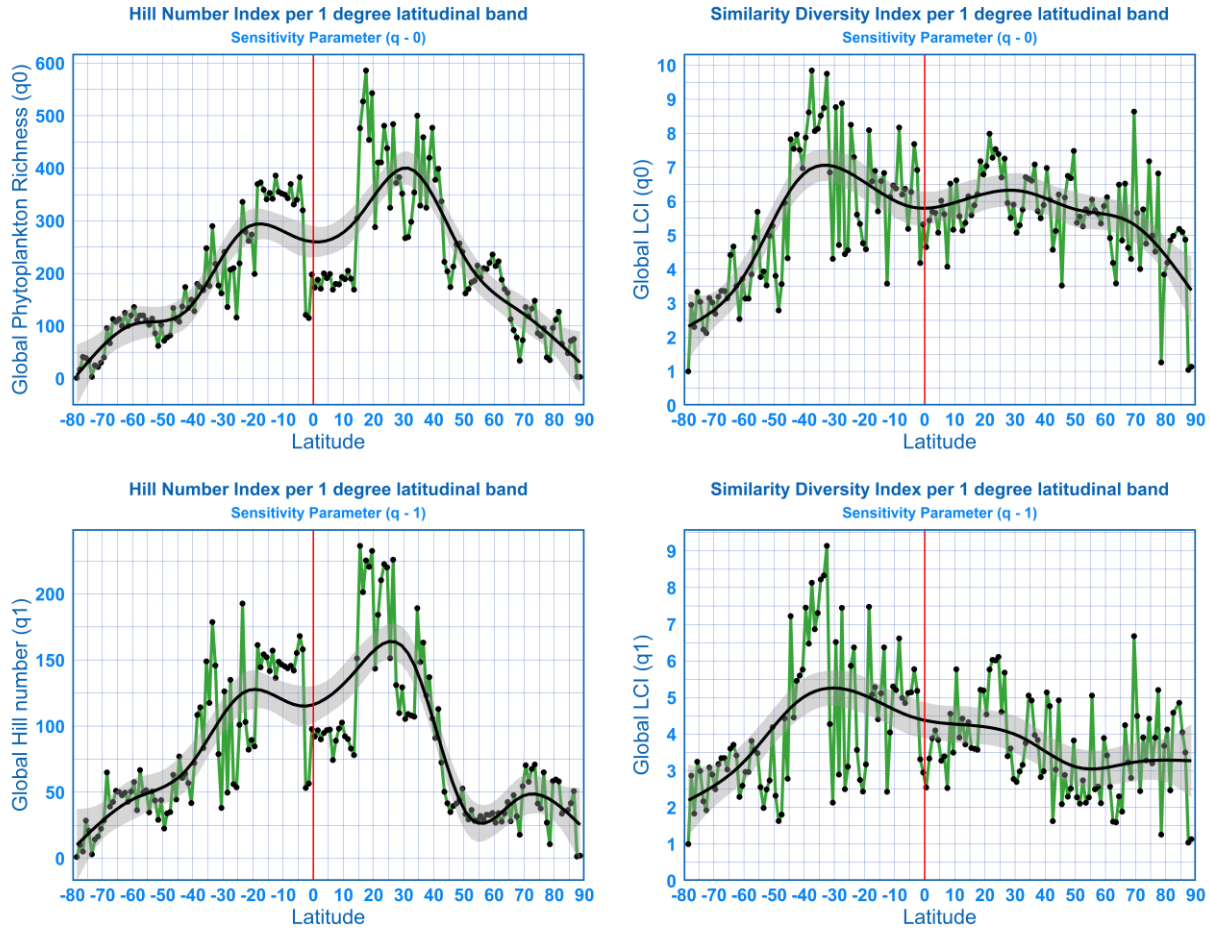


Figure 3.3 - Coarse-grain scale latitudinal diversity assessment of marine phytoplankton using LCI (A) is the naïve Leinster and Cobbold index (LCI) with no taxonomic similarity at q0, (B) is the naïve LCI with no taxonomic similarity at q1, (C) is the similarity-sensitive LCI at q0, (D) similarity sensitive at q1. The vertical red line indicates 0° latitude. Coarse-scale diversity per 1° latitude with a generalised additive model (GAM) smoothing function is illustrated.

The introduction of the similarity diversity component introduces an increase in variability to the analysis of LDG of marine phytoplankton. Figure 3.4 details a similar result of greater percentage decline in diversity in the northern temperate region as seen in the Hill number case (Figure 2.7). The difference in this assessment of percentage decline is the wider distribution in values. This decline is also reduced overall compared to the results of chapter 2, where the majority of values decreased > 75%, whereas the inclusion of similarity shows a reduced decline across latitudes with all values < 75% decline.

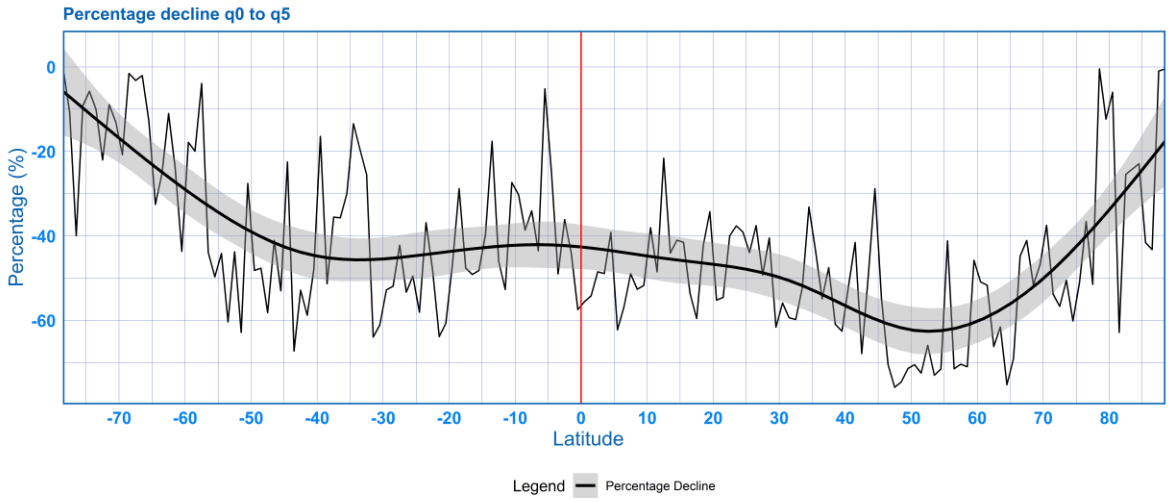


Figure 3.4 – Similarity-sensitive LCI percentage decline in diversity from q0 to q5 per latitude. The plot illustrates coarse-grain scale percentage decline from q0 to q5 per 1° latitude (thin black line) with a generalised additive model (GAM) smoothing function applied, shown as associated black line with grey shading representing 95% confidence intervals. The horizontal red line indicates the equator at 0° latitude.

The noted increase in variability is highlighted in the percentage decline verses Pielou index (Figure 3.5). The reduction in the R^2 value from 0.17 in the Hill number assessment to the 0.08 in the similarity-sensitive assessment highlights this. In general, the larger spread of percentage decline values in the similarity assessment highlights a decoupling of the relationship of diversity value from *q*0 to *q5* and the evenness of the latitudinal coarse-grain communities.

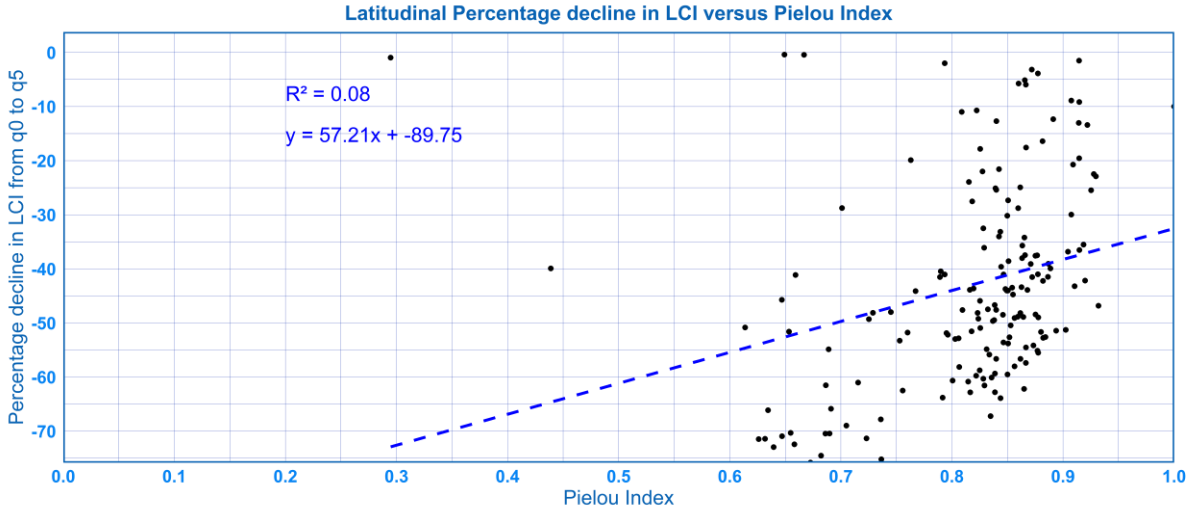


Figure 3.5 – Similarity-sensitive correlation plot for percentage decline and Pielou-index. Blue dashed line is linear relationship of variables with an R^2 value of 0.08.

3.3.2 Leinster and Cobbold index fine-scale grain assessment

The fine-scale grain scale similarity-sensitive case of the LCI at *q0* and *q1* (Figure 3.6) continues to show the prominent peak is in the southern sub-tropics and spreads into the southern temperate region as previously observed in the coarse-grain scale assessment. The distinctive northern sub-tropical peak of the Hill number assessment (Section 2.3.4) has been reduced in this fine-grain scale assessment with the inclusion of taxonomic similarity, where high similarity in taxonomic information reduces the overall diversity measured by the similarity-sensitive LCI.

Contrastingly, the distribution of marine phytoplankton similarity-sensitive LCI diversity differs from that of the Hill number cases of the LCI at fine-grain scale. In the similarity-sensitive fine-grain scale assessment, the southwestern Pacific, near Australia and the Persian Gulf system shows greater diversity with the coastal upwelling system of Peru remaining high in diversity. Additionally, the Drake Passage and Southern Ocean show moderately high diversity in the similarity-sensitive LCI diversity assessment.

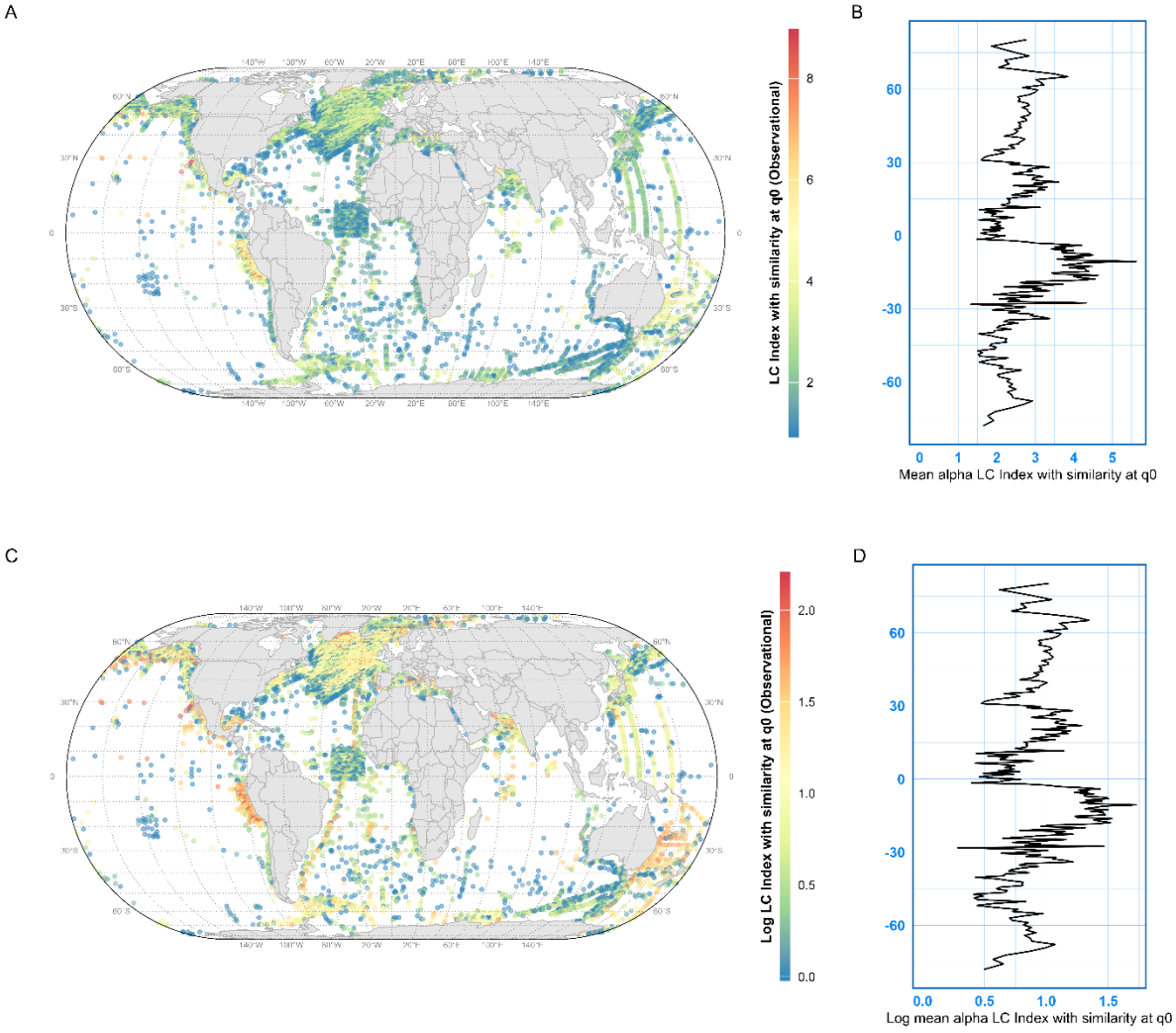


Figure 3.6 - Phytoplankton LCI diversity on fine-grain scale. (A) Marine phytoplankton diversity (LCI – q0), (B) LCI diversity using fine-grain scale, (C) Logarithmic (log) marine phytoplankton diversity (LCI – q0), (D) Log LCI diversity using fine-grain scale.

3.3.3 Chen and Grinfeld decomposition of Leinster and Cobbold index

Figure 3.7 details the naïve effective richness results from the decomposition of the LCI. The decomposed Hill number (naïve LCI) case, *q0* pattern reconstructs the species richness profile of the marine phytoplankton LDG as observed previously in Figure 2.6. The LDG has two bimodal peaks in diversity centered at the sub-tropical regions. As observed in the Hill number assessment, the southern tropical region (-20°S to the equator) and the northern temperate region [50°N, 70°N] deviate the greatest from richness (*q0*).

The similarity-sensitive decomposed LCI shows a similar pattern to the naïve case with a general exception of having a shallower decline in diversity towards the polar regions. The northern hemisphere continues to show a high deviation of diversity values measured with $q >$

0 compared to the southern hemisphere except for the southern tropical region which continues to show deviation for values of $q > 0$.

3.3.3.1 Decomposed richness

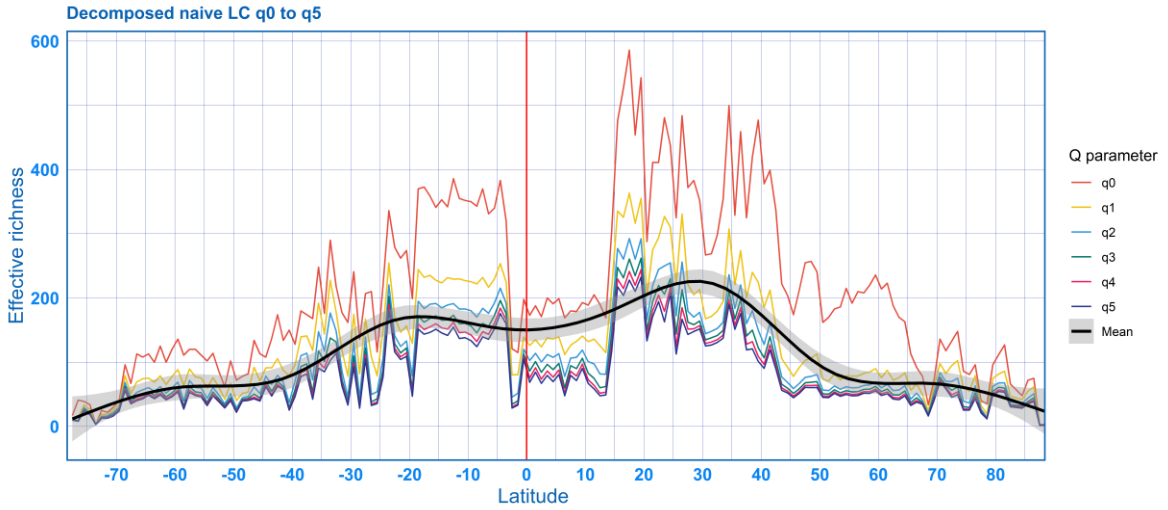


Figure 3.7 - Coarse-grain scale global marine phytoplankton Chen and Grinfeld decomposed naïve LCI effective richness. Q sensitivity parameter ranging from 0 to 5 in integers of 1 per 1° latitude (coloured lines). The black line is a generalised additive model smoothing function applied to the results of the decomposed LCI diversity assessment with grey shading detailing the 95% confidence interval across all q values. The horizontal red line shows the equatorial line at 0° latitude.

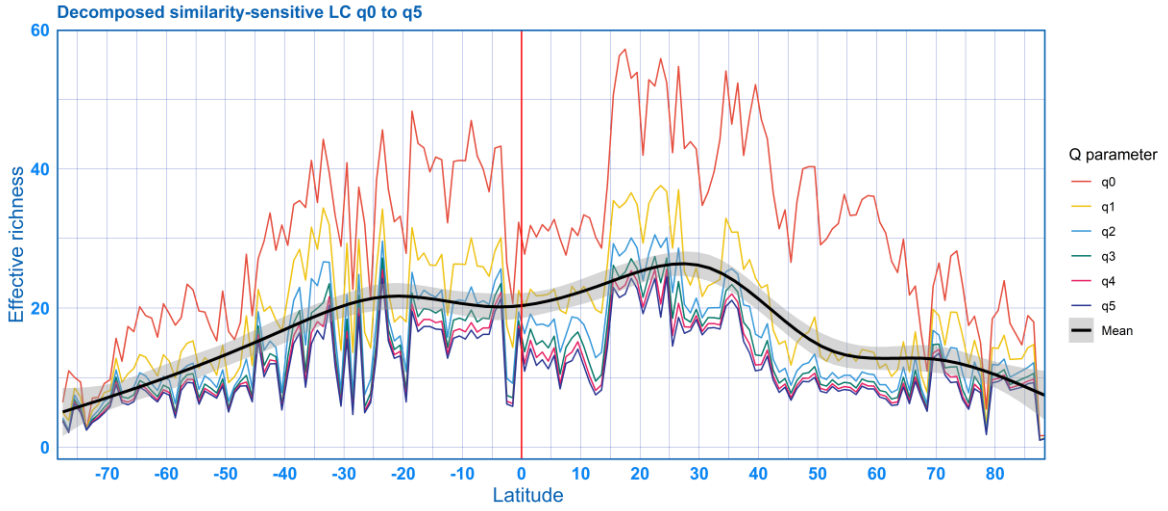


Figure 3.8 - Coarse-grain scale global marine phytoplankton Chen and Grinfeld decomposed similarity-sensitive LCI effective richness. Q sensitivity parameter ranging from 0 to 5 in integers of 1 per 1° latitude (coloured lines). The black line is a generalised additive model smoothing function applied to the results of the decomposed LCI diversity assessment with grey shading detailing the 95% confidence interval across all q values. The horizontal red line shows the equatorial line at 0° latitude.

3.3.3.2 Decomposed evenness

The decomposed LCI for the marine phytoplankton diversity (Figure 3.9 & Figure 3.10) have a comparatively similar distribution to each other. The balance, i.e. evenness, of the relative abundances of the coarse scale 1° latitudinal communities varies with latitude in both cases with a notable reduction in balance in the northern hemisphere. This northern dip ranges from [~40°N, 70°N] in both the naïve and similarity-sensitive decomposition cases. This general mirroring of diversity component responses to latitude continues here following the decomposed effective richness patterns. Similarly, there are general differences the Hill number

and the similarity-sensitive LCI cases; the similarity-sensitive case appears more variable for all latitudes. It also is shown that the southern hemisphere in general [-70°S, 0°S] appears to be more balanced than the northern hemisphere overall [0°N, 88°N]. Notably, in the decomposed naïve case at sensitivity parameter $q = 0$, the evenness of the communities is not quantified; therefore, the value for balance is maximally even at a value of 1 for all latitudes assessed.

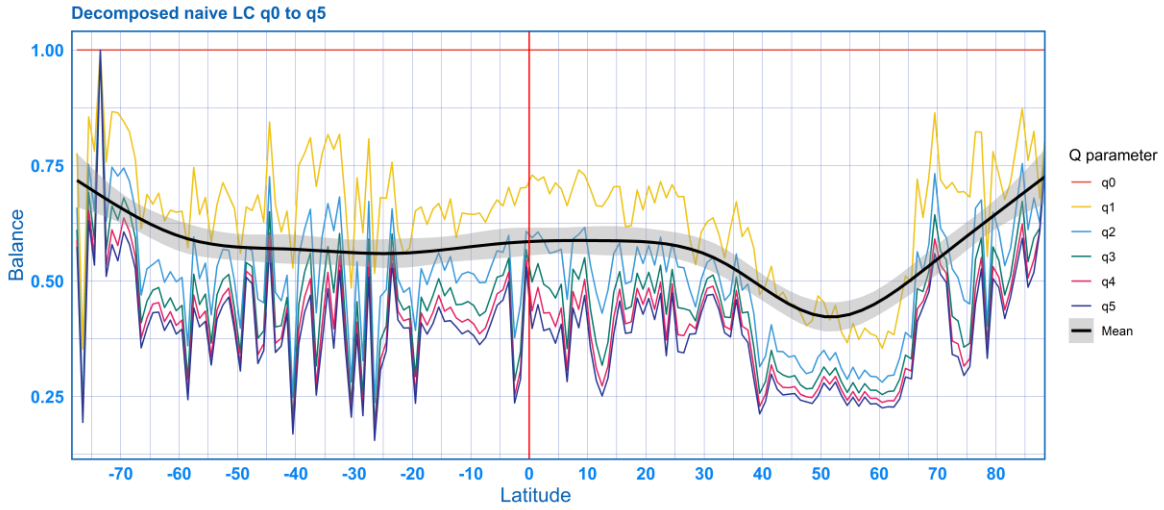


Figure 3.9 - Coarse-grain scale global marine phytoplankton Chen and Grinfeld decomposed naive LCI balance. Q sensitivity parameter ranging from 0 to 5 in integers of 1 per 1° latitude (coloured lines). The black line is a generalised additive model smoothing function applied to the results of the decomposed LCI diversity assessment with grey shading detailing the 95% confidence interval across all q values. The horizontal red line shows the equatorial line at 0° latitude.

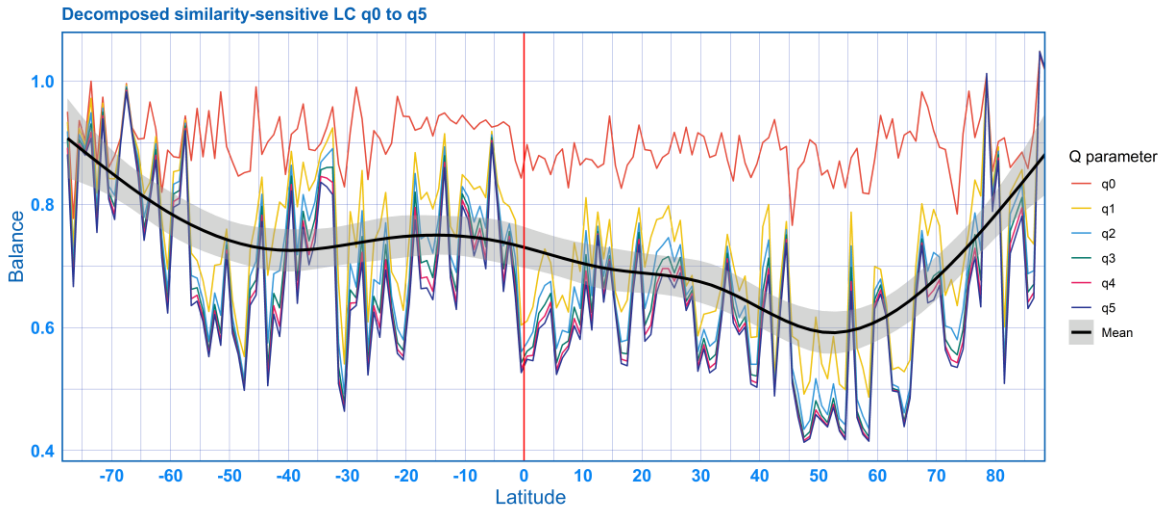


Figure 3.10 - Coarse-grain scale global marine phytoplankton Chen and Grinfeld decomposed similarity-sensitive LCI balance. Q sensitivity parameter ranging from 0 to 5 in integers of 1 per 1° latitude (coloured lines). The black line is a generalised additive model smoothing function applied to the results of the decomposed LCI diversity assessment with grey shading detailing the 95% confidence interval across all q values. The horizontal red line shows the equatorial line at 0° latitude.

3.3.3.3 Decomposed similarity

Figure 3.11 details the decomposed similarity-sensitive LCI measured dissimilarity of the global marine phytoplankton diversity. This diversity component quantifies the pairwise dissimilarity, it ranges from 0-1 where 0 represents a fully similar community, i.e. 1 species, and 1 represents a fully dissimilar community of unrelated species. From this assessment it can

be observed that the southern hemisphere, despite the general variability throughout the assessment, southern marine phytoplankton communities are more dissimilar. Notably, the latitudinal zone of [-50°S, -25°S] representing the southern temperate region and some of the sub-tropics sees a general increase in dissimilarity compared to that of the northern hemisphere. In the northern hemisphere, the region of [20°N, 40°N], the northern sub-tropics is notably more similar than all other latitudes globally.

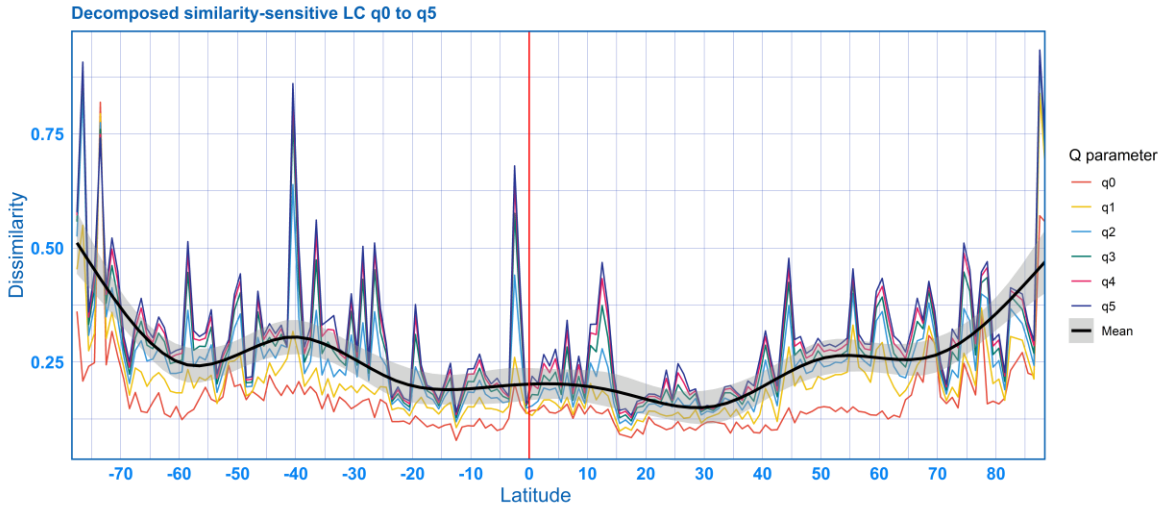


Figure 3.11 - Coarse-grain scale global marine phytoplankton Chen and Grinfeld decomposed LCI Dissimilarity (Chen & Grinfeld, 2024; Leinster & Cobbold, 2012). This graph details the coarse-grain scale marine phytoplankton decomposed dissimilarity assessment with sensitivity parameters ranging from 0 to 5 in integers of 1 per 1° latitude (coloured lines). The black line is a generalised additive model smoothing function applied to the results of the decomposed LCI diversity assessment with grey shading detailing the 95% confidence interval. The horizontal red line shows the equatorial line at 0° latitude.

3.3.4 Latitudinal taxonomic composition and evenness

As observed in Figure 3.2, the Hill number diversity of marine phytoplankton shows a bimodal distribution. In figure 3.12A, this pattern is shown again with species richness represented as a stacked bar plot. The phylum composition of each individual species at coarse-grain scale, the latitudinal communities are mostly composed of three phyla; Bacillariophyta, Miazoa and Haptophyta.

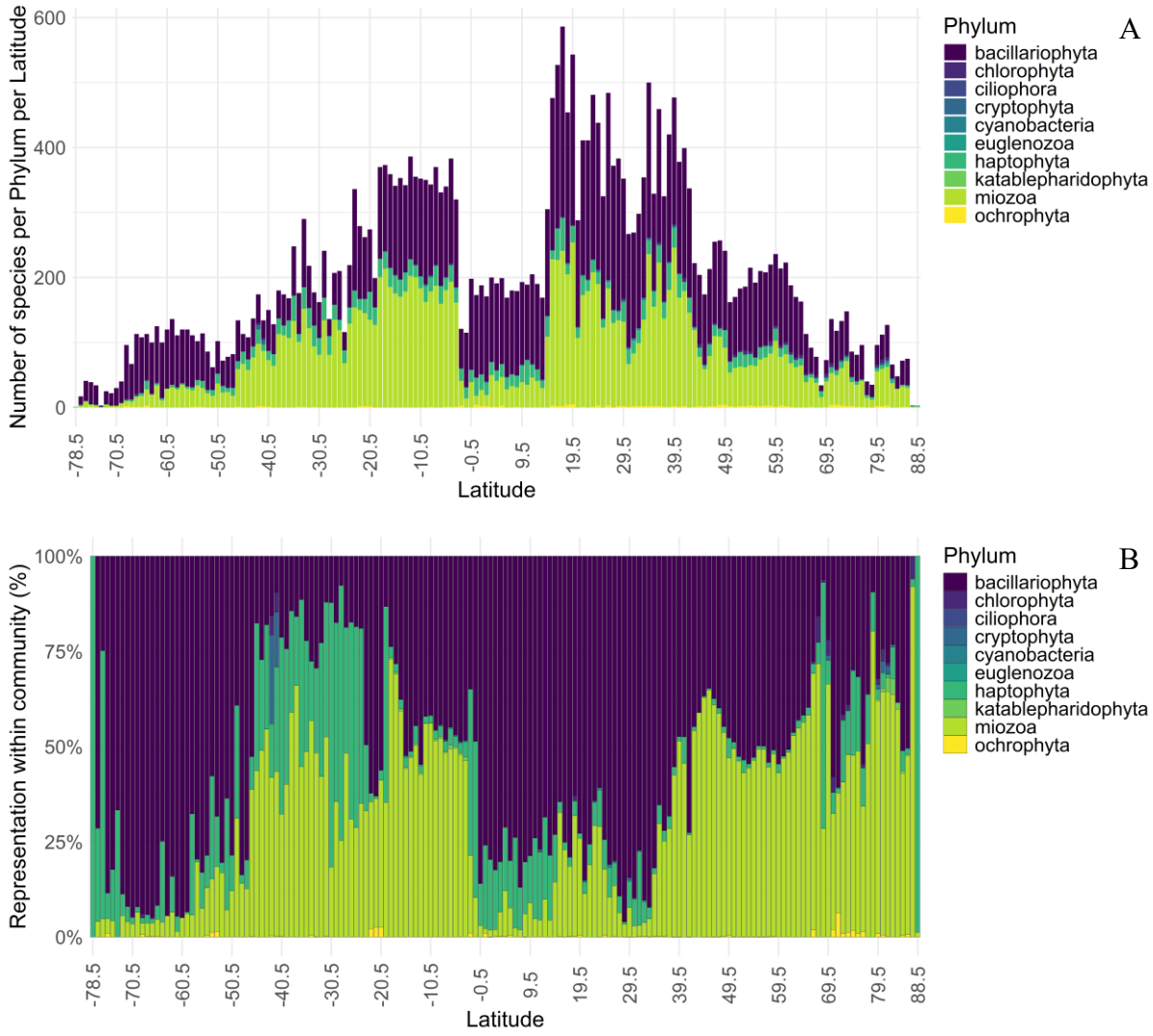


Figure 3.12 – Coarse-grain scale global marine phytoplankton species count per phylum (A). Coarse-grain scale global marine phytoplankton taxonomic (phylum) individual composition represented as a percentage (B).

This pattern hides the number of individuals contained within each coarse-grain community as it does not show the number of individuals represented by each phylum. When the phylum representation is shown as a percentage of the community individuals at the coarse-grain scale, this pattern shifts. As seen in figure 3.12A, although the number of species for Haptophyta is low in the southern temperate region between [-50.5°S, -20.5°S], the number of individuals is high for that same region in figure 3.12B. Species of the same phyla are more likely to have similar genera, family, order and class and as such are more similar. As can be seen in figure 3.12A, the region of the northern temperate between [19.5°N, 59.5°N] is comprised of two genera, Bacillariophyta and Miozoa. This when, compared to the three genera in the southern

temperate region, Bacillariophyta, Haptophyta and Miozoa is less diverse in phylum by comparison. Moreso, the evenness of the phylum distribution in the northern hemisphere is typically less even with Bacillariophyta heavily represented by individuals between [-0.5°S, 29.5°N].

To examine in further detail, calculating Pielou's evenness for each taxonomic level, where a value of 1 is maximally even, for all levels of taxonomy is more even. This is also shown in the decomposed balance data in figure 3.10, where the decomposed balance is greatest is highest between [-50°S, -20°S] and the general diversity values described in figure 3.1. In particular, the distribution of phylum across species (pink line) is shown to be most even in distribution in this region.

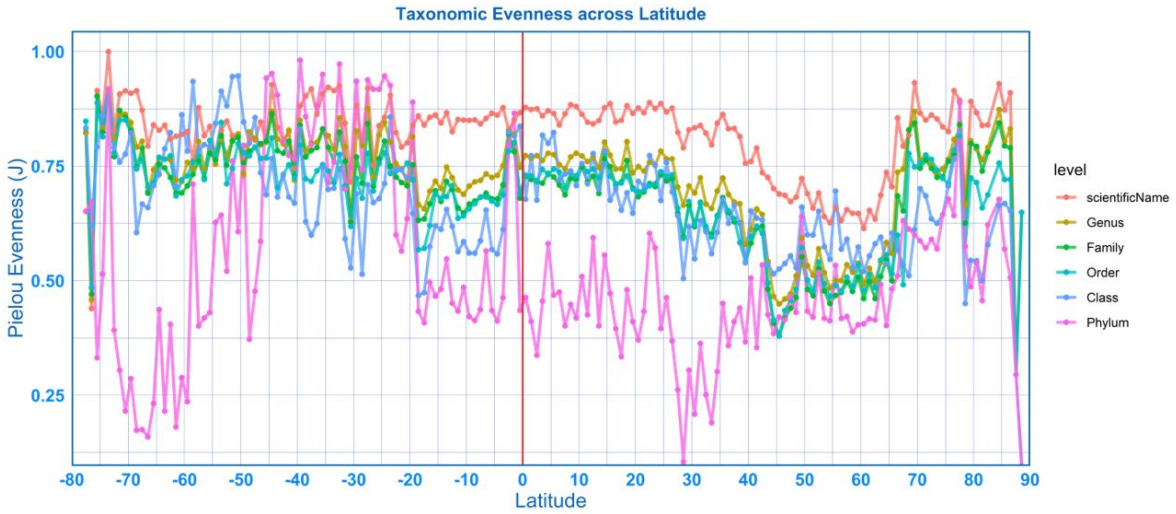


Figure 3.13 - Coarse-grain scale global marine phytoplankton taxonomic Pielou's evenness per taxonomic level across latitude.

As a result, the LDG of marine phytoplankton changes its modality and peaks in the southern hemisphere when species richness, evenness and taxonomic similarity is considered. While richness in northern and southern hemispheres is comparable, the southern hemisphere is taxonomically more diverse, through greater evenness and higher diversity of taxonomy represented in the southern region.

3.3.5 Sensitivity of marine phytoplankton LDG pattern with data removal

To examine the effect of sampling effort on the assessed LDG pattern of marine phytoplankton. The Sir Alister Hardy Foundation for Ocean Science (SAHFOS) operates the extensive CPR sampling regime in the North Atlantic. In the PhytoBase dataset the number of observations with this institution code is 281,722 (22.23% of the total observations: 1,267,490). While the SAHFOS dataset contained 263 species, there is only 3 unique species contained within this subset of data. After removal the dataset comprised of 958,768 observations with 1558 species. The resulting change in the LDG of marine phytoplankton after this removal is minimal (Figure 3.14A and B). There is a decline in species richness between [65°N, 70°N] but overall, the pattern remains largely unchanged in species richness distribution.

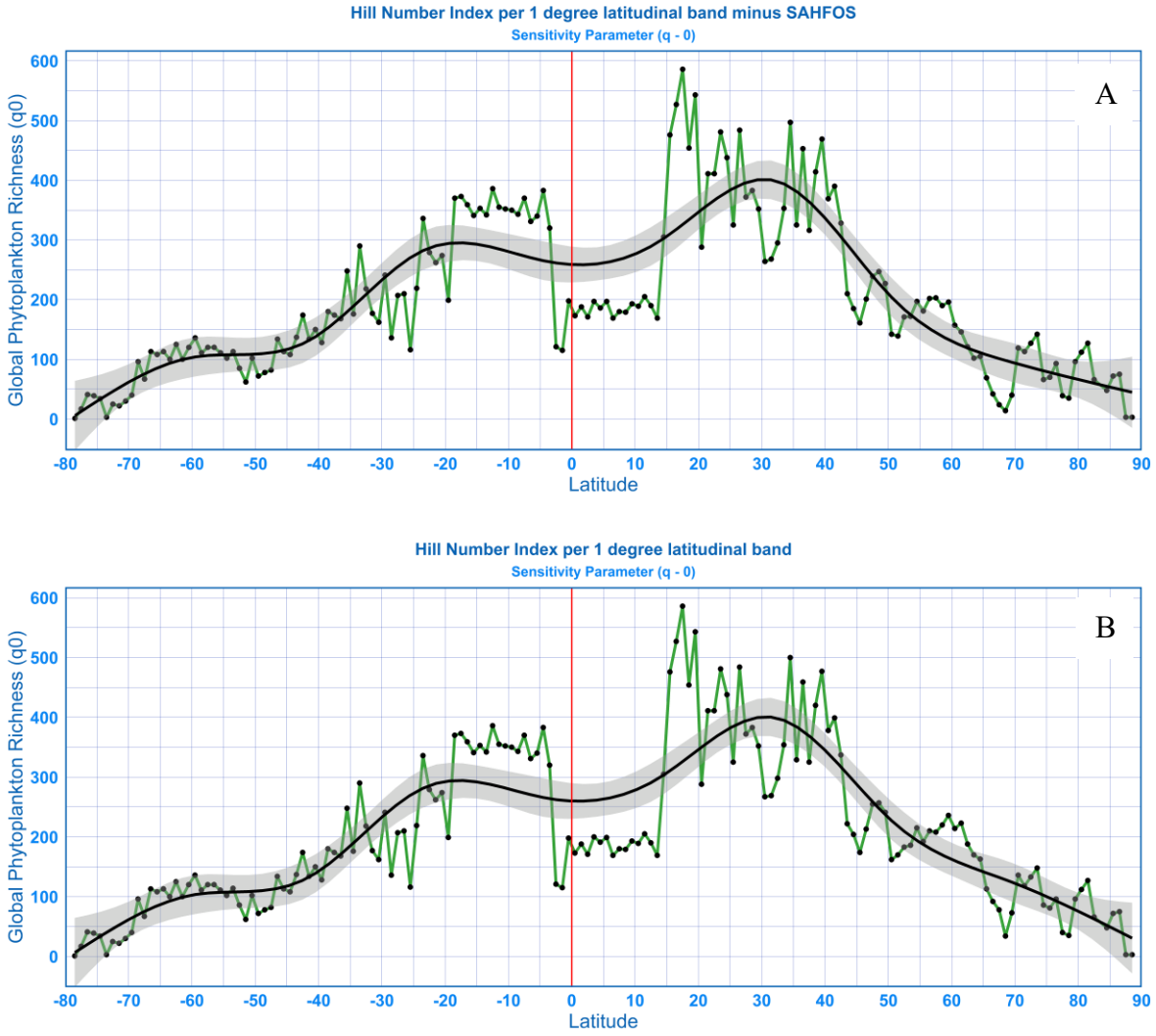


Figure 3.14 - Global marine phytoplankton species richness across latitudes with SAFHOS data removed derived from a Generalised Additive Model (GAM) prior to sampling correction (A). Global marine phytoplankton richness across latitudes derived from a Generalised Additive Model (GAM) (B). Red vertical line denotes equator, black dots symbolise raw richness and grey shading is 95% confidence interval from GAM.

The similarity sensitive results using the LCI also shows minimal deviation from the full analysis after the removal of the SAHFOS dataset (Figure 3.15A and B). There is a decrease in similarity-sensitive richness between [65°N, 70°N] again, and an increase in diversity between [70°N, 80°N]. In addition, the largest change is the overall value of diversity across all latitudes reducing from an effective diversity number of 9.8 species maximum in Figure 3.15A to 7.7 in Figure 3.15B.

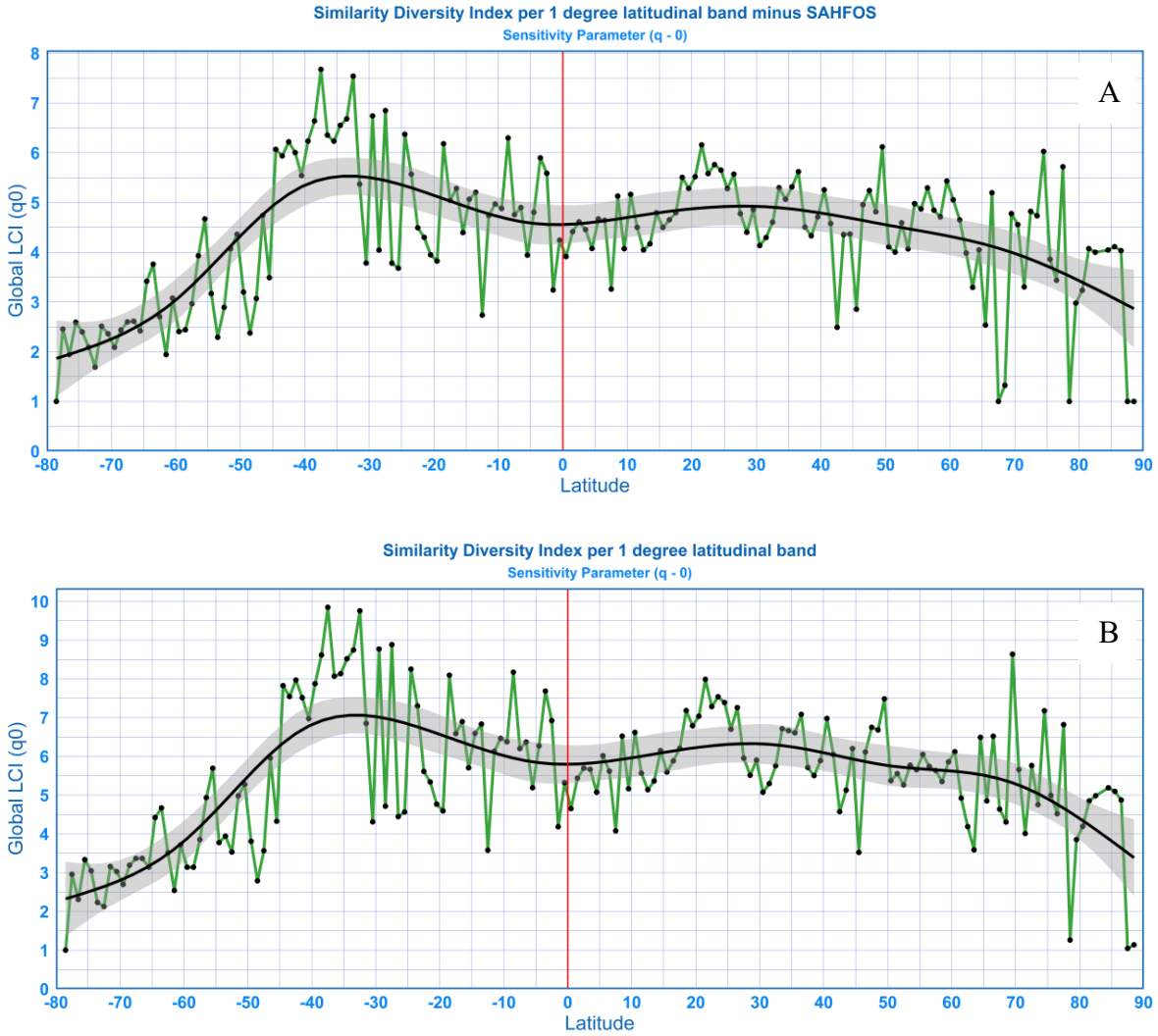


Figure 3.15 - Global marine phytoplankton similarity-sensitive richness across latitudes with SAFHOS data removed derived from a Generalised Additive Model (GAM) prior to sampling correction (A). Global marine phytoplankton richness across latitudes derived from a Generalised Additive Model (GAM) (B). Red vertical line denotes equator, black dots symbolise raw richness and grey shading is 95% confidence interval from GAM.

3.3.6 Similarity-sensitive LCI Rarefaction

Figure 3.16, shows how the data was split for the rarefaction diversity reassessment. The southern temperate and polar region (STemp, blue) spans [-78.04°S, -40°S]. The southern tropical and sub-tropical region (STrop) is covered by the next region between latitudes, [-40°S, 0°]. The northern tropical and sub-tropical region (NTrop) is bounded by [0°N, 40°N] latitudes. Lastly, the northern temperate and polar region (NTemp) is covered by the truncated latitudes of [40°N, 80°N].

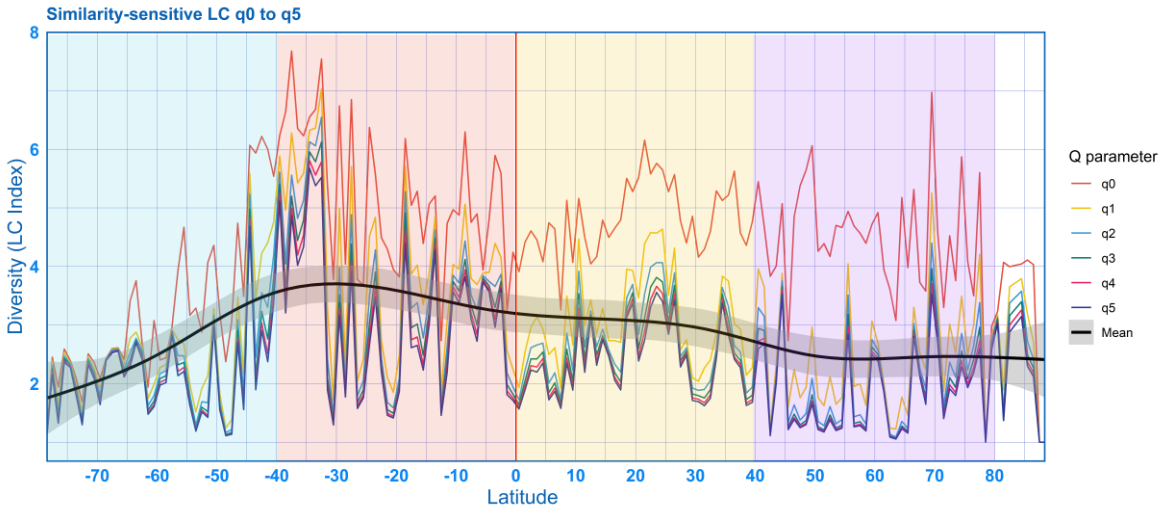


Figure 3.16 - Similarity-sensitive LC index LDG assessment. Colour coded for the rarefaction assessment regions. Blue block is southern temperate, red block is southern tropical, yellow block is north tropical and the purple block is north temperate.

The rarefaction results for each region and the suite of similarity-sensitive LCI across *q* are shown in Figure 3.17 and Figure 3.18. The region with the greatest similarity-sensitive richness is the southern temperate region, estimating an effective number of 12 species sampled in the first 1000 random observations selected (10 replicates)(Figure 3.17). The lowest similarity-sensitive richness is the northern temperate region. Both tropical regions show similar values when similarity is considered. Notably all the similarity-sensitive rarefaction assessments saturate at or near a visual asymptote within this sub-sampling regime. The naïve Hill number rarefaction assessment of the previous chapter showed less asymptotic curves.

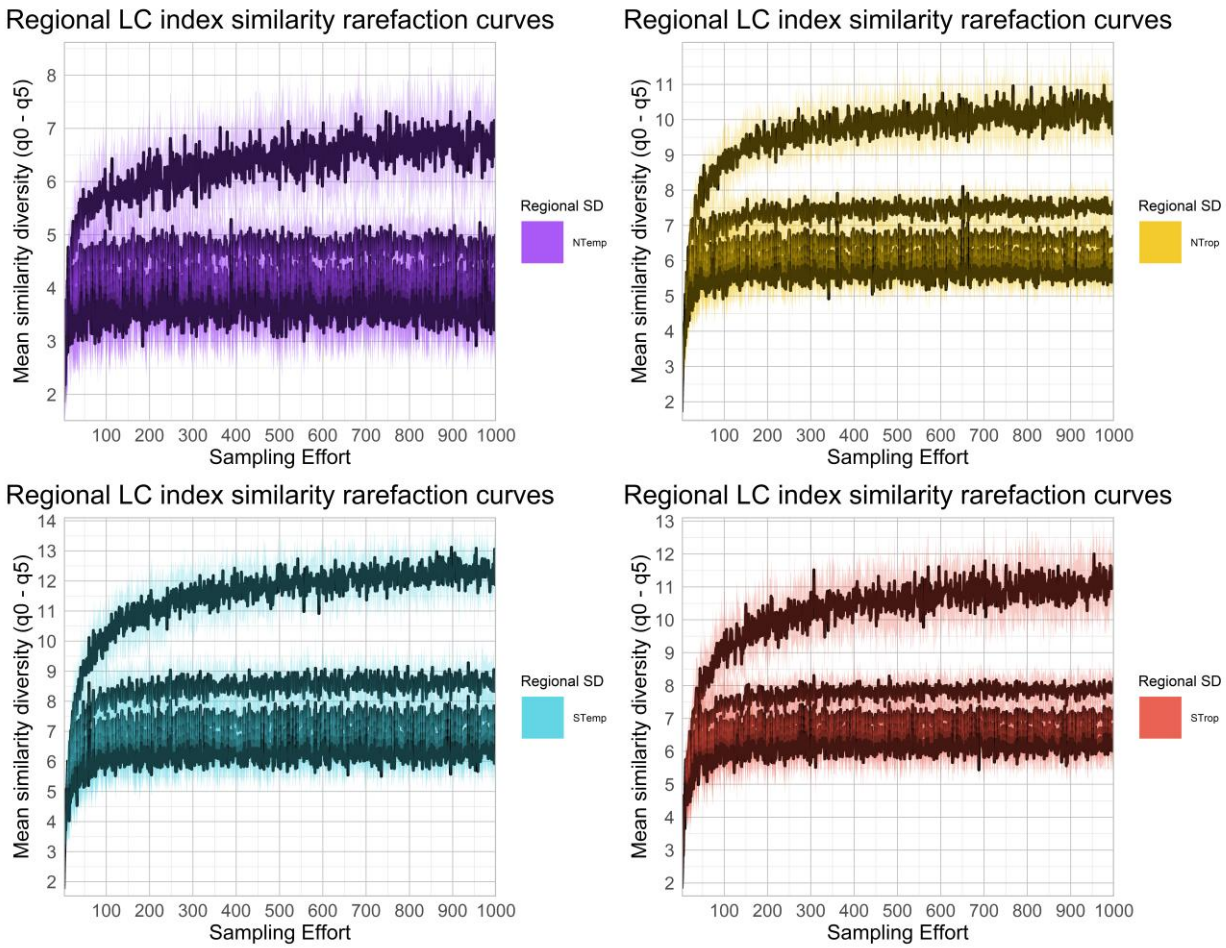


Figure 3.17 - Similarity-sensitive rarefaction assessment from 0 to 1000 samples repeated 10 times. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

As observed in Figure 3.17, the southern temperate and tropical regions detail the greatest similarity-sensitive richness, with the northern temperate region detailing the lowest.

The 0 to 20,000 discontinuous rarefaction results for each region and suite of LCI is shown in Figure 3.18. Here it is further confirmed that the largest rarified richness is the southern temperate region with a estimation that >13 species are sampled in 20,000 random observation samples (replicated 50 times). The lowest richness rarefied is the northern temperate region showing > 7 species.

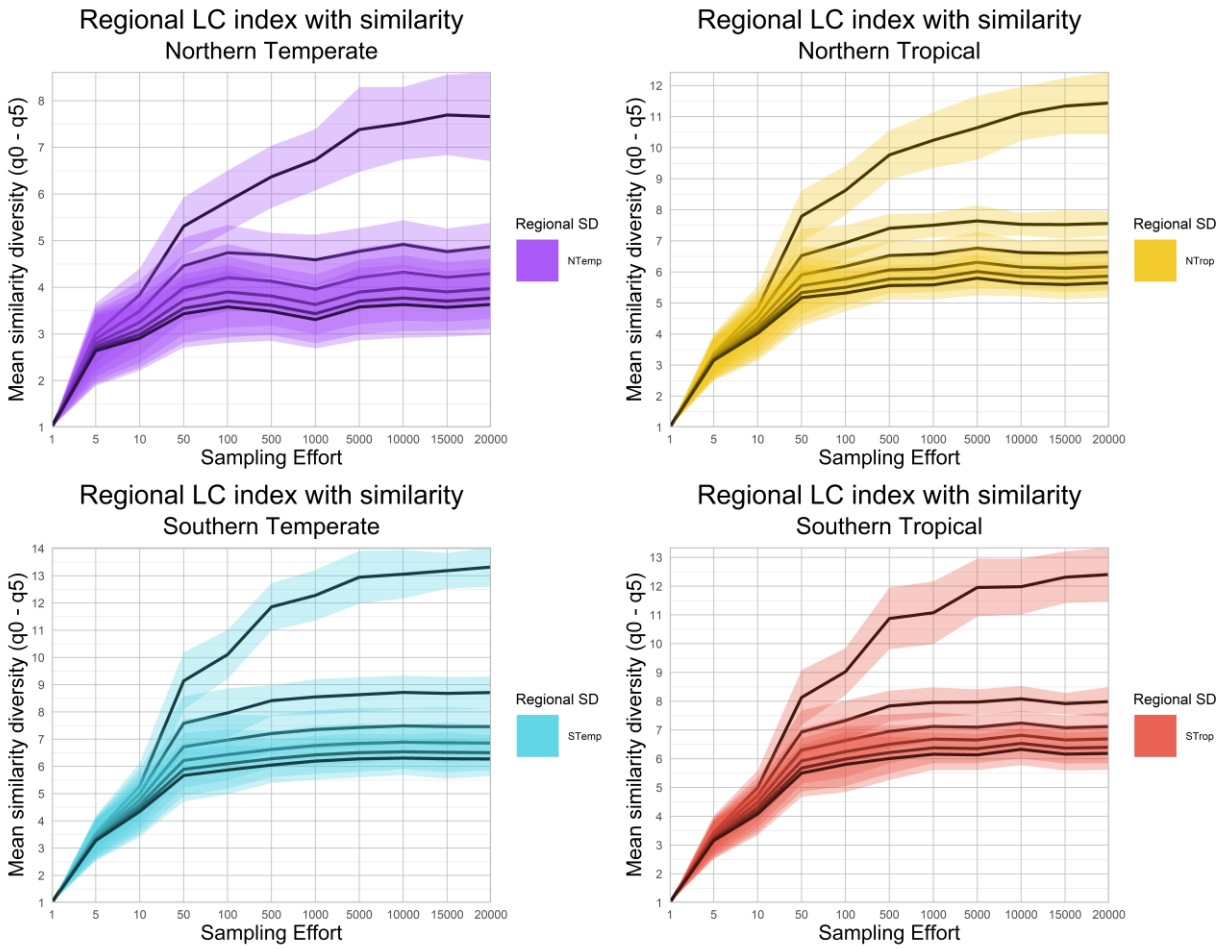


Figure 3.18 -Similarity-sensitive rarefaction assessment from 0 to 20,000 samples (discontinuous, 50 replicates. Colour coded for the rarefaction assessment regions. Blue is southern temperate, red is southern tropical, yellow is north tropical and the purple is north temperate. Each curve on each plot in descending order of value is q0 to q5 respectively.

3.4 Discussion

Here, I sought to examine the widely and thoroughly interpreted pattern of the LDG observed in other groups using a novel similarity-sensitive diversity index (Leinster & Cobbold, 2012a) and a global marine phytoplankton dataset. The aim of this work was to examine what effect the inclusion of taxonomic similarity had on the LDG of marine phytoplankton, if there was no deviation from the species richness pattern. I expected that considering similarity held no benefit at this scale. Yet this is not what I observed, there was a substantial shift in the LDG pattern of marine phytoplankton from bimodal to unimodal distribution, from equatorial focus to a southern hemisphere dominance. This answered my initial exploratory question of does the incorporation of community similarity in our diversity assessments change our perspective of the LDG. The implications for the why this occurs and how it occurs highlights the limitations of work at this scale. The effect of sampling bias on this diversity index is not currently clear; does community similarity respond to effort in the same way as richness is a key question for instance.

My findings provide a general picture of what is currently observed globally for marine phytoplankton diversity. It is likely that future sampling effort may change this LDG pattern observed here (Menegotto & Rangel, 2018) and the saturation curve drawn in FFigure 2.3. The LDG assessments of marine phytoplankton performed in this chapter and the decomposition of the LCI provide baselines to develop future analysis and diversity assessments which consider a broader definition of diversity, providing an evolutionary grounded perspective by including species relatedness. To date there have been very limited global assessments of marine phytoplankton diversity (Righetti et al., 2019, 2023), often fragmented latitudinal studies (Chust et al., 2013; O'Brien et al., 2016; Olguín Salinas et al., 2015) and differing time points and subgroups of phytoplankton assessed resulting in difficulties aligning general patterns or environmental responses for all marine phytoplankton. This work adds to the global perspective of marine phytoplankton diversity in a substantial manner, with the aim of resolving LDG disparities in observational assessments of marine phytoplankton.

It is not possible to directly compare the novel diversity assessment to existing LDG patterns, as this approach incorporates richness, evenness and similarity. At least from the general assessment method of the LCI, my decomposition work does allow this comparison to be made in a more direct manner. The results of the Hill number species richness decomposition, obtained by using a naïve assumption of similarity, prompted comparing the LDG species richness patterns found in this study with others. What I found is that there is a bimodal distribution for the marine phytoplankton LDG which has been observed and explained as a response to climate-driven temperatures, niche truncation at the equator and reduced diversity (Chaudhary et al., 2021) or by lack of sampling effort in the tropics (Menegotto & Rangel, 2018). The disparity in these explanations is cause for concern, as they have different implications for next steps. Either the LDG is truly equatorially unimodal and as such models approximating this macroecological diversity should reflect this reality (Righetti et al., 2019) or global diversity is reshaped actively by human impacts to the potential detriment or change of global ecological function.

While climate-driven species distribution shifts are altering marine ecosystems globally (Barton et al., 2016; Perry et al., 2005), global changes in general species richness patterns add an additional change in marine ecosystems as a result of human activities (Tittensor et al., 2021). As for how my similarity-sensitive latitudinal relationship relates to potential climate-driven species shifts and wider LDG patterns is not directly apparent at this stage. The novelty of this approach at this scale presents the groundwork from which these important relationships should be resolved or considered in the field of macroecology.

Specifically, these findings raise important questions surrounding what the general LDG distribution is, unimodal or bimodal, and what implications this has on the drivers of macroecological diversity patterns. Generally, biological interactions which also constrain species distributions are difficult to resolve at macroecological scales for specious taxa (Early & Keith, 2019), therefore abiotic variables are commonly used to infer and draw correlations in species distributions (Guisan et al., 2017). When considering the modality of the LDG, if the unimodal peak at the equator is the expected LDG shape, this directly affects which abiotic variables correlate with the LDG of a taxon (Rohde, 1992). This would indicate a different mechanism is leading to the change in the modality of the LDG observed. Further obfuscated by differences in sampling effort and methodologies across the PhytoBase dataset (Menegotto & Rangel, 2018).

Furthermore, a key finding is the alpha diversity giving rise to longitudinal diversity distributions are important and are likely to be more prescriptive of drivers of diversity beyond the sanitised latitudinal relationship often sought (Procheş et al., 2023). As the results of the alpha diversity assessments show, coastal regions may be hotspots of diversity for species richness although sampling effort spatial bias also contributes to high richness estimates (Figure 3.6). My findings complicate the latitudinal perspective by including two more diversity components to consider leading to changes in the diversity hotspots of marine phytoplankton (Section 3.3.2). However, when considering the lack of agreement on drivers of species richness to date (Dutkiewicz et al., 2020; Righetti et al., 2019; Tittensor et al., 2010b; Visser et al., 2014), which is known to likely arise from a multitude of factors in tandem, the inclusion of more information channels of diversity information may lead to new hypothesises which complement ecological understanding.

Existing theories for the formation of the LDG broadly point towards thermally affected metabolic rates with richness (Brown et al., 2004; Righetti et al., 2019; Tittensor et al., 2010b; Tittensor & Worm, 2016) , yet no currently held general theories of macroecological scale consider evenness and similarity in tandem with richness. Similarity is not fully translated to global natural systems and settings as of yet and work remains to extend neutral theory of ecology into real-world dataset approximation and modelling (Hubbell, 2001; Tittensor & Worm, 2016). These results dispute this when similarity is considered. If temperature was the core driver of diversity, then a unimodal distribution should also arise from the similarity inclusive assessment. However, I found a southern peak where the key difference in hemisphere is circulation and potential connectivity between marine phytoplankton is increased (Ward et al., 2021).

An additional finding was a southernly shift in the diversity maxima using the similarity-sensitive LCI from the equatorial centered sub-tropical peaks observed in the Hill number assessment. The drive of this shift was explored using the LCI decomposition method (Chen & Grinfeld, 2024). The inclusion of the similarity matrix in the LCI leads to a change in the modality of the LDG of marine phytoplankton. This matrix interacts with the relative abundance and richness of marine phytoplankton communities to scale the diversity in a way that does not occur in the Hill number approach (Leinster & Cobbold, 2012a). The reason for this shift is unclear and requires deeper exploration as the sensitivity of the LCI to changes of each diversity component requires further research. The difficulty is the interactive relationship of the three different components; richness, evenness and similarity (Daly et al., 2018), as observed in the decomposition work. Each component exhibits a different pattern latitudinally, but it is not clear, what is ecologically driving the changes of each of these components individually at a latitudinal scale. The data does indicate that the taxonomic evenness and dissimilarity of the taxonomy in the southern hemisphere is greatest, leading to higher LCI effective numbers of species and peak in the southern hemisphere. Further methodological and scientific data exploration is required to access specific changes in these diversity components at sufficient spatial and temporal detail to explain why this shift occurs.

There are potential ecological explanations, species distribution modelling has indicated marine phytoplankton communities of temperate regions have greater species turnover than the low latitudes of the equatorial regions (Righetti et al., 2019). These communities are situated in locations of greater seasonality than their tropical counterparts (Acevedo-Trejos et al., 2015a). Additionally, the southern hemisphere differs from the northern hemisphere in two notable ways. Firstly, the unique ocean circulation of the southern ocean and the Antarctic circumpolar current (ACC) is a continuous oceanographic transport system which connects the southern Pacific, Indian, Atlantic and Southern Oceans (Sanches et al., 2016). This is not the case in the northern hemisphere where the circulatory system is constrained by the Arctic ocean seasonal icesheet and the shallow Bering Strait channel in the Bering Sea connecting the Atlantic and Pacific oceans (Timmermans & Marshall, 2020). Which directly affects phytoplankton population connectivity and community diversity. This may contribute to the lower taxonomic diversity in the northern hemisphere compared to that of the southern hemisphere. Secondly, the southern hemisphere has two prominent upwelling regions; Humboldt current system on the coastal shelf of Peru and the Benguela Current system on the coastal shelf system of Angola. Upwelling systems and their role in the diversity of marine phytoplankton has been modelled and shown to increase diversity and turnover due to variability in nutrient availability selecting for different size proxies of marine phytoplankton (Dutkiewicz et al., 2020). The increased species turnover in the southern hemisphere and the observed higher similarity-sensitive LCI diversity values could be a direct result of the ACC connectivity in the southern hemispheric oceans. The increased connectivity leads to greater dispersal capacity of marine phytoplankton species from the Pacific, Indian, Atlantic and possibly Southern Ocean. As the likelihood of communities being dissimilar increases as distance increases (Cermeño et al., 2010), if distinct communities in each of these oceans are connected, it could be hypothesised that the taxonomic diversity is greater in the southern hemisphere as observed in the decomposition of the LCI (Figure 3.11). What the decomposition

does highlight is the potential sensitivity of the index to potential changes in similarity and evenness resulting in large changes in the overall diversity measurement. Future work should investigate the sensitivity of these higher information indices to each component and the possible implications this has for interpretation of differences in diversity measured.

Here we have assessed three components of diversity and examined the potential relationships between these and sampling effort. This early work and exploration will open research avenues into examining these relationships in greater detail and how these relationships change through time. Furthermore, these findings highlight the necessity to consider similarity and sampling effort in our diversity assessments as it provides an ecological nuance to our measurement of diversity which has been missing until now. This chapter expanded substantially the perspective of the marine phytoplankton LDG through the incorporation of similarity. What remains to be seen is how the LDG is shaped when we model the distributions of species with environmental relationships using SDM's to account for the latitudinal and longitudinal sampling in tandem. There is also the expansion of the widely used species distribution framework and a similarity-sensitive diversity index to assess how we bridge the gap between 'traditional' species richness distributions towards the more ecologically grounded similarity-sensitive assessment.

3.4.1 Limitations and assumptions

The limitations of this study primarily fall into two categories: practical constraints related to data coverage and diversity estimation, and theoretical challenges arising from the novelty of the approach. The drawback of this novel approach is the interpretation of what the measured value of diversity represents in practice in the context of sampling effort. Specifically what diversity component of the community assessed; richness, evenness or similarity, leads to changes in the diversity measured by the LCI. For example, a community of high evenness but low similarity could lead to a LCI value of equal magnitude of a community with low evenness but high similarity given the same measurement of richness. Comparatively, species richness is readily understood as a discrete unit of diversity allowing direct comparisons in time and geographic locations. While the more complex measurements of diversity which include evenness may face the same drawback as the LCI the difference is that evenness can be measured independently with indices such as the Pielou index or assessing ranked abundance profiles. Therefore, the LCI presents a challenge but decomposition methods noted earlier reduce the challenge of interpretation substantially.

From a theoretical standpoint, this work represents the first application of a similarity-sensitive diversity index to the latitudinal diversity gradient (LDG) problem. As a result, interpretation presents a challenge. A key issue is that the effect of common biases, such as sampling effort, on the measured similarity of a community is not well understood. This is particularly relevant when considering evenness and similarity independently of richness. Both are continuous functions of community composition and contribute to diversity in different ways. While high evenness typically increases diversity in traditional indices, especially when richness is moderate, high species similarity can reduce diversity, even under conditions of high evenness. Consequently, it is not immediately clear whether a lower-than-expected diversity value is

driven by low evenness or high similarity, making it difficult to interpret LDG patterns in marine phytoplankton. The decomposition work of Chen and Grinfeld has helped clarify these challenges, significantly mitigating this limitation. However, further investigation is required to extract specific insights into local changes in the three diversity components. Additionally, this study is limited to a single time frame, providing no insight into how richness, evenness, or similarity may change over time.

There are four methodological choices which likely influence the interpretation of these results. The effect of area on diversity measurement which is contributed by the choice of area delineation in our analysis. Area of assessment has been shown to directly affect the estimation of diversity, where as area increases so does the number of species observe. This occurs colinearly with the number of samples in area. The effect of our choice to use 1-degree latitudinal area in our coarse-grain area assessment is directly influenced by the area effect with lower latitudes having larger areas than that of higher latitudes due to the spherical shape of Earth. We aimed to rectify this by assessing diversity on fin-grain scales with an equal area grid and the later rarefaction work. The area relationship with diversity is common and we have not approached scaling our diversity relationship with latitude directly, but this is an avenue for future research.

The second choice is this definition of taxonomy as our quantification of similarity. There are a wide array of similarity definitions which will likely lead to difference in the LDG of marine phytoplankton. We selected taxonomy as it is directly comparable in readily comparable ‘chunks’ of information. There is a methodological complexity that is introduced when you move from a fixed discrete comparison to a continuous comparison such as body size. Other definitions of similarity such as genetic may be more readily integrated if these similarities are available as percentages of similarity between species.

The third choice is the compression of time into a single timepoint and deriving relative abundance estimates across time therefore reducing the ecological-environmental averages to general averages. Due to the disparity in general effort over time across space, comparing and assessing the diversity of global phytoplankton at equivalent timepoints would present a likely fractured viewpoint of the LDG. Our aim was to assess a general macroecological pattern with the assumption that repeated sampling would characterise a general level of diversity at that location.

The fourth and last notable methodological choice to remove species with incomplete taxonomic linages, affecting the overall diversity estimation. At the top of all our assessments of diversity, species are removed and this will change richness, the relative abundance and community similarity measured for locations which otherwise would have had these species counted. While this would influence the richness, the maximum number of species lost is minimal and as such does not present a substantial change in the richness pattern observed. The relative abundance and similarity measurements could be more affected for these species, but this is an issue of phytoplankton diversity assessments in general not one that could be rectified with the scope of our work. The broader issue is the lack of methodological bridge between genetic diversity and traditional taxonomic classification. Phytoplankton sit at the edge of

feasibility of taxonomic classification, while some species are easily observed under microscope others are sufficiently small to be indistinguishable without genetic assessment.

4 Species distribution modelling marine phytoplankton for Chlorophyll-*a* prediction

4.1 Introduction

Species distribution modeling (SDM) is a statistical approach used to predict the geographic distribution of species based on environmental covariates and species occurrence data (Guisan et al., 2017). It relies on correlations between known species locations and environmental predictors of species occurrence, such as temperature, salinity, or habitat type. SDMs range from simplistic linear regression models to complex non-linear machine-learning techniques, and they help identify patterns in species distributions across spatial and temporal scales (Thuiller et al., 2009). These models are useful for understanding the environmental preferences of species and forecasting their potential distributions under changing environmental conditions (Báez et al., 2020; Record et al., 2013).

Estimation of a species distribution using this method improves upon previous methods of species distribution using the raw observations alone. Only using observational data in species distribution estimation is limited due to ability of research programs to sample extensively to capture the full distribution of a single species typically, unless an extremely small distribution is apparent. Instead, SDMs use the environmental covariates of observations that are available to estimate the environmental niche of the species. This modelled environmental niche is then used to extrapolate in geographic space the probability of a species occurring in a location given the environmental conditions of that location, where no observation has been recorded. The product is a species distribution shaped by known observations and ecological niche theory. SDMs are commonly applied in terrestrial and marine ecosystems for various purposes, including conservation planning, fisheries management, and climate change impact assessments (Guisan et al., 2013; Johnson et al., 2023; Panzeri et al., 2024). In terrestrial systems, they are used to map the distribution of endangered species, guide reforestation efforts, and predict shifts in species ranges due to habitat loss (Eyre et al., 2022; Van Daele et al., 2021). In marine science, SDMs are widely used to model the distributions of fish, marine mammals, and benthic species (Tittensor et al., 2010b). However, their application to microscopic marine organisms such as phytoplankton and zooplankton has been relatively limited (Egorova et al., 2025; Righetti et al., 2019), despite the critical role these organisms play in ocean ecosystems, carbon cycling, and global climate regulation.

Notably, another methodological benefit of the SDM method is estimation of species distributions given limited observations. This was a key driver of the selection of this method to examine the marine phytoplankton LDG and potential relationships of the resulting data product. As examined in Chapters 2 and 3, the observational sampling bias present in the PhytoBase dataset raised questions of the realism of the patterns drawn. This research chapter

examines the marine phytoplankton LDG of the most well sampled species within the PhytoBase dataset to address these sampling biases.

Extending SDMs to marine phytoplankton is essential because these microscopic organisms form the base of the marine food web and are highly sensitive to environmental changes (Brotas et al., 2023; Reynolds et al., 2002). Unlike larger species, phytoplankton exist globally, yet their distributions remain understudied on a broad scale while large marine species benefit from extensive monitoring networks, such as birdwatching from vessels, marine mammal surveys, and aerial observations (Hammond et al., 2021; Sullivan et al., 2009). Phytoplankton species diversity cannot be estimated using remote sensing techniques, instead reflected wavelengths associated with phytoplankton pigments are remotely sensed allowing groups such as coccolithophores to be distinguished. Current satellite data provides insights into productivity but lack the taxonomic resolution needed to estimate species-level diversity accurately (El Hourany et al., 2024). Even within existing datasets, significant effort is still required to integrate observations across different methodologies and spatial scales to form a cohesive understanding of phytoplankton diversity.

By developing SDMs tailored to phytoplankton, I aim to improve our understanding of their spatial variability and enhance predictions of primary productivity and diversity patterns in the world's oceans. This work will contribute to a more comprehensive view of marine biodiversity and inform conservation and management strategies in a rapidly changing ocean. Specifically, by examining how well SDMs may meet this aim in the context of limited observational coverage using the PhytoBase dataset.

4.2 Methodology

4.2.1 Introduction

The taxonomically-joined AlgaeBase and PhytoBase dataset constructed previously (Section: 2.2.2.3) is used in the following diversity assessment. All statistical analysis, data processing and graphical presentations were performed using R statistical software (RStudio, 2020). The method was adapted for marine phytoplankton from methods outlined in (Egorova et al., 2025; Reygondeau et al., 2018). The SDM ensemble method and BIOMOD2 (Thuiller et al., 2009) parameterisation follows those applied in Egorova et al. (2025) differences are found in data depth filtering, environmental variables used and species range constraints due to different biota modelled. The method was originally designed for mesozooplankton and here it is applied to marine phytoplankton.

Here I pre-processed the taxonomic PhytoBase data using available sample depth data, keeping observations within the epipelagic zone. Relevant environmental variables were selected for marine phytoplankton occupying the epipelagic layer versus the mesopelagic zone modelled in the original application of the method. I did not elect to perform any expert species range constraints for marine phytoplankton due to lack of available knowledge on species specific distributions. Here I elected to extend the method to run on a high-performance computing cluster rather than a localised computing approach. I also took a conservative approach to the minimum number of species occurrences required for modelling (≥ 44 observations) leading

to a highly selective number of species modelled. This number was chosen as an arbitrary value to select enough observations to satisfy model fitting requirements. I will now detail my application of this method for the modelling of global marine phytoplankton diversity with respect to the work of (Egorova et al., 2025; Reygondeau et al., 2018).

4.2.2 Data

4.2.2.1 Environmental

Table 6 - Available databases of the environmental parameters used in this study.

Data source	Environmental variables	Layer type	Time period	Units	Spatial resolution
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=t	Temperature	DECAV	1955-2017	Celsius (°C)	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=s	Salinity	DECAV	1955-2017	Unitless	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=i	Silicate	ALL	1878-2017	µmol/kg	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=p	Phosphate	ALL	1878-2017	µmol/kg	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=n	Nitrate	ALL	1878-2017	µmol/kg	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=o	Dissolved Oxygen	ALL	1878-2017	µmol/kg	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=M	Mixed Layer Depth	DECAV81B0	1981-2010	Metres	1°
World Ocean Atlas 18 https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=A	Apparent Oxygen Utilisation	ALL	1878-2017	µmol/kg	1°
BioOracle v3.0 https://www.bio-oracle.org/downloads-to-email.php	Iron	Mean	2000 – 2018	µmol/m^3	0.05°
BioOracle v3.0 https://www.bio-oracle.org/downloads-to-email.php	Surface pH	Mean	2000 – 2018	pH	0.05°
BioOracle v3.0 https://www.bio-oracle.org/downloads-to-email.php	Surface velocity	Mean	2000 – 2018	ms^1	0.05°
AQUA MODIS Level-3 Mapped Photosynthetically Available Radiation, Version 2024, 2024	Surface photosynthetically active radiation	Mean	2002 - 2024	Einstein ms^{-2} day^{-1}	$4km^2$

ALL = An average of all available data, DECAV = Average of six decadal means (1955-1964, 1965-1974, 1975-1984, 1985-1994, 1995-2004, and 2005-2017) and DECAV81B0 = 1981-2010.

Global environmental layers relevant to the ocean euphotic zone were collected to characterise the environmental niche of marine phytoplankton (Supplementary material 7.4). These climatologies were selected based on known relationships between environmental factors and phytoplankton growth and productivity (Guisan et al., 2017). Temporal variability in environmental conditions was not explicitly included, as the aim of this study was to capture broad-scale spatial patterns rather than short-term fluctuations. This approach provides a generalisable framework for understanding species-environment relationships at large scales, while recognising that finer-scale temporal dynamics could refine niche characterisations in future studies.

Temperature strongly influences metabolic rates, constraining the thermal range a species can occupy (Brown et al., 2004). If temperatures exceed optimal limits, metabolic function declines, leading to competitive exclusion or physiological constraints, for example denaturing of cells via metabolic breakdown ad higher temperatures which leads to competitive exclusion via other species more thermally capable at higher temperatures. While marine phytoplankton are not typically exposed to lethal thermal limits in natural conditions, temperature still affects their global distribution when species are not within their natural optima (Chen, 2015). Salinity, in contrast, has a smaller effect on global phytoplankton distributions due to its relatively narrow range of variation. However, in regions with high freshwater input, such as the Amazon River, reduced salinity can favor freshwater phytoplankton over marine species (Goes et al., 2014).

Macronutrients such as silicate, nitrate, and phosphate are essential for phytoplankton growth and reproduction. Diatoms require silicate for frustule formation, nitrate supports amino acid and nucleic acid synthesis, and phosphate is critical for ATP production, which drives metabolic processes. The availability of these nutrients directly influences phytoplankton abundance and community composition and vice versa, phytoplankton growth reduces nutrient availability. Oceanographic mixing and the role of stratification affects the supply rate of nutrients in the euphotic zone determining nutrient selectivity of different groups and thus species of marine phytoplankton (Villamaña et al., 2019). Additionally, micronutrients such as Iron is known to be a limiting factor in the growth of marine phytoplankton.

Light availability for photosynthesis is paramount to growth of marine phytoplankton. The availability of light is variable globally due to differences in cloud cover, turbidity of the water and depth in the water column. The variance of light availability results in adaptive strategies of marine phytoplankton to occupy different niche spaces within the range of light availability (Litchman, 1998).

The next parameter, pH, is another important environmental variable for consideration when modelling marine phytoplankton distributions. Changes to the biochemistry of the ocean and the effect on sea surface pH is leading to increasing acidification of the upper ocean. As climate change progresses, the concentration of carbon dioxide (CO_2) is increasing in the atmosphere and due to the buffering process of the air-sea flux, the ocean is uptaking atmospheric CO_2. This increase in dissolved CO_2 results in increasing levels of carbonic acid with breaks down

to free hydrogen ions and bicarbonate (Doney et al., 2009). The increased formation of bicarbonate due to this process reduces the bioavailability of calcium carbonate. This reduction in availability may affect the distribution of coccolithophores which use calcite to form calcite disks on their outer cellular surface (Kaiser & Williams, 2011).

All of the above direct abiotic factors are directly affected by the mixing and turbulence of the water column (Villamaña et al., 2019). To serve as a proxy for this, the addition of sea surface velocity was included in the suite of environmental layers for characterisation of marine phytoplankton diversity.

The World Ocean Atlas 18 (WOA18) dataset is a comprehensive oceanographic dataset. The world ocean atlas as a product has been publicly available from the National Centers for Environmental Information (NCEI), a subdivision of the National Oceanic and Atmospheric Administration (NOAA) since 1994. There have been eight iterations of the World Ocean Atlas as new parameters and instrumentation have come online. WOA18 is the seventh iteration, with a more recent update to the dataset as of 2023, WOA23, but at the time of research for this work this version was not available so I have not used it.

The WOA18 is comprehensive with key oceanographic parameters which we have used in the subsequent analyses. For this analysis we have used, Temperature, Salinity, Silicate, Phosphate, Oxygen, Nitrate and Mixed Layer Depth (MLD) with a one-degree grid resolution. Each dataset contains multiple temporal ranges and averages; annual, seasonal and monthly climatologies. The temporal range we have used for our analyses is an extension of the 30-year period (1981-2010) 'climate normal' reference period defined by the World Meteorological Organization (WMO). The 'climate-normal' period is defined by the usage of remote sensing of oceanographic parameters so is used to ensure comparison among different data types for consistency. Climatological normal datasets are used widely to provide a prediction of conditions most likely to occur in a given location (World Meteorological Organization, 2017). In addition, they provide a comparison case for climate anomalies in years post the normal. For the purpose of this research, I used the greatest temporal coverage available for each climatology (Table 1). The exception being Mixed Layer Depth which is the climate normal period.

A key environmental parameter for phytoplankton is sunlight availability and intensity. To incorporate this into our models, I used a full mission composite of photosynthetically available radiation (sPAR) obtained from the AQUA-MODIS (Moderate Resolution Imaging Spectroradiometer) managed by NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group (*AQUA MODIS Level-3 Mapped Photosynthetically Available Radiation, Version 2024*, 2024). The composite is of the duration of the mission from 04/07/2002 to 04/07/2024 with a resolution of $4km^2$.

Remote sensing chlorophyll-a is facilitated by detecting reflected wavelengths of light from the ocean surface. A NASA operated satellite, Aqua-MODIS (Moderate Resolution Imaging Spectroradiometer), is used to remotely sense ocean colour (Eight spectral bands ranging from 405 to 887*nm*)(Savtchenko et al., 2004). Chlorophyll-a is one of a few pigments used in photosynthesis by marine phytoplankton, reflecting the colour green. Measuring the ratio of

blue-green reflectance spectra in the upper ocean details the relative concentration of marine phytoplankton chlorophyll-a. The key benefit of remote sensing is the global data output compared to limited in-situ fluorescence measurements.

Data from the AQUA-MODIS satellite was obtained for the full mission to date (Start: 2002/07/04 to Current: 2024/12/31). The data product is an average chlorophyll-a measurement per 4km^2 globally for the full mission. The data was then gridded to the equal area grid to ensure 1:1 value matching with the modelled marine phytoplankton data. Any MODIS data with a proportional of land in the grid cell was removed prior to correlative analysis. In addition, a global shapefile with the economic exclusion zones (marineregions.org) was acquired to filter and remove MODIS and modelled marine phytoplankton data. This was motivated by issues in chlorophyll-a estimation in coastal regions due to interactions with bottom reflectance and detritus in the water column.

The BioOracle platform was built with the motivation to provide high-resolution global datasets for ecological modelling practices such as species distribution models. The key benefit of the BioOracle is the synthesis of 24 data layers to the same high resolution compared to obtaining this data from different providers which use different resolutions (Tyberghein et al., 2012). This reduces the strain of mapping all required layers for analysis to the same grid for analysis. Recently, BioOracle was updated to include outputs from the Sixth Phase of the Coupled Model Intercomparison Project (CMIP6) to provide future scenarios for all layers (Assis et al., 2024). As our work is currently focused on the distribution of marine phytoplankton in the present day, we have used present day layers for dissolved iron, surface pH and sea surface velocity (https://www.biooracle.org). The data for the present-day layers we have selected we obtained by the BioOracle group from the Global Ocean Physics Reanalysis and Forecast and the Global Ocean Biogeochemistry Analysis and Forecast, two pre-processed re-analyses of Copernicus Marine Environment Monitoring Service (https://data.marine.copernicus.eu/products). Where they integrate both satellite and in-situ data into each data layer (Jean-Michel et al., 2021). The data from the Copernicus group is available in 0.08^0 and 0.25^0 degree resolution whereas the BioOracle group chose to provide 0.05^0 resolution. The BioOracle used an inverse distance weighting algorithm which interpolates missing information on the new grid resolution while weighting values (power of 2) from the Copernicus grid (8) closest to the new desired location higher than those further away (Assis et al., 2024).

To match environmental parameters to the species occurrences for niche modelling, the discrepancies in resolution across these data layers required standardisation for comparison. All environmental layers and the species occurrences were rasterised and interpolated to a standardised equal area grid. This was performed using the nearest neighbour method using the RANN package in the R statistical software. The equal area grid (55km^2) created using ArcGIS software was obtained from Gabriel Reygondeau in collaboration of this methodological development.

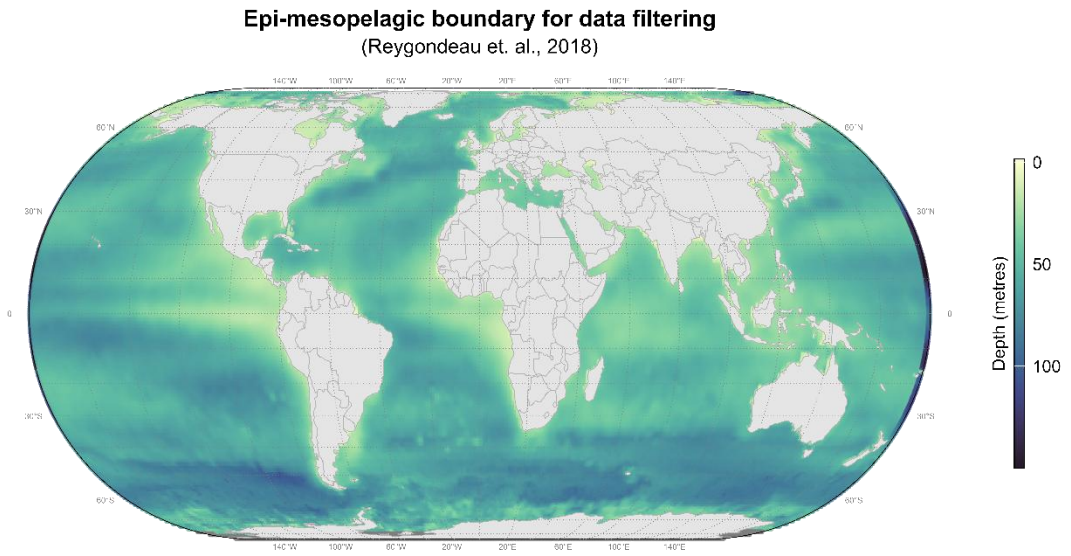


Figure 4.1 - Epi-mesopelagic boundary from Regondeau et al., 2018, depth value indicates boundary of mesopelagic depth.

Phytoplankton are found in the euphotic zone of the upper ocean water column. The environmental layers obtained for this modelling procedure contains depths across the water column from surface to benthos. To ensure the spatial environmental layers are relevant to the ecological habitat of marine phytoplankton, the layers were depth filtered to the epipelagic zone. The depth cut off for the epipelagic zone was sourced from the 3D biogeochemical approach and associated produced epi-mesopelagic boundary data product (*Figure 4.1*) (Reygondeau et al., 2018). All environmental data shallower than this depth boundary were averaged. The mean values were then used as the environmental variables to characterise the species niches used in this analysis.

4.2.2.2 Biotic

The marine phytoplankton dataset created in previous chapters was used in the following analysis. The dataset created previously was adopted from the compiled PhytoBase dataset (Righetti et al., 2020). I sourced the full taxonomic linages of the species therein and constructed an enhanced version of the PhytoBase dataset for the previous diversity assessments and this subsequent analysis using AlgaeBase (Guiry, M.D. & Guiry, G.M., 2023). The specific methodology for doing so is outlined in Chapter 2 section 2.2.2.3.

For the SDM analysis, the data was filtered to include only species with recorded depths of less than 250 m and at least 50 observations. The dataset from previous chapters was reduced to include only the relevant columns: scientific name, species, genus, family, order, class, phylum, kingdom, domain, latitude, observation depth, year, and month. This filtering ensured that only observations within the euphotic zone were used for modeling. Each species was cross-validated with the World Register of Marine Species (WoRMS) to confirm valid species names and obtain an Aphia ID. Species not found in this procedure were excluded from further analysis see section 7.3.2 for a list of modelled species.

All species data, excluding occurrence information, along with the Aphia ID and taxonomic details, was stored in a META file. Occurrences were then binned into the equal-area grid. To reduce potential bias from species observed multiple times in a single grid cell, occurrence data

was filtered to include species occupying 44 or more grid cells. Species with fewer occurrences were excluded from the modelling process.

4.2.2.3 Species Distribution Modelling using BIOMOD2 and HPC

Parallel processing is implemented to efficiently run SDMs across multiple species. The number of processing cores is determined based on the computing environment, whether running locally or on a high-performance computing (HPC) cluster. Batch processing is employed to handle species in groups of 50 at a time, ensuring computational efficiency. Each parallel process reloads the necessary libraries to maintain functionality, and logging functions are set up to track errors or issues during execution. The next step involves defining the data paths for species occurrences, model evaluation results, probability distributions, and variable importance measures. The modeling process begins by selecting relevant models within BIOMOD2, including Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), Maximum Entropy (MaxEnt), Random Forest (RF), Classification Tree Analysis (CTA), Surface Range Envelope (SRE), and Artificial Neural Networks (ANNs).

To ensure robust model calibration, pseudo-absence selection is conducted, where absence points are generated based on species occurrence patterns. The dataset is split into training and testing subsets according to predefined ratios, which vary based on the number of occurrences (e.g., a 70/30 split for abundant species). The modeling process (Figure 4.2) is run in duplicate using two pseudo-absence replicates to enhance predictive reliability. BIOMOD2 is then used to execute the SDMs, specifying cross-validation procedures and evaluation metrics, including Receiver Operating Characteristic (ROC) curves and True Skill Statistic (TSS). The modeling output includes individual projections for each species, which are subsequently combined using ensemble modeling techniques.

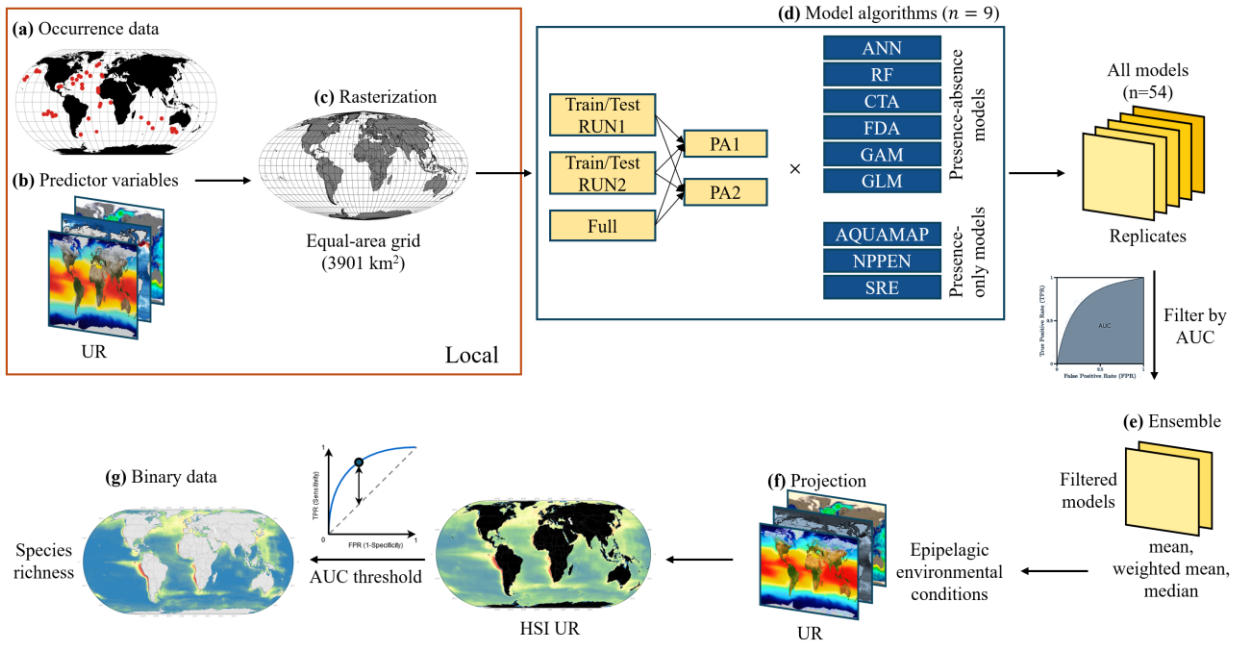


Figure 4.2 – The species distribution model workflow depicts the different analytical steps performed for each species. Depending on the sampling size the train/test split is different. AUC - area under the receiver operating curve; HSI – habitat suitability index; UR – Unrestricted range; PA – sets of background points; TSS – true skill statistic; Local – Steps performed on local machine. Figure adapted from (Egorova et al., 2025) with permission.

The ensemble modeling step aggregates predictions across different algorithms to improve accuracy and stability. Several ensemble metrics are calculated, including the ensemble mean, median, and weighted mean, with a significance threshold of 0.6 applied for model inclusion. Additionally, variable importance is assessed to determine the relative contribution of environmental predictors to species distributions. The final step involves generating global distribution projections, ensuring that models are correctly applied to the environmental space. These projections exclude clamping masks to avoid extrapolation biases. Model outputs are stored systematically, including probability distributions, evaluation metrics, and comparative results from Aquamaps and BIOMOD2. A final evaluation file compiles all modeling results, parameter settings, and variable importance scores for further analysis.

Habitat suitability index (HSI) was calculated as a sum of individual HSI for each species in each grid cell. The total mean HSI is all species HSI summed per grid cell which is then normalised to a value between 0-1. Modelled species richness is a binary conversion of the probability of occurrence outputted by BIOMOD2. The conversion is done by using a cutoff of model performance (Area under curve (AUC) ≥ 0.6)(Araújo et al., 2005; Elith et al., 2006), if a species is predicted to occur above this cutoff it is given a binary presence of 1 and if below, 0 occurrence. The value of AUC 0.6 was chosen to be above the 0.5 value which indicates the model has no discrimination ability above random prediction.

4.2.2.4 Atlantic Meridional Transect data filtering from PhytoBase

The PhytoBase dataset compiled data sources of marine phytoplankton diversity data. Notably included in the PhytoBase dataset, the Atlantic Meridional Transect is a long running in-situ oceanographic research cruise. It spans the northern and southern Atlantic Ocean, operating since 1995 from the National Oceanography Center, UK. Within the PhytoBase dataset there are 4748 observations tagged with AMT identifiers. To perform the subsequent correlative analysis, the modelled marine phytoplankton data and MODIS data was filtered by the grid cells associated with the AMT stations. This dataset was used to cross-validate the modelled distributions with observational data.

4.2.2.5 Standard error calculation

To assess the spatial correlation of the AQUA-MODIS chlorophyll-a data and the modelled habitat suitability, a linear model was constructed and the standard error of the mean performed for each grid cell. To do this, linear model was constructed using the lm() function in R between the log chlorophyll-a and log HSI data. The predict() function was then applied to the constructed linear model with standard error calculation. The standard error statistics is calculated on logarithmic values therefore the error estimate is also logarithmic in scale.

4.3 Results

4.3.1 Global marine phytoplankton modelled richness and habitat suitability index

The modelled species richness of marine phytoplankton (Figure 4.3A) shows high richness in upwelling regions with boundary currents. Peru, Benguela and Canary current systems show high projected richness for the 283 species modelled. The modelled LDG (Figure 4.3B) details a bimodal LDG with a decline in modelled richness at the equator. Additionally, the mean richness shows a peak in the northern hemisphere. This arises due to moderate richness modelled in the region and reduced pelagic system observations present compared to lower latitudes and the southern hemisphere. This details an area effect of the modelled results. The HSI results (Figure 4.3C) generally follow the modelled species richness in locations of suitability hotspots. The HSI result models the suitable habitat of the 283 modelled marine phytoplankton but differs from the species richness map (Figure 4.3A) as the latter performs a cutoff to convert the probability of occurrence into a binary presence/absence value.

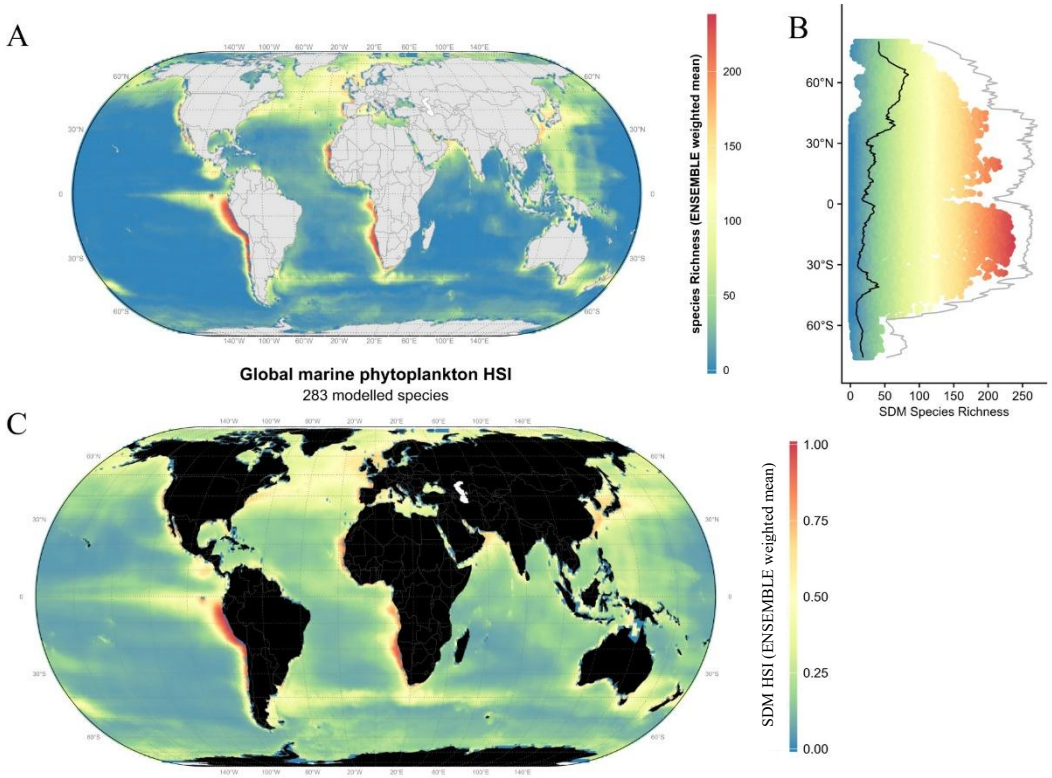


Figure 4.3 - Marine phytoplankton species distribution model weighted mean ensembles. (A) Modelled species richness for 283 species. (B) Total species richness per latitude (55km^2), grey line and Mean species richness per latitude (55km^2), black line and (C) Modelled mean habitat suitability index for marine phytoplankton (283 species).

The environmental predictors for the species distributions and their relative importance per phyla is shown in Figure 4.4. The diffuse attenuation coefficient (Zeu) has strong importance for the majority of marine phytoplankton phyla with the exception of Euglenozoa. Euglenozoa shows high importance for locations of oxygen utilization. Chlorophyta and Ochrophyta have the greatest relative importance of Zeu as a predictor for their distributions but also the greatest variability across species within the phyla. Dissolved silica is a strong environmental predictor

of importance for Ochrophyta. Sea surface velocity has the lowest relative importance across all phyla. Temperature shows minimal relative importance across all phyla.

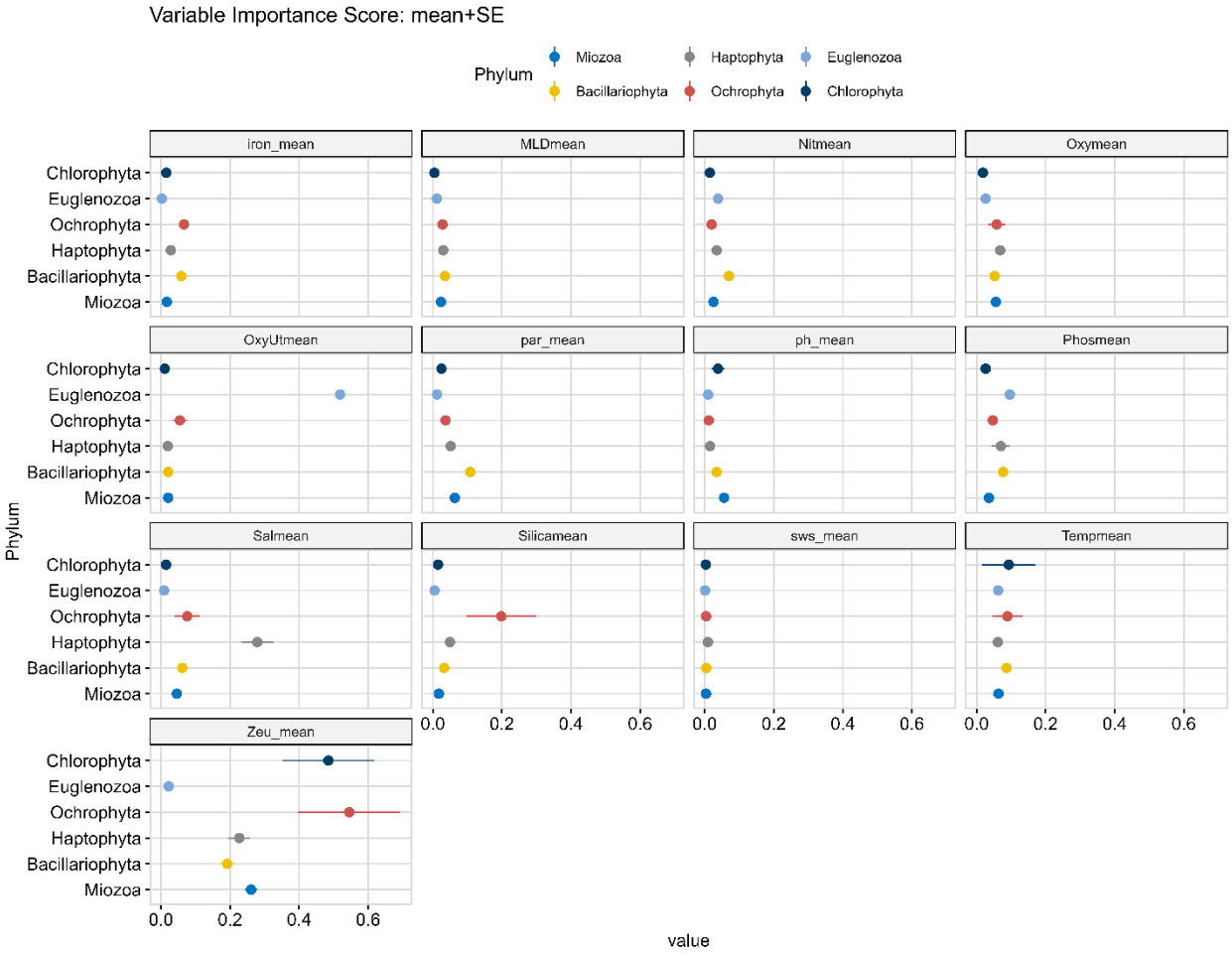


Figure 4.4 – Modelled environmental variables importance for each marine phytoplankton class (283 species). Mean value shown by coloured circles and line represents variable importance standard error across all species within class. MLD – Mixed layer depth, OxyUT – Oxygen Utilisation, PAR – Photosynthetically active radiation, Phos – Phosphate, Nit – Nitrate, Temp – sea surface temperature, Sal – Salinity, sws_mean – mean surface velocity and Zeu – Diffuse attenuation coefficient.

4.3.2 Global Chlorophyll-a relationship with modelled marine phytoplankton habitat suitability

Below, the results of the ensemble modelling of marine phytoplankton are shown (*figure 4.5B*). The log HSI details globally that coastal regions are of high suitability to marine phytoplankton. Notably, the northern hemisphere and the North Atlantic shows high suitability. The pelagic Pacific Ocean shows generally lower suitability with the exception of the equatorial current system off the coast of Peru. The global chlorophyll-a (log transformed) data shows similar spatial distribution to that of the modelled habitat suitability. The correlation plots assess the relationship between global chlorophyll-a and the modelled HSI prior to log transformation. In *figure 4.5C* the correlation is drawn between log chlorophyll-a and modelled HSI. The relationship is weak (R^2 = 0.17) due to locations of high HSI at regions of intermediate log chlorophyll-a. To rescale the relationship, the HSI data is log transformed. Figure 4.5D shows the log transformed HSI and the log chlorophyll-a shows a stronger relationship (R^2 = 0.49) after rescaling.

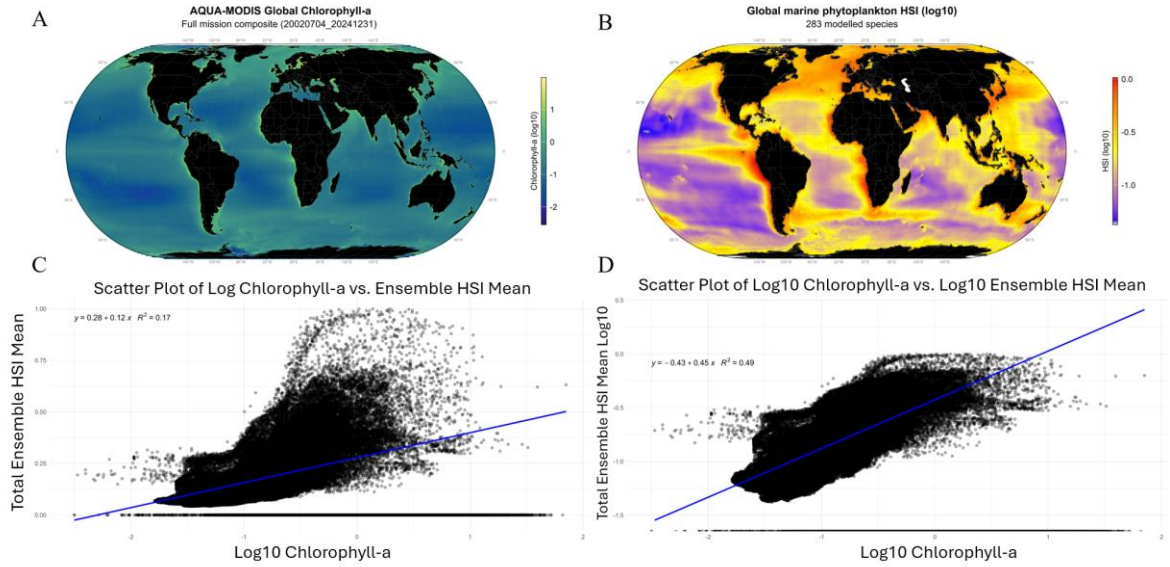


Figure 4.5 – Global chlorophyll-a and modelled HSI maps and correlation plots. (A) Full AQUA-MODIS mission composite of satellite chlorophyll-a (log), (B) Global HSI (log) for 283 modelled species, (C) Scatterplot of Log chlorophyll-a versus mean ensemble HSI, (D)) Scatterplot of Log chlorophyll-a versus Log mean ensemble HSI.

4.3.3 Truncated Chlorophyll-a and modelled marine phytoplankton habitat suitability

Removal of the coastal regions using EEZ filtering and truncating the data above and below 50° leads to stronger relationship initially (Figure 4.6, $R^2 = 0.41$). The removal of coastal data was driven by coastal regions typically having over-estimated chlorophyll-a measurements due to high turbidity (Moradi & Zoljoodi, 2023). The constrained relationship holds when the log transformed truncated HSI and chlorophyll-a are assessed. The truncated data details a global relationship between modelled marine phytoplankton HSI and chlorophyll-a holds irrespective of data removal of regions with higher uncertainty in estimation. The uncertainty arises in coastal systems and high polar latitudes due to satellite coverage limitations and optical restraints.

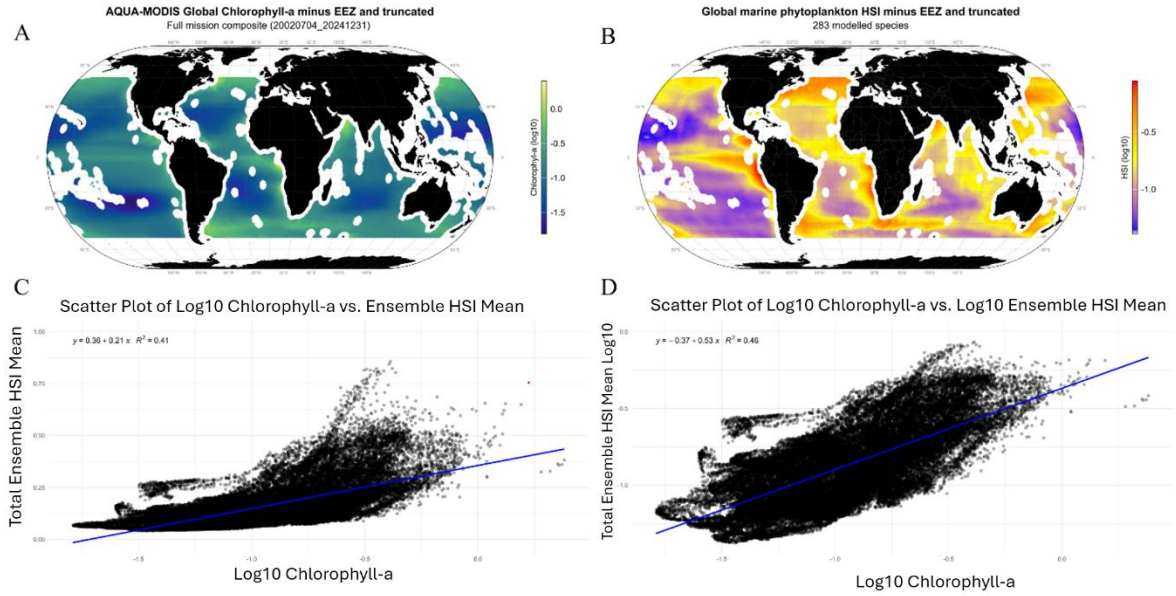


Figure 4.6 – Truncated global chlorophyll-a and modelled HSI maps and correlation plots. (A) Truncated AQUA-MODIS mission composite of satellite chlorophyll-a (log), (B) Truncated HSI (log) for 283 modelled species, (C) Scatterplot of truncated Log chlorophyll-a versus mean ensemble HSI, (D) Scatterplot of truncated Log chlorophyll-a versus Log mean ensemble HSI.

The full standard error map (Figure 4.7A) highlights locations where the predicted linear relationship has greater standard error estimated from the linear model. Locations where that error is reduced indicates where that estimated error is small for the relationship of chlorophyll-a and HSI values. In Figure 4.7A the standard error is concentrated in coastal regions, particularly the upwelling regions with boundary currents; Peru, Benguela and Canary current systems. Notably, pelagic systems of the northern and southern Pacific and Atlantic subtropical gyre systems show higher predicted error from the standard mean of the linear model. In Figure 4.7B the distribution of the standard error is altered by the truncation of the northern and southern temperate and polar regions and removal of the EEZ globally. This map underlines the noted locations noted for the full spatial distribution of the standard error (Figure 4.7B).

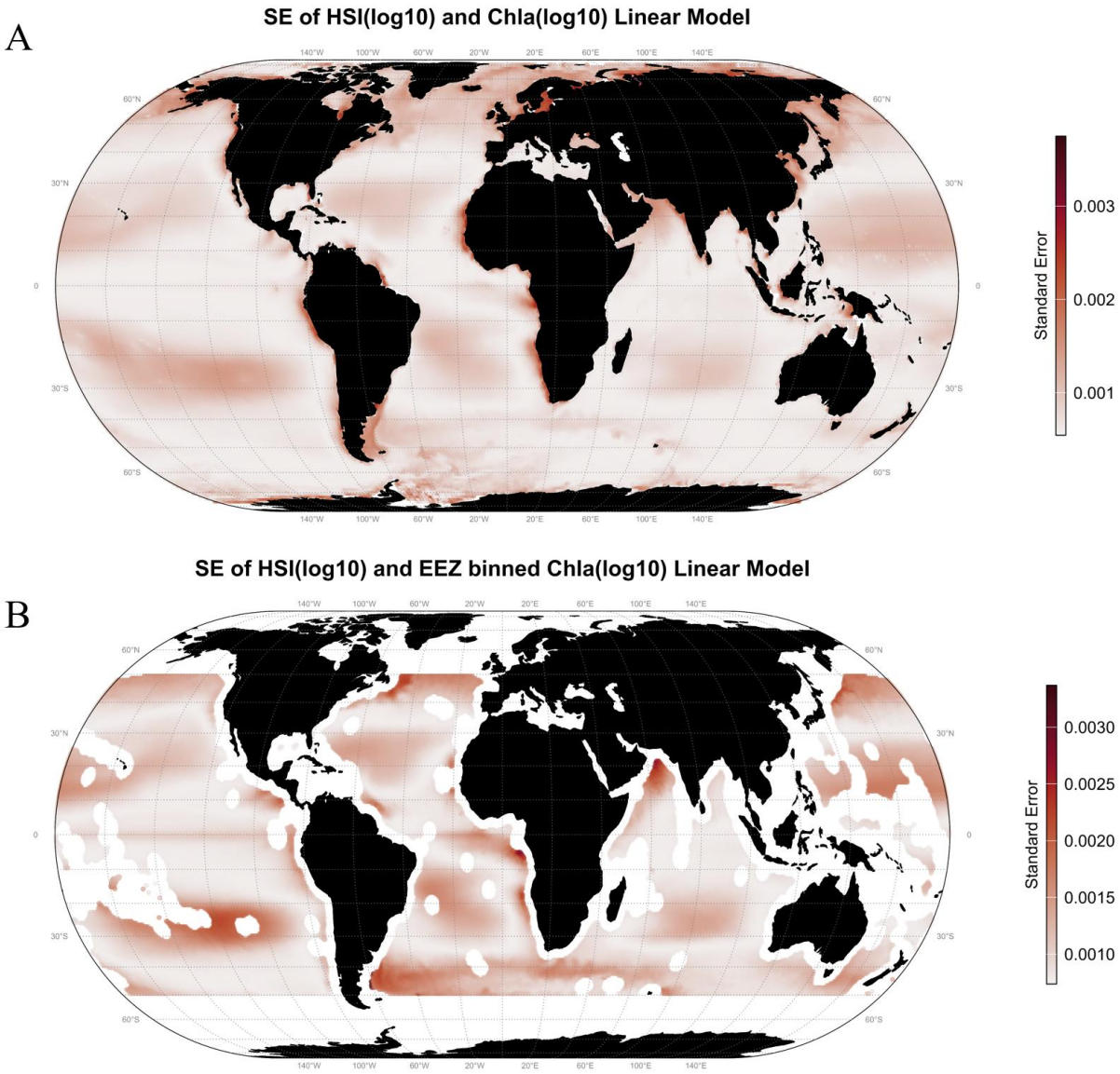


Figure 4.7 – Spatial standard error of log HIS and log chlorophyll-a. Standard error for full modelled HIS(log) and AQUA-MODIS chlorophyll-a(log) correlation, (B) Truncated and EEZ removed standard error for modelled HSI(log) and AQUA-MODIS chlorophyll-a(log) correlation.

4.3.4 Atlantic meridional transect analysis

To cross-validate the global and truncated analyses, the above figures detail the observational data from the AMT. The AMT spans a substantial portion of the Atlantic Ocean from Southampton, UK to the south American and African continents (Figure 4.8). The relationship between the mixed variable collection (log) suggests modelled richness and observed richness show similar relationships with latitude. They both have peaks in the northern coastal sub-tropics and the southern coastal Argentina. The AMT satellite chlorophyll-a (log) shows similar peaks with a key relationship with the log transformed modelled HSI. This relationship is extremely strong and detailed further in Figure 4.8C, with an of $R^2 = 0.94$.

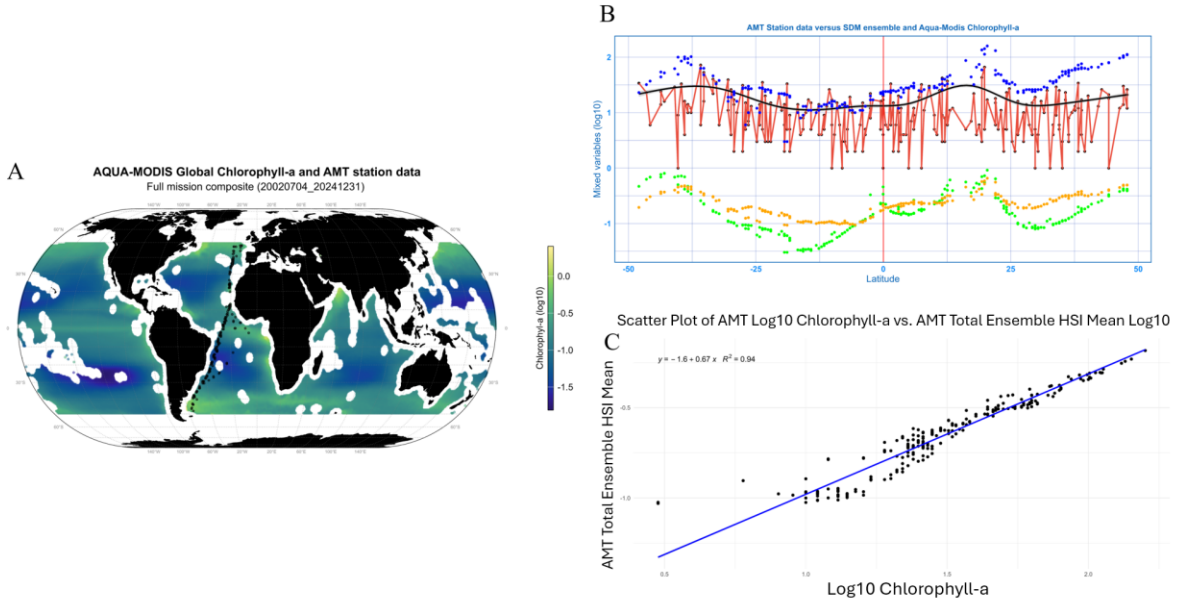


Figure 4.8 - Log observations of species richness (red line), log modelled species richness (blue circles), log Chlorophyll-a (green circles) and log modelled habitat suitability index at AMT stations (orange circles). Figure 31B, Full mission composite Aqua-MODIS chlorophyll-a truncated and EEZ removed map with AMT stations (black circles) shown. Figure 31C shows the correlation of log chlorophyll-a and log HSI.

In Figure 4.9A the non-logarithmic values for modelled species richness and observed AMT species richness show generally high correlation in lower latitudes. There are two notable peaks in AMT species richness at -37.5° and 18.5° latitudes. The modelled species richness mirrors these peaks in latitudinal diversity but the number of species with unrestricted ranges predicted to occur is greater. The notable deviation in this relationship occurs in the northern sub-tropical region (37.5° - 50°), where the modelled species richness increases and the observed AMT richness plateaus.

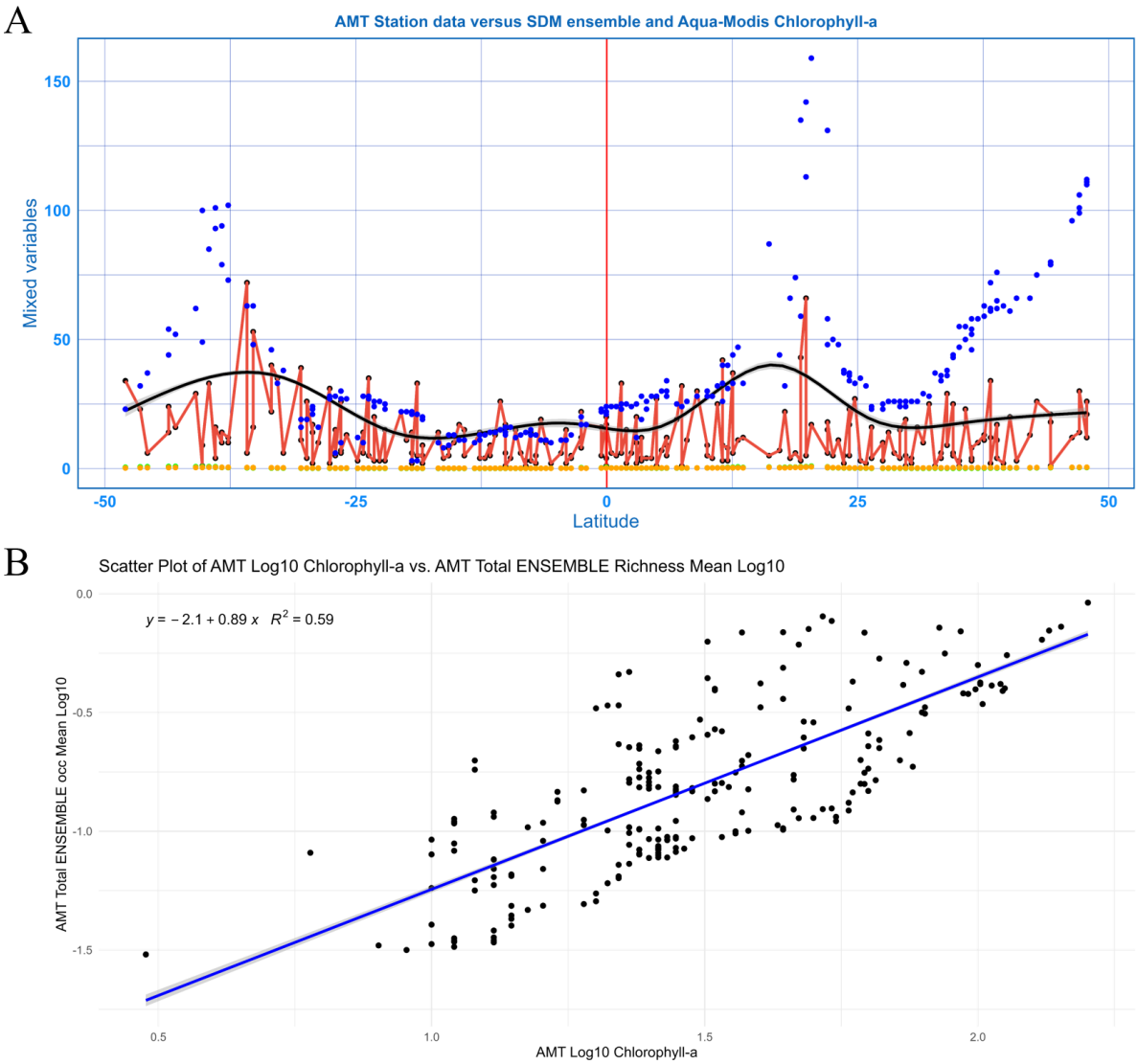


Figure 4.9 – AMT station data vesus SDM ensemble and AQUA-MODIS chlorophyll-a. (A) Non-logarithmic values of modelled species richness (blue circles), AMT species richness from PhytoBase (red line and black points), HSI (orange circles), Chlorophyll-a (green circles), (B) Linear regreassion of Log Chlorophyll-a and Log modelled species richness. GAM fitted to AMT species richness.

The AUC values (Figure 4.10) detail the variance in relative predictive performance of each of the individual SDMs and the resulting ensembles. Overall, the surface range envelope (SRE) model had low predictive capacity across all phyla. Figure 4.10 also highlights the differing number of species per phyla modelled. Bacillariophyta and Miozoa were most numerous in this modelling selection. The ensemble model performed consistently better with greater AUC values than the individual models.

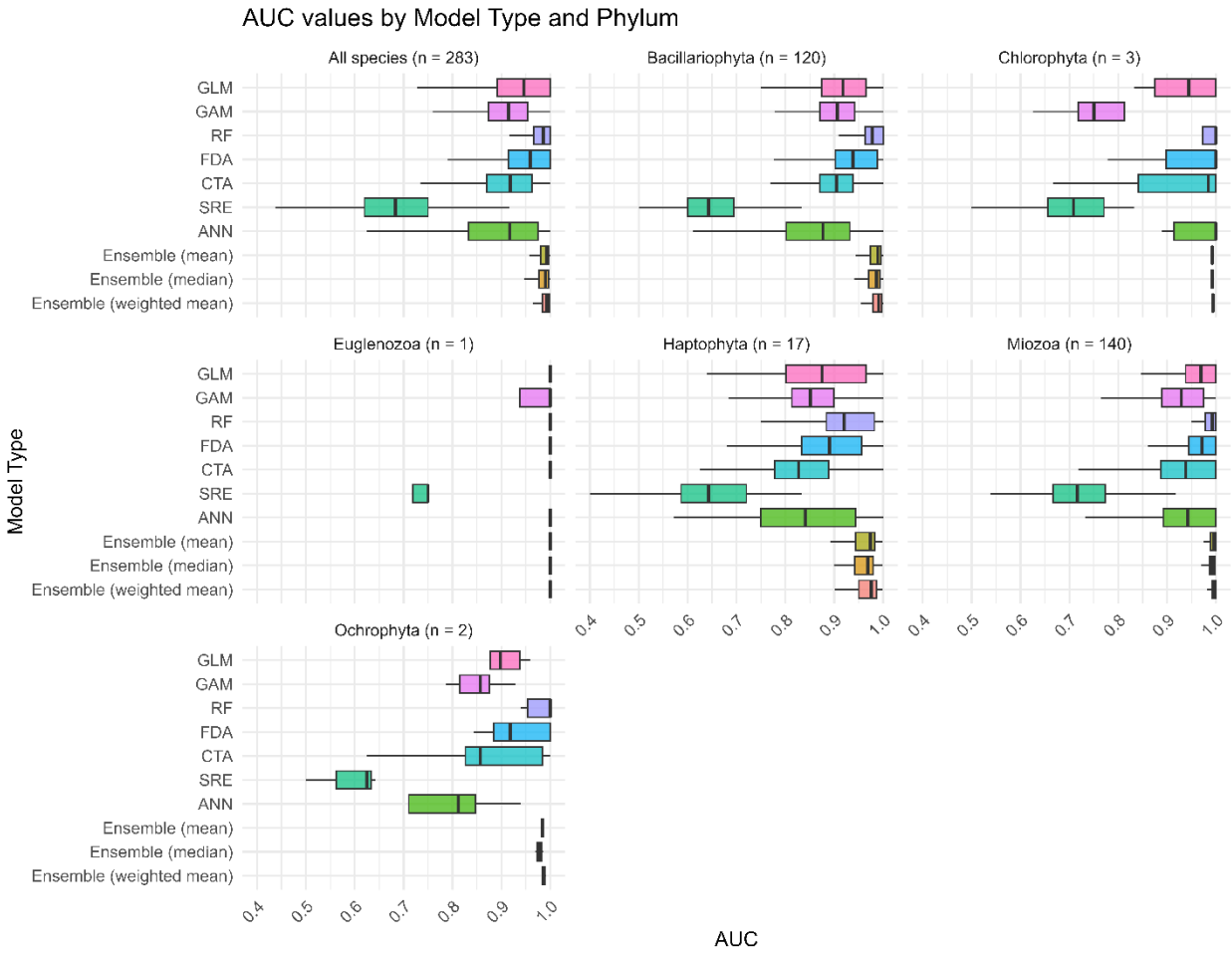


Figure 4.10 – Distributions of area under curve (AUC) scores of projections from individual species distribution algorithms and their ensembles. Boxplots presented for all species and disaggregated by model performance across phyla. The boxplots detail the distribution of data based on the five-number summary: minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The box spans the interquartile range (IQR), with a line indicating the median. Inspiration for data presentation see Egorova et al., 2025.

4.4 Discussion

The rationale of this chapter was to apply the SDM method to estimate the LDG of marine phytoplankton and assess potential uses of the resulting distributions. The driver to use this method was to tackle the sampling biases found in the PhytoBase dataset. As noted, the SDM can estimate distributions given sufficient observations to characterise the niche of a species. Due to the limited number of observations for many species in the PhytoBase dataset, not all species were modelled in this approach. This is a core issue of this approach, in this ensemble species modelling approach, I modelled the distributions of 283 marine phytoplankton species with high spatial coverage but severely reducing the total diversity contained in the PhytoBase dataset. Importantly, the species modelled here represent only the most spatially widespread taxa in the PhytoBase dataset, which contains 1,591 species in total. This subset limitation constrains the ability to infer global phytoplankton diversity purely from species distribution models.

The resulting latitudinal diversity pattern of species richness from the SDM approach (Figure 4.3B) deviates from those observed in Chapter 2 and 3. The SDM approach generates a distribution with broad latitudinal regions of high richness from sub-tropical regions, dropping sharply in the polar regions. There is a slight reduction in diversity at the equator, similar to the Hill number species richness assessment of Chapter 2. This suggests that the latitudinal diversity gradient for marine phytoplankton may be generally high for all regions except polar regions, where sunlight is seasonally limited. The SDM approach does not model interactive effects of competition for resources such as nutrients or light, therefore this also may drive the broader distributions of species and thus a generally higher richness across a wider range of latitudes. It may also be a factor of the generalist nature of the species modelled, which are highly abundant, which when modelled with no interactive terms for distribution result in large spatial distributions.

The resulting habitat suitability patterns show a global distribution but biased to coastal regions, with peak suitability in upwelling zones and coastal shelves. High suitability is also observed at the southern tips of South America and South Africa, whereas open ocean pelagic systems, such as the Pacific subtropical gyres, exhibit low suitability. However, given their high spatial coverage, these selected species are assumed to represent a significant portion of the productive fraction of marine phytoplankton biomass in line with the larger size spectra modelled by Dutkiewicz et al in 2020. It is also assumed that the these species modelled using this SDM approach are representative of the larger size class of marine phytoplankton which show a stronger correlation with biomass (Li, 2002). This assumption underpinned the subsequent correlative analysis with remotely sensed chlorophyll-a, a widely used proxy for phytoplankton productivity. The analysis revealed an intermediate global linear relationship between modelled habitat suitability (HSI) and chlorophyll-a (Figure 4.9B). This suggests that the ensemble modelling approach has predictive capacity for estimating chlorophyll-a community composition from phytoplankton diversity data, a major step toward using species distribution models to estimate global phytoplankton biomass.

To test the robustness of this relationship, I removed Exclusive Economic Zone (EEZ) coastal data from both the chlorophyll-a and HSI datasets and re-ran the correlation analysis. The intermediate relationship persisted, further supporting the model's predictive capability beyond coastal influences. To further validate the model output, I compared it against observational data from the Atlantic Meridional Transect (AMT) cruise, a substantial dataset comprising 4,748 observations. Here, the relationship between HSI and remotely sensed chlorophyll-a at AMT stations was strongly linear ($R^2 = 0.94$). This reinforces the hypothesis that modelled phytoplankton diversity can reliably predict chlorophyll-a satellite measurement. The integration of species diversity information into biomass estimation models could enable future work incorporating relative abundance into SDMs, thereby improving biomass predictions. If successful, this would significantly enhance estimates of marine primary producer biomass. While further research is needed to confirm whether this method can provide direct biomass estimations, these early results suggest strong potential to do so.

The SDM results overall suggest high diversity relationship to coastal systems but as observed in the fine-grain scale analysis of Chapter 2 - Figure 2.5, the sampling intensity is also concentrated in coastal systems. The saturation curve and rarefaction that upwards of 20,000 observations are required to saturate richness. In general, this is not the case and as such this approach is underestimating species richness of pelagic systems and biasing distributions to coastal systems. While a key advantage of species distribution models is their efficiency compared to mechanistic approaches but the high number of observations required and the observational data available reduces this benefit. This current database of marine phytoplankton and being unable to model rare species reduces confidence in this approach. While SDMs offer shorter modelling times and can be used to hindcast past phytoplankton diversity states, as well as predict future shifts in distributions, the do so for a limited subset of species. Future work should explore the capacity of SDMs to hindcast past chlorophyll-a concentrations by ground-truthing with in situ measurements. This is particularly relevant given the relatively short operational times of Earth observation satellites, which limit long-term global chlorophyll-a monitoring. In contrast, mechanistic models require extensive spin-up times, often extending beyond satellite observational records. If SDMs can reliably estimate historical chlorophyll-a levels, they may provide a valuable ground-truthing tool for mechanistic climate model projections.

5 Thesis discussion and conclusions

5.1 Key findings

The broad aim of this thesis was to assess the alignment of marine phytoplankton diversity to the LDG macroecological pattern observed broadly across marine biological taxa (Tittensor et al., 2010b). I found that the marine phytoplankton LDG pattern deviates from the expected unimodal relationship (peak of diversity at the equator) and that choice of diversity index influences the location of the peak of marine phytoplankton diversity. Marine phytoplankton species richness peaks in the sub-tropical regions latitudinally and in equal area assessments in line with recent reanalysis of the marine LDG (Chaudhary et al., 2021). In tandem, this occurs concurrently with a North Atlantic sampling bias of high sampling compared to that of low northern latitudes and southern hemisphere latitudes (Righetti et al., 2020).

The Hill Number LDG assessment showed that this bimodal signature in the distribution of marine phytoplankton occurs at all levels of the *q* sensitivity parameter. These findings detail that the bimodal pattern occurs across the spectrum of rarity. To confirm this finding, I standardised the sampling effort across four broad latitudinal regions and applied the rarefaction method. The rarefaction results confirmed that the north and south tropical regions, which include the subtropical regions, have the highest diversity of marine phytoplankton compared to that of the temperate and polar regions.

Comparatively, the re-analysis of the LDG of marine phytoplankton with species richness, evenness and taxonomic similarity gives a different result. The LCI LDG assessment details a unimodal peak located in the southern hemisphere when taxonomic similarity is added, also observed in terrestrial grasses (Visser et al., 2014). At the level of index quantification, application of the decomposition of the LCI (Chen & Grinfeld, 2024) details how each component of diversity shapes the LDG of marine phytoplankton using the LCI. Sampling effort likely strongly influences the bimodal diversity pattern, but not entirely as high rates of sampling in the northern temperate region do not produce high species richness values. Evenness is fairly consistent except for the Northern temperate region, where evenness declines. Similarity is greater in the northern hemisphere and decreases in the southern hemisphere.

Following the axioms of diversity (Daly et al., 2018), diversity measurement should be maximised by greater richness in tandem with greater evenness and lower similarity (Leinster & Cobbold, 2012b). These axioms state; Symmetry: that the order of relative abundances in a vector does not alter the diversity value, Continuity: Relative abundances are a continuous function, Evenness: diversity is maximal for fixed number of species when all relative abundances are equal, Principle of transfers: Where the transference of abundance from one species of high abundance to another of lower abundance, the resulting diversity value must increase in line with the evenness axiom, Monotonicity in number of species: Addition of a new species must increase diversity and the Replication principle which states the pooled sample of maximally diverse and even communities is *n* times the number of subsetted samples of maximally diverse and even communities.

As observed in the analysis performed in Chapter 3, this explains our findings when the LCI is calculated for the LDG of marine phytoplankton. The communities of the southern temperate latitudes have lower similarity, are relatively even and exhibit richness on a par with the northern temperate region. The lower evenness, higher similarity and similar richness in the northern temperate region results in a lower LCI diversity value. In the tropical regions, higher similarity of communities there, despite high species richness reduces the overall LCI diversity measurement. Therefore, when taxonomic diversity is considered in tandem with richness and evenness, diversity is greatest in the southern hemisphere with the PhytoBase dataset.

Alongside the latitudinal diversity assessment, I performed analysis of fine-grain scale diversity globally using Hill numbers and LCI. The alpha species richness of marine phytoplankton has hotspots in coastal regions. Due to the greater accessibility of the coastal system to sampling effort, it is possible high richness in these regions is due to increased effort. The equivalent alpha sampling effort analysis suggests this may be the case with greater sampling effort observed in marine phytoplankton richness hotspots. In the LCI alpha diversity assessment, the results are mixed with respect to the sampling effort analysis. In the LCI alpha diversity assessment the diversity hotspots shift from coastal systems and spread to temperate regions. Notably, the diversity of marine phytoplankton in the eastern China Sea reduces when taxonomic similarity is considered. The hotspot of diversity in Peru remains. Temperate locations such as the coastal and southern Pacific region around Australia, coastal Greenland and the Persian Gulf show hotspots of LCI diversity.

These assessments of marine phytoplankton diversity beyond species richness by applying the Hill Number framework and LCI are the first to be performed. This research highlights the importance of careful consideration of diversity indices and will inform future work on marine phytoplankton macroecology.

The results of the species distribution modelling of marine phytoplankton global diversity show a similar bimodal LDG species richness pattern to those observed using the observational data which is biased spatially. The 283 species modelled were the most spatially abundant species, the choice of these species was driven by the assumption they represent the most productive species within the PhytoBase dataset. Therefore, as seen in the LDG assessment using Hill numbers, their diversity pattern aligns with the observational diversity analysis when $q = 5$, representing the diversity of the most abundant species. There is limited research on the global macroecology of marine phytoplankton and subsequent diversity patterns for cross validation of the results observed in this SDM ensemble research.

The only existing work on the global distribution of marine phytoplankton using SDM methods (Righetti et al., 2019) differs from the results of my work in my third chapter. Firstly, I modelled 283 species while they modelled 567 species while also performing an aggregated average of monthly model runs compared to my approach of decadal averages. This may lead to the following differences. The SDM results of my third chapter indicate species richness declines at the equator with sub-tropical maxima compared to Righetti et. al's work shows a tropical maximum. Furthermore, there is a high general spatial spread of marine phytoplankton richness across longitudes, compared to my SDM results showing high diversity in coastal systems and

locations of oceanographic boundaries. This difference and findings in my SDM approach led to the development of the SDM diversity-productivity analysis performed using AQUA-MODIS Chl-a satellite data as a proxy for biomass. While future work is required to assess SDM approaches used here and those used by Righetti et al's group to understand the specific modelling parameterisation that lead to these longitudinal differences in distribution of marine phytoplankton.

The finding of a strong correlative analysis of SDM diversity and the full mission composite of global chlorophyll-a suggests that there is a diversity-productivity relationship for global marine phytoplankton. This indicated that locations of high primary productivity support high diversity of marine phytoplankton. To date no such relationship has been found at a global scale for marine phytoplankton. As with all macroecological studies there are numerous contributing factors which may enhance or reduce diversity as species level. Notably, these results indicate that the metabolic theory of ecology is less pronounced than that of the productivity hypothesis.

5.2 Choosing a diversity index, is there no wrong choice?

The core benefit of assessing the LDG of marine phytoplankton while expanding the traditional approach of species richness measurement is increased discernability for spatial conservation decisions. Species richness is a restrictive measurement of diversity, when comparing two locations for relative diversity, richness alone cannot differentiate between two locations of equal richness (Leinster & Cobbold, 2012a). By incorporating richness and evenness in my global assessment of marine phytoplankton diversity using Hill numbers, I observed the benefit of greater information gained regarding the macroecology of marine phytoplankton that would be otherwise unavailable with species richness alone. The decline in evenness of communities in the north temperate region warrants further investigation of the reason for this deviance from the initial richness assessment. It is of note that the North Atlantic sampling regime uses the Continuous Plankton Recorder system which does not capture smaller species, reducing the overall community diversity sampled. This may be the reason for the lower evenness, if smaller species are consistently more abundant, which may also rectify the dip in taxonomic evenness observed in Chapter 3 - Figure 3.13.

When the richness, evenness and similarity inclusive LCI is applied, the deviance leads to higher diversity of marine phytoplankton located in southern temperate regions. The movement of the maxima of marine phytoplankton diversity observed depending on the mathematical construction of diversity index chosen suggests there is more to be learned beyond marine phytoplankton macroecology. While sampling biases reduce the strength of these results, the rarefaction shows that less samples are required to saturate LCI diversity measures. As such, this thesis has focused on marine phytoplankton, do we find similar differences in global diversity patterns of other groups such as pinnipeds, mammals or birds when the LCI is used. As noted, the LCI has a higher capacity to distinguish different diversity values in different regions owing to the inclusion of taxonomic similarity, how this affects other group macroecological patterns is of interest if the sampling biases can be overcome in future research development. This work advances the field of macroecology beyond the species richness approach which has characterised global macroecology studies to date. The current disparity

in the lack of uptake of more meaningful measures of diversity currently applied is now more readily circumvented though the method devised and originated in the first two chapters of this thesis.

All diversity indices have their benefits and drawbacks, notably species richness is readily performed for easily observed species. The LCI index requires deeper information, relative abundances and a measure of similarity. This could restrict uptake by the broader ecological scientific community yet as I have shown for marine phytoplankton, it is attainable. Furthermore, for species larger than marine phytoplankton, taxonomic information and classification only increases. Community evenness is also not easily attained and requires appropriate counting methods to quantify. This too is readily performed at local scales and as observed in my first two chapters, can be approximated sufficiently. The results of my first two chapters highlight how this gap in diversity index complexity and scale of diversity assessment can be rectified.

In a period of time where we are driven to protect and conserve biodiversity substantially, with 30% of land and marine area protected by 2030 (Stephens, 2023). We are then faced with the choice of where to protect and conserve. I would argue that the results of my first two research chapters of using a range of diversity indices, species richness alone is not up to the task and more complex diversity indices such as LCI are required to quantify and differentiate conservation priorities. I state this as the number of observations to saturate species richness is substantially more than that of the LCI, therefore the LCI is more efficient. The LCI also presents a wider picture of diversity than that of richness as recognising the relatedness of species is a truer picture of the evolutionary history leading to species richness.

5.3 LDG research in the 21st Century

Despite 300 years of observation and research, a unifying driver of the LDG across fauna and flora has not been found. Recent research points to the effect of temperature on metabolism (MTE) as the main driver (Righetti et al., 2019; Tittensor et al., 2010b) but there are deviations to this expected relationship. Additionally, there is potentially also diversity-productivity relationships observed (Dutkiewicz et al., 2020) and in the research findings of the third research chapter of this thesis. The wide array of possible drivers has been reviewed extensively (Pianka, 1966; Rohde, 1992). The confounding issue of LDG studies is the ‘same effect-many drivers’ outcome of the latitudinal biological-abiotic correlative relationship analysis. Latitudinal gradients, from northern to southern polar regions, of predictors such as temperature, seasonality, photosynthetically active radiation and ice age glacial coverage decline lead to covarying predictors of species richness.

One of the longstanding challenges in macroecology is identifying a unifying driver of biodiversity patterns across taxa. This challenge is even more pronounced due to the absence of a universally agreed-upon latitudinal diversity gradient (LDG) pattern. Traditionally, LDG research assumed a peak in diversity at the equator (Barton et al., 2010; Righetti et al., 2019; Tittensor & Worm, 2016), but recent studies have demonstrated variations in this pattern (Chaudhary et al., 2021; Visser et al., 2014), including latitudinal shifts and the influence of sampling biases (Menegotto & Rangel, 2018). These inconsistencies make it difficult to

attribute diversity trends to abiotic factors, particularly when the observed patterns change spatially in response to predictor variables.

Recent modeling efforts suggest that phytoplankton diversity is broadly distributed across latitudes but exhibits spatial variations influenced by environmental conditions (Righetti et al., 2019). Our research, using similar datasets and methodologies, supports these findings while also highlighting a key discrepancy: the spatial contribution of diversity is particularly sensitive to the inclusion of spatially dominant species which may lead to broad distributional ranges due to the lack of interactivity between species in the SDM approach. This raises fundamental questions about how species with widespread versus localised distributions shape global LDG patterns. Additionally, collinearity among predictor variables, variations in sampling effort, and the scale-dependent nature of biodiversity assessments continue to pose challenges for LDG studies. These issues have persisted for centuries, underscoring the complexity of the problem.

A critical challenge in LDG research is how we define and quantify diversity. Traditional species distribution models have primarily focused on species richness, but they have yet to incorporate relative abundance and similarity measures effectively. The development of the LCI to the macroecology of marine phytoplankton is a promising avenue, but it introduces further complexity in distinguishing which components of diversity lead to deviations from the richness patterns observed. The application of the decomposition method resolves this issue (Chen & Grinfeld, 2024). As it can detail the independent components of diversity, richness, evenness and similarity from the LCI.

A major unresolved issue in phytoplankton research is whether these organisms are globally dispersed or whether species distributions are more constrained than previously assumed (Ward et al., 2021). While preliminary data suggest that both patterns may coexist, some species are truly cosmopolitan, while others are localised, this hypothesis has yet to be tested explicitly.

Given the persistent uncertainties and methodological limitations in LDG research, it may be time to reconsider how we conceptualise and study these patterns. Rather than searching for a single unifying driver of biodiversity, a more productive approach may be to embrace a multi-method framework that accounts for taxon-specific dispersal abilities, habitat preferences, and evolutionary constraints on a taxonomic group basis. This could be potentially achieved by combining mechanistic dispersal modelling approaches from Dutkiewicz et al, 2020, increased tropic interactions shown in Neil Bana's work in 2011. In addition, the SDM approach while limited to model tropic interactions and mechanistic dispersal, does retain species identities and is based in observations, providing a connection to present and future observational data collection. If more functional data such as size-spectra can be included in the observational data for the SDM approach, I hypothesise this could provide a link of observational species identification to the size-based approaches of Neil Banas and the Dutkiewicz group.

Over the next several decades, climate change will disrupt established biodiversity patterns (Garcia et al., 2014; Trew & Maclean, 2021; Yasuhara et al., 2020), making it even more difficult to identify universal LDG drivers. Increasing variability of global temperatures, particularly in marine environments, could alter the spatial organisation of biodiversity and weaken previously observed latitudinal trends. If LDG patterns become less structured over

time, this could further obscure efforts to distinguish the relative contributions of climate change versus preexisting environmental gradients.

To advance LDG research in the 21st century, we must refine our methodological approaches, acknowledge the inherent complexity of biodiversity patterns, and integrate local-scale processes into global models. My findings indicate that more sampling is required to describe the latitudinal diversity gradient of marine phytoplankton. Until then, the LCI approach is more insightful than that of the species richness approach and future macroecological work should consider using more descriptive measures of diversity. In addition, while SDMs are limited by data availability, useful data products with predictive capacity can be obtained such as chlorophyll-a community composition. While the environmental drivers of marine phytoplankton diversity continue to be examined, we should use available knowledge to assess community composition on macroecological scales, albeit for a limited subset of species of the total species pool. Addressing the longstanding issues of sampling effort, predictor collinearity, and taxon-specific responses will be critical. Furthermore, as climate change continues to reshape biodiversity distributions, LDG research must adapt to capture these dynamic shifts which means applying and refine approaches which are more descriptive of the observed world such as the LCI and SDM approaches used in this thesis. While these approaches are limited by sampling effort, they are less so than that of Hill numbers owing to utilisation of information such as taxonomy which is broadly available for described species. Ultimately, resolving these challenges requires a shift in focus from seeking a singular explanation for LDG patterns to developing a more nuanced understanding of biodiversity drivers across spatial and temporal scales.

5.4 General limitations

5.4.1 Observational data biases

The analyses performed in this thesis are reliant on observational data collection. To date, global marine phytoplankton data is concentrated to coastal and the north Atlantic region. As a result, conclusions drawn from spatial analysis of marine phytoplankton diversity have to reckon with these biases. The observed marine phytoplankton diversity is substantially lower than that expected with global asymptotic estimation (Mora et al., 2011). The rarefaction approach was performed on broad latitudinal zones, which lead to general conclusions regarding the validity of the LDG patterns drawn. These patterns were drawn at 1-degree latitudes and partitioned equal area grid assessments. The rarefaction method was not applied to this resolution due to the low areas of sampling effort per diversity area partitioning method nor were they applied at finer scales owing to time constraints in this research. This serves as a starting point to example localised diversity patterns in response to sampling effort performed in these settings. To ensure the rarefaction method was comparable, the number of subsamples performed was limited to 1000 continuous and 20000 discontinuous. Therefore, the effect of sampling effort was assessed in a limited manner and further work is required to expand on this methodological approach.

It follows that the sampling effort disparity globally for marine phytoplankton requires attention. The low sampling effort in the pelagic system of the Pacific Ocean and the Indian

ocean leads to substantial gaps in confidence in macroecological patterns drawn. To rectify the spatial disparity, I applied the SDM methodology to redistribute well observed species globally based on global abiotic predictors of habitat suitability. I assume that the most well observed species are more likely to have sufficient niche space characterisation to model their distributions. The result of the SDM approach continued to show high diversity in regions of upwelling and large boundary current systems. Whether this is a result of sampling biases or a representation of the macroecology of well observed species of marine phytoplankton can only be rectified by increased sampling in regions of low sampling.

A large component of the work in this thesis is reliant on taxonomic classification of marine phytoplankton species. As a result, species with limited taxonomic classification were removed from the analysis. Two major genera of phytoplankton were removed as a result, *Prochlorococcus* and *Synechococcus*. These genera were recognised relatively recently as major contributors of marine primary productivity, so to not model these species or include their diversity is a limitation of the LDG correlative relationships drawn. The rationale for their removal was that they were not classified to species level. This is due to their very small sizes, representing a limitation of reliance on taxonomic diversity for marine phytoplankton. Phytoplankton representative of a group of species which require different methods of classification due to their size. Consolidating this disparity in methods for classification is required to capture a broader perspective on the LDG of marine phytoplankton diversity.

Functional diversity has been assessed broadly for larger species due to higher data availability and societal preferences, where charismatic species gain more attention and thus funding(Etard et al., 2020; Troudet et al., 2017). Marine phytoplankton functional diversity is an important area of research for ecosystem functioning and health. The LCI is a useful measure in this regard with the flexibility of the definition of similarity. It is possible to perform the same diversity assessment with genetic or functional similarity information. I did not elect to do so as these definitions of similarity require explicit formulation of similarity quantification. Taxonomic diversity is partitioned into an organisational structure of comparison i.e. species to species, genus to genus and so on. Genetic similarity requires in depth genetic knowledge to ensure appropriate areas of genetic code are compared and quantified. Functional diversity operates in continuous space predominantly i.e. body size is a continuous spectrum, how this is partitioned in a non-biased manner requires further research.

5.4.2 Lack of temporality in diversity assessments performed

The research of this thesis focused on time-independent analysis of the LDG of marine phytoplankton. The only element of time was capturing all time points sampled both in observational data and environmental covariates of marine phytoplankton diversity. As a result, these results are highly temporally generalised. As a result, the findings are constrained by this approach and are unable to describe how the LDG operates seasonally. For a group of marine phytoplankton with communities structured by seasonal changes in nutrients, mixed layer depth and temperature, the lack of seasonality may lead to a limited perspective on how the LDG changes temporally on an annual basis. The choice to exclude time from this analysis was motivated by the longstanding existence of LDG patterns spanning beyond the time period of

data collection. Therefore, any patterns drawn using the methodology outlined in this thesis assume the pattern is a general snapshot but can not detail the effect of temporality on the LDG.

5.5 Applications

This thesis was motivated by the increased capacity to estimate differences in diversity using a similarity-sensitive index. The applicability of the methods designed in this research will serve to assess other fauna and flora diversity. The benefit of assessing a complex problem like diversity with indices which contain more information of the system studied is deeper understanding of the problem itself. Furthermore, the ability of this approach to differentiate areas of diversity that have similar richness and evenness patterns is paramount to selection of areas for protection and conservation. In general, any research assessing diversity can do so with greater characterisation using the methods outlined in this thesis. I show in this thesis that while the LCI index confounds three diversity components into a single value, the decomposition computational method designed for marine phytoplankton here can be applied to any taxonomic group.

Richness is often used due to its simplicity but as we now aim to conserve far broader and general systems outside of the typical high richness regions like coral reef systems, we need to move beyond richness. We should aim to conserve as much ‘information’ as possible in our protective area networks. As we have observed in our work, the richness paradigm does not necessarily line up the ecologically grounded assessment of diversity while considering the relative similarity of each species in the community. Therefore, we should aim to consider a deeper level of diversity to maximise our potential conservation of ‘information’. It may be crude to reduce the diversity of form and function of life into pure information, but it is this information that was are seeking to conserve and protect. We have finite resources to do so and are beholden to societal perspectives on how much we protect and conserve, so any protection is able to achieve to prioritise the maximum amount of information conserved possible.

5.6 Future research

There are a broad range of future directions motivated by the results of this thesis. Firstly, assessing the generality of the findings observed for marine phytoplankton LDG using the LCI. This work has focused on marine phytoplankton, the next step is assessing other fauna and flora to assess whether the deviation of the richness LDG to the LCI LDG is a common pattern. Currently this is unknown and warrants further research. The added benefit is other taxa may have more complete species inventories such as sharks and coastal benthic fauna such as bivalves. As result, if the southern peak in LCI diversity is observed in other taxa it may point to a broader general pattern or determine the role of sampling effort in marine phytoplankton analysis.

Additionally, this thesis was focused on drawing macroecological patterns, particularly the LDG of marine phytoplankton. The next step is to assess how correlations in environmental predictors differ depending on choice of diversity index. As the LDG changes from Hill numbers to LCI, it is of interest to see how abiotic predictors may change depending on index used. This next step in this research may aid in distinguishing between environmental drivers

of diversity and evolutionary drivers using contemporary data availability. This may arise as taxonomic diversity may be greater in regions where species originated versus where there is greater environmental niche space to host more species. Secondly, this work has focused one functional group, marine phytoplankton, there is not explicit limitation in assessing the diversity of ecosystems. The benefit of this approach may highlight how diversity differs from benthic, coastal and pelagic ecosystems and the degree of redundancy in these ecosystems. Thirdly, the SDM correlative finding with satellite chlorophyll-a suggests a broader applicability of SDM methods to estimate future and hindcast historic chlorophyll-a patterns. The limited window of satellite observation limits our capacity for predictions of primary productivity.

5.7 Concluding statements

The aim of this thesis was to assess the marine phytoplankton macroecology for alignment to the LDG and the effect of diversity index on patterns observed. Furthermore, the thesis was also motivated to assess potential drivers of marine macroecological patterns. In conclusion, marine phytoplankton follow the LDG richness pattern if a bimodal pattern is included in the LDG paradigm. If not, then marine phytoplankton deviate from the expected pattern of the LDG with a drop in diversity at the equator. The effect of choice of diversity index is apparent with a shift in the location of peak diversity in marine phytoplankton. Going forward, macroecological diversity assessments need to broaden their horizons and move beyond richness.

6 Bibliography

Acevedo-Trejos, E., Brandt, G., Bruggeman, J., & Merico, A. (2015a). Mechanisms shaping size structure and functional diversity of phytoplankton communities in the ocean. *Scientific Reports*, *5*(1), 8918. https://doi.org/10.1038/srep08918

Acevedo-Trejos, E., Brandt, G., Bruggeman, J., & Merico, A. (2015b). Mechanisms shaping size structure and functional diversity of phytoplankton communities in the ocean. *Scientific Reports*, *5*(1), 8918. https://doi.org/10.1038/srep08918

Acevedo-Trejos, E., Marañón, E., & Merico, A. (2018a). Phytoplankton size diversity and ecosystem function relationships across oceanic regions. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20180621. https://doi.org/10.1098/rspb.2018.0621

Acevedo-Trejos, E., Marañón, E., & Merico, A. (2018b). Phytoplankton size diversity and ecosystem function relationships across oceanic regions. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20180621. https://doi.org/10.1098/rspb.2018.0621

Alatalo, R., & Alatalo, R. (1977). Components of Diversity: Multivariate Analysis with Interaction. *Ecology*, *58*(4), 900–906. https://doi.org/10.2307/1936226

Anderson, C. B. (2018). Biodiversity monitoring, earth observations and the ecology of scale. *Ecology Letters*, *21*(10), 1572–1585. https://doi.org/10.1111/ele.13106

AQUA MODIS Level-3 Mapped Photosynthetically Available Radiation, Version 2024. (2024). [Dataset]. NASA Ocean Biology Distributed Active Archive Center.

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology*, *11*(9), 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Aristotle. (1887). *Aristotle's History of Animals: In Ten Books*. George Bell & Sons.

Assis, J., Fernández Bejarano, S. J., Salazar, V. W., Schepers, L., Gouvêa, L., Fragkopoulou, E., Leclercq, F., Vanhoorne, B., Tyberghein, L., Serrão, E. A., Verbruggen, H., & De Clerck, O. (2024). Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Global Ecology and Biogeography*, *33*(4), e13813. https://doi.org/10.1111/geb.13813

Atkinson, A., Rossberg, A. G., Gaedke, U., Sprules, G., Heneghan, R. F., Batziakas, S., Grigoratou, M., Fileman, E., Schmidt, K., & Frangoulis, C. (2024). Steeper size spectra with decreasing phytoplankton biomass indicate strong trophic amplification and future fish declines. *Nature Communications*, *15*(1), 381. https://doi.org/10.1038/s41467-023-44406-5

Avise, J. C. (2000). *Phylogeography: The History and Formation of Species*. Harvard University Press.

Báez, J. C., Barbosa, A. M., Pascual, P., Ramos, M. L., & Abascal, F. (2020). Ensemble modeling of the potential distribution of the whale shark in the Atlantic Ocean. *Ecology and Evolution*, *10*(1), 175–184. https://doi.org/10.1002/ece3.5884

Banas, N. S. (2011). Adding complex trophic interactions to a size-spectral plankton model: Emergent diversity patterns and limits on predictability. *Ecological Modelling*, *222*(15), 2663–2675. https://doi.org/10.1016/j.ecolmodel.2011.05.018

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. (2010). Patterns of Diversity in Marine Phytoplankton. *Science*, *327*(5972), 1509–1511. https://doi.org/10.1126/science.1184961

Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities.

Proceedings of the National Academy of Sciences, *113*(11), 2964–2969. https://doi.org/10.1073/pnas.1519080113

Beardall, J., Allen, D., Bragg, J., Finkel, Z. V., Flynn, K. J., Quigg, A., Rees, T. A. V., Richardson, A., & Raven, J. A. (2009). Allometry and stoichiometry of unicellular, colonial and multicellular phytoplankton. *New Phytologist*, *181*(2), 295–309. https://doi.org/10.1111/j.1469-8137.2008.02660.x

Beaugrand, G., Edwards, M., Raybaud, V., Goberville, E., & Kirby, R. R. (2015). Future vulnerability of marine biodiversity compared with contemporary and past changes. *Nature Climate Change*, *5*(7), 695–701. https://doi.org/10.1038/nclimate2650

Benedetti, F., Wydler, J., Clerc, C., Knecht, N., & Vogt, M. (2025). Emergent Relationships Between the Functional Diversity of Marine Planktonic Copepods and Ecosystem Functioning in the Global Ocean. *Global Change Biology*, *31*(3), e70094. https://doi.org/10.1111/gcb.70094

Bernot, J., Ahyong, S., Boyko, C. B., Bailly, N., Bieler, R., Brandão, S. N., Daly, M., De Grave, S., Gofas, S., Hernandez, F., Hughes, L., Neubauer, T. A., Paulay, G., Boydens, B., Decock, W., Dekeyzer, S., Goharimanesh, M., Vandepitte, L., Vanhoorne, B., … Zullini, A. (2024, August 1). *World register of marine species (WoRMS)*. WoRMS Editorial Board. https://www.marinespecies.org

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., Moyes, F., Magurran, A., McGill, B., Myers-Smith, I. H., Winter, M., Bjorkman, A. D., Bowler, D. E., Byrnes, J. E. K., Gonzalez, A., Hines, J., Isbell, F., Jones, H. P., Navarro, L. M., … Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, *366*(6463), 339–345. https://doi.org/10.1126/science.aaw1620

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 326–349. https://doi.org/10.2307/1942268

Brooks, J. J., Markegard, S. I., Langdon, S. J., Anderstrom, D. S. É., Douville, M. G., George, T. A., Jackson, M. K., Jackson, S. G., Mills, T. K̲aachkutí, Ramos, J. D., Rowan, J. Y., Sanderson, T., & Smythe, C. (2024). Indigenous knowledge and species assessment for the Alexander Archipelago wolf: Successes, challenges, and lessons learned. *The Journal of Wildlife Management*, *88*(6), e22563. https://doi.org/10.1002/jwmg.22563

Brotas, V., Ferreira, A., Veloso, V., Tracana, A., Guerreiro, C. V., Tarran, G. A., Woodward, E. M. S., Ribeiro, L., Netting, J., Clewley, D., & Groom, S. B. (2023). Assessing phytoplankton community composition in the Atlantic Ocean from in situ and satellite observations. *Frontiers in Marine Science*, *10*, 1229692. https://doi.org/10.3389/fmars.2023.1229692

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, *41*(1), 8–22. https://doi.org/10.1111/jbi.12228

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*(7), 1771–1789. https://doi.org/10.1890/03-9000

Brun, P., Vogt, M., Payne, M. R., Gruber, N., O'Brien, C. J., Buitenhuis, E. T., Le Quéré, C., Leblanc, K., & Luo, Y.-W. (2015). Ecological niches of open ocean phytoplankton taxa. *Limnology and Oceanography*, *60*(3), 1020–1038. https://doi.org/10.1002/lno.10074

Burrows, M. T., Schoeman, D. S., Buckley, L. B., Moore, P., Poloczanska, E. S., Brander, K. M., Brown, C., Bruno, J. F., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., Kiessling, W., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F. B., Sydeman, W. J., & Richardson, A. J. (2011). The Pace of Shifting Climate in Marine

and Terrestrial Ecosystems. *Science*, *334*(6056), 652–655. https://doi.org/10.1126/science.1210288

Carmichael, B., Koumou, G. B., & Moran, K. (2023). Unifying Portfolio Diversification Measures Using Rao's Quadratic Entropy. *Journal of Quantitative Economics*, *21*(4), 769–802. https://doi.org/10.1007/s40953-023-00368-5

Cazzolla Gatti, R., Amoroso, N., & Monaco, A. (2020). Estimating and comparing biodiversity with a single universal metric. *Ecological Modelling*, *424*, 109020. https://doi.org/10.1016/j.ecolmodel.2020.109020

Cermeño, P., De Vargas, C., Abrantes, F., & Falkowski, P. G. (2010). Phytoplankton Biogeography and Community Stability in the Ocean. *PLoS ONE*, *5*(4), e10037. https://doi.org/10.1371/journal.pone.0010037

Cermeño, P., Marañín, E., Harbour, D., Figueiras, F. G., Crespo, B. G., Huete-Ortega, M., Varela, M., & Harris, R. P. (2008). Resource levels, allometric scaling of population abundance, and marine phytoplankton diversity. *Limnology and Oceanography*, *53*(1), 312–318. https://doi.org/10.4319/lo.2008.53.1.0312

Cermeño, P., Rodríguez-Ramos, T., Dornelas, M., Figueiras, F., Marañón, E., Teixeira, I., & Vallina, S. (2013). Species richness in marine phytoplankton communities is not correlated to ecosystem productivity. *Marine Ecology Progress Series*, *488*, 1–9. https://doi.org/10.3354/meps10443

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 297–324. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Chaudhary, C., Richardson, A. J., Schoeman, D. S., & Costello, M. J. (2021). Global warming is causing a more pronounced dip in marine species richness around the equator. *Proceedings of the National Academy of Sciences*, *118*(15), e2015094118. https://doi.org/10.1073/pnas.2015094118

Chaudhary, C., Saeedi, H., & Costello, M. J. (2016). Bimodality of Latitudinal Gradients in Marine Species Richness. *Trends in Ecology & Evolution*, *31*(9), 670–676. https://doi.org/10.1016/j.tree.2016.06.001

Chen, B. (2015). Patterns of thermal limits of phytoplankton. *Journal of Plankton Research*, *37*(2), 285–292. https://doi.org/10.1093/plankt/fbv009

Chen, B., & Grinfeld, M. (2024). Decomposing diversity into measures of evenness, similarity, and richness. *Ecology and Evolution*, *14*(2), e10952. https://doi.org/10.1002/ece3.10952

Chust, G., Irigoien, X., Chave, J., & Harris, R. P. (2013). Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography*, *22*(5), 531–543. https://doi.org/10.1111/geb.12016

Clayton, S., Dutkiewicz, S., Jahn, O., & Follows, M. J. (2013). Dispersal, eddies, and the diversity of marine phytoplankton. *Limnology and Oceanography: Fluids and Environments*, *3*(1), 182–197. https://doi.org/10.1215/21573689-2373515

Colwell, R. K., & Coddington, J. A. (1994). Estimating Terrestrial Biodiversity through Extrapolation. *Philosophical Transactions: Biological Sciences*, *345*(1311), 101–118.

Colwell, R. K., Rahbek, C., & Gotelli, N. J. (2004). The Mid-Domain Effect and Species Richness Patterns:What Have We Learned So Far? *The American Naturalist*, *163*(3), E1–E23. https://doi.org/10.1086/382056

Currie, D. J., & Kerr, J. T. (2008). Tests of the Mid-Domain Hypothesis: A Review of the Evidence. *Ecological Monographs*, *78*(1), 3–18. https://doi.org/10.1890/06-1302.1

Daly, A. J., Baetens, J. M., & De Baets, B. (2018). Ecological Diversity: Measuring the Unmeasurable. *Mathematics*, *6*(7), 119. https://doi.org/10.3390/math6070119

Darwin, C., 1809-1882. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859. https://search.library.wisc.edu/catalog/9934839413602122

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., … Karsenti, E. (2015). *First Tara Oceans V9 Rdna Metabarcoding Dataset.* https://doi.org/10.5281/ZENODO.15600

Deser, C., Alexander, M. A., Xie, S.-P., & Phillips, A. S. (2010). Sea Surface Temperature Variability: Patterns and Mechanisms. *Annual Review of Marine Science*, *2*(1), 115–143. https://doi.org/10.1146/annurev-marine-120408-151453

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471), eaax3100. https://doi.org/10.1126/science.aax3100

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press.

Doney, S. C., Fabry, V. J., Feely, R. A., & Kleypas, J. A. (2009). Ocean Acidification: The Other CO_2 Problem. *Annual Review of Marine Science*, *1*(1), 169–192. https://doi.org/10.1146/annurev.marine.010908.163834

Dutkiewicz, S., Boyd, P. W., & Riebesell, U. (2021). Exploring biogeochemical and ecological redundancy in phytoplankton communities in the global ocean. *Global Change Biology*, *27*(6), 1196–1213. https://doi.org/10.1111/gcb.15493

Dutkiewicz, S., Cermeno, P., Jahn, O., Follows, M. J., Hickman, A. E., Taniguchi, D. A. A., & Ward, B. A. (2020). Dimensions of marine phytoplankton diversity. *Biogeosciences*, *17*(3), 609–634. https://doi.org/10.5194/bg-17-609-2020

Early, R., & Keith, S. A. (2019). Geographically variable biotic interactions and implications for species ranges. *Global Ecology and Biogeography*, *28*(1), 42–53. https://doi.org/10.1111/geb.12861

Egorova, Y., Reygondeau, G., Cheung, W. W. L., & Pakhomov, E. A. (2025). Species Distribution Models for Mesopelagic Mesozooplankton Community. *Journal of Biogeography*, *52*(1), 13–26. https://doi.org/10.1111/jbi.15011

El Hourany, R., Pierella Karlusich, J., Zinger, L., Loisel, H., Levy, M., & Bowler, C. (2024). Linking satellites to genes with machine learning to estimate phytoplankton community structure from space. *Ocean Science*, *20*(1), 217–239. https://doi.org/10.5194/os-20-217-2024

Elith, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., … E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*, *29*(12), 2143–2158. https://doi.org/10.1111/geb.13184

Eyre, A. C., Briscoe, N. J., Harley, D. K. P., Lumsden, L. F., McComb, L. B., & Lentini, P. E. (2022). Using species distribution models and decision tools to direct surveys and identify potential translocation sites for a critically endangered species. *Diversity and Distributions*, *28*(4), 700–711. https://doi.org/10.1111/ddi.13469

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & Taylor, F. J. R. (2004). The Evolution of Modern Eukaryotic Phytoplankton. *Science*, *305*(5682), 354–360. https://doi.org/10.1126/science.1095964

Fenton, I. S., Aze, T., Farnsworth, A., Valdes, P., & Saupe, E. E. (2023). Origination of the modern-style diversity gradient 15 million years ago. *Nature*, *614*(7949), 708–712. https://doi.org/10.1038/s41586-023-05712-6

Ferrer-Valero, N. (2018). Measuring geomorphological diversity on coastal environments: A new approach to geodiversity. *Geomorphology*, *318*, 217–229. https://doi.org/10.1016/j.geomorph.2018.06.013

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Finlay, B. J. (2002). Global Dispersal of Free-Living Microbial Eukaryote Species. *Science*, *296*(5570), 1061–1063. https://doi.org/10.1126/science.1070710

Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology*, *12*(1), 42–58. https://doi.org/10.2307/1411

Gagné, T. O., Reygondeau, G., Jenkins, C. N., Sexton, J. O., Bograd, S. J., Hazen, E. L., & Van Houtan, K. S. (2020). Towards a global understanding of the drivers of marine and terrestrial biodiversity. *PLOS ONE*, *15*(2), e0228065. https://doi.org/10.1371/journal.pone.0228065

Garcia, R. A., Cabeza, M., Rahbek, C., & Araújo, M. B. (2014). Multiple Dimensions of Climate Change and Their Implications for Biodiversity. *Science*, *344*(6183), 1247579. https://doi.org/10.1126/science.1247579

Gardner, J. L., Peters, A., Kearney, M. R., Joseph, L., & Heinsohn, R. (2011). Declining body size: A third universal response to warming? *Trends in Ecology & Evolution*, *26*(6), 285–291. https://doi.org/10.1016/j.tree.2011.03.005

Gaston, K. J., & Blackburn, T. M. (1999). A Critique for Macroecology. *Oikos*, *84*(3), 353–368. https://doi.org/10.2307/3546417

Gaston, K. J., & Chown, S. L. (1999). Why Rapoport's Rule Does Not Generalise. *Oikos*, *84*(2), 309–312. https://doi.org/10.2307/3546727

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of Size and Temperature on Metabolic Rate. *Science*, *293*(5538), 2248–2251. https://doi.org/10.1126/science.1061967

Gleason, H. A. (1922). On the Relation Between Species and Area. *Ecology*, *3*(2), 158–162. https://doi.org/10.2307/1929150

Goericke, R., & Welschmeyer, N. A. (1993). The marine prochlorophyte Prochlorococcus contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *40*(11–12), 2283–2294. https://doi.org/10.1016/0967-0637(93)90104-B

Gonçalves-Araujo, R., De Souza, M. S., Mendes, C. R. B., Tavano, V. M., & Garcia, C. A. E. (2016). Seasonal change of phytoplankton (spring vs. Summer) in the southern

Patagonian shelf. *Continental Shelf Research*, *124*, 142–152. https://doi.org/10.1016/j.csr.2016.03.023
Grattepanche, J.-D., Jeffrey, W. H., Gast, R. J., & Sanders, R. W. (2022). Diversity of Microbial Eukaryotes Along the West Antarctic Peninsula in Austral Spring. *Frontiers in Microbiology*, *13*, 844856. https://doi.org/10.3389/fmicb.2022.844856
Guiry, M.D. & Guiry, G.M. (2023). AlgaeBase. *World-Wide Electronic Publication, National University of Ireland, Galway*. https://www.algaebase.org
Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models: With Applications in R* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781139028271
Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435. https://doi.org/10.1111/ele.12189
Hammond, P. S., Francis, T. B., Heinemann, D., Long, K. J., Moore, J. E., Punt, A. E., Reeves, R. R., Sepúlveda, M., Sigurðsson, G. M., Siple, M. C., Víkingsson, G., Wade, P. R., Williams, R., & Zerbini, A. N. (2021). Estimating the Abundance of Marine Mammal Populations. *Frontiers in Marine Science*, *8*, 735770. https://doi.org/10.3389/fmars.2021.735770
Hardin, G. (1960). The Competitive Exclusion Principle. *Science*, *131*(3409), 1292–1297.
Harley, C. D. G., Randall Hughes, A., Hultgren, K. M., Miner, B. G., Sorte, C. J. B., Thornber, C. S., Rodriguez, L. F., Tomanek, L., & Williams, S. L. (2006). The impacts of climate change in coastal marine systems. *Ecology Letters*, *9*(2), 228–241. https://doi.org/10.1111/j.1461-0248.2005.00871.x
Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, *12*(1), 5372. https://doi.org/10.1038/s41467-021-25699-w
Hill, M. O. (1973a). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352
Hill, M. O. (1973b). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352
Hillebrand, H. (2004). On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist*, *163*(2), 192–211. https://doi.org/10.1086/381004
Hillebrand, H., Acevedo-Trejos, E., Moorthi, S. D., Ryabov, A., Striebel, M., Thomas, P. K., & Schneider, M.-L. (2022). Cell size as driver and sentinel of phytoplankton community structure and functioning. *Functional Ecology*, *36*(2), 276–293. https://doi.org/10.1111/1365-2435.13986
Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., & Yamanaka, Y. (2011). Synoptic relationships between surface Chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, *8*(2), 311–327. https://doi.org/10.5194/bg-8-311-2011
Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
Humboldt, Alexander von, Baxter, George, Bohn, Henry G., & Otté, Elise. C. (1850). *Views of nature, or, Contemplations on the sublime phenomena of creation: With scientific illustrations*. Henry G. Bohn, London. Biodiversity Heritage Library.
Hutchinson, G. E. (1961). The Paradox of the Plankton. *The American Naturalist*, *882*.

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., Endo, H., Gasol, J. M., Gregory, A. C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco, E., Soviadan, D., Zayed, A. A., Zingone, A., … Wincker, P. (2019). Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*, *179*(5), 1084-1097.e21. https://doi.org/10.1016/j.cell.2019.10.008

Irigoien, X., Huisman, J., & Harris, R. P. (2004). Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature*, *429*(6994), 863–867. https://doi.org/10.1038/nature02593

Irwin, A. J., Nelles, A. M., & Finkel, Z. V. (2012). Phytoplankton niches estimated from field data. *Limnology and Oceanography*, *57*(3), 787–797. https://doi.org/10.4319/lo.2012.57.3.0787

Jean-Michel, L., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L. G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D., & Pierre-Yves, L. T. (2021). The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. *Frontiers in Earth Science*, *9*, 698876. https://doi.org/10.3389/feart.2021.698876

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, *110*(28). https://doi.org/10.1073/pnas.1302251110

Johnson, S., Molano-Flores, B., & Zaya, D. (2023). Field validation as a tool for mitigating uncertainty in species distribution modeling for conservation planning. *Conservation Science and Practice*, *5*(8), e12978. https://doi.org/10.1111/csp2.12978

Jonkers, L., Meilland, J., Rillo, M. C., de Garidel-Thoron, T., Kitchener, J. A., & Kucera, M. (2022). Linking zooplankton time series to the fossil record. *ICES Journal of Marine Science*, *79*(3), 917–924. https://doi.org/10.1093/icesjms/fsab123

Jost, L. (2010). The Relation between Evenness and Diversity. *Diversity*, *2*(2), Article 2. https://doi.org/10.3390/d2020207

Kaiser, M. J., & Williams, P. J. leB (Eds.). (2011). *Marine ecology: Processes, systems, and impacts* (2nd ed). Oxford University Press.

Kallimanis, A. S., Mazaris, A. D., Tzanopoulos, J., Halley, J. M., Pantis, J. D., & Sgardelis, S. P. (2008). How does habitat diversity affect the species–area relationship? *Global Ecology and Biogeography*, *17*(4), 532–538. https://doi.org/10.1111/j.1466-8238.2008.00393.x

Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, *9*(8), 605–618. https://doi.org/10.1038/nrg2386

Keith, S. A., Webb, T. J., Böhning-Gaese, K., Connolly, S. R., Dulvy, N. K., Eigenbrod, F., Jones, K. E., Price, T., Redding, D. W., Owens, I. P. F., & Isaac, N. J. B. (2012). What is macroecology? *Biology Letters*, *8*(6), 904–906. https://doi.org/10.1098/rsbl.2012.0672

Kok, M. T. J., Kok, K., Peterson, G. D., Hill, R., Agard, J., & Carpenter, S. R. (2017). Biodiversity and ecosystem services require IPBES to take novel approach to scenarios. *Sustainability Science*, *12*(1), 177–181. https://doi.org/10.1007/s11625-016-0354-8

Ladau, J., Sharpton, T. J., Finucane, M. M., Jospin, G., Kembel, S. W., O'Dwyer, J., Koeppel, A. F., Green, J. L., & Pollard, K. S. (2013). Global marine bacterial diversity peaks at high latitudes in winter. *The ISME Journal*, *7*(9), 1669–1677. https://doi.org/10.1038/ismej.2013.37

Lawton, J. H. (1999). Are There General Laws in Ecology? *Oikos*, *84*(2), 177–192. https://doi.org/10.2307/3546712

Leinster, T., & Cobbold, C. (2012a). Measuring diversity: The importance of species similarity. *Ecology*, *93*, 477–489. https://doi.org/10.2307/23143936
Leinster, T., & Cobbold, C. A. (2012b). Measuring diversity: The importance of species similarity. *Ecology*, *93*(3), 477–489. https://doi.org/10.1890/10-2402.1
Li, W. K. W. (2002). Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature*, *419*(6903), 154–157. https://doi.org/10.1038/nature00994
Linné, 1707-1778, Carl von. (1789). *Systema naturae per regna tria naturae: Secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Vol. 1, pt. 1 (13th edition)*. Lugduni : Apud J. B. Delamolliere. http://hdl.handle.net/10713/1921
Litchman, E. (1998). Population and community responses of phytoplankton to fluctuating light. *Oecologia*, *117*, 247–257.
Litchman, E., de Tezanos Pinto, P., Edwards, K. F., Klausmeier, C. A., Kremer, C. T., & Thomas, M. K. (2015). Global biogeochemical impacts of phytoplankton: A trait-based perspective. *Journal of Ecology*, *103*(6), 1384–1396. https://doi.org/10.1111/1365-2745.12438
Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, *11*(10), 995–1003. https://doi.org/10.1111/j.1461-0248.2008.01229.x
Lu, M., & Jetz, W. (2025). Fine-Grain Data Reveal Vulnerability of Biodiversity to Climate Change. *Global Change Biology*, *31*(12), e70627. https://doi.org/10.1111/gcb.70627
Macarthur, R. H. (1965). Patterns of Species Diversity. *Biological Reviews*, *40*(4), 510–533. https://doi.org/10.1111/j.1469-185X.1965.tb00815.x
MacArthur, R., & Wilson, E. (1967). *The Theory of Island Biogeography*.
Machac, A., Zrzavý, J., Smrckova, J., & Storch, D. (2012). Temperature dependence of evolutionary diversification: Differences between two contrasting model taxa support the metabolic theory of ecology. *Journal of Evolutionary Biology*, *25*(12), 2449–2456. https://doi.org/10.1111/jeb.12019
Magurran, A. E. (2004). *Measuring Biological Diversity* (1st ed.). Wiley-Blackwell, Oxford.
Malfertheiner, L., Tackmann, J., Matias Rodrigues, J. F., & Von Mering, C. (2026). Community conservatism is widespread across microbial phyla and environments. *Nature Ecology & Evolution*, *10*(2), 232–245. https://doi.org/10.1038/s41559-025-02957-4
Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., De Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences*, *113*(11). https://doi.org/10.1073/pnas.1509523113
Mannion, P. D. (2020). A deep-time perspective on the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences*, *117*(30), 17479–17481. https://doi.org/10.1073/pnas.2011997117
Mayr, E. (1996). What Is a Species, and What Is Not? *Philosophy of Science*, *63*(2), 262–277.
McCormick, W. P., Lyons, N. I., & Hutcheson, K. (1992). Distributional properties of jaccard's index of similarity. *Communications in Statistics - Theory and Methods*, *21*(1), 51–68. https://doi.org/10.1080/03610929208830764
McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I., & White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories

to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/https://doi.org/10.1111/j.1461-0248.2007.01094.x

Menegotto, A., & Rangel, T. F. (2018). Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nature Communications*, *9*(1), 4713. https://doi.org/10.1038/s41467-018-07217-7

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, *9*(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Moradi, M., & Zoljoodi, M. (2023). Validation of standard ocean-color chlorophyll-a products in turbid coastal waters: A case study on statistical evaluation and quality control tests in the Persian Gulf. *Journal of Marine Systems*, *240*, 103875. https://doi.org/10.1016/j.jmarsys.2023.103875

Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., Meiners, T., Müller, C., Obermaier, E., Prati, D., Socher, S. A., Sonnemann, I., Wäschke, N., Wubet, T., Wurst, S., & Rillig, M. C. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, *4*(18), 3514–3524. https://doi.org/10.1002/ece3.1155

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nielsen, J., Hedeholm, R. B., Heinemeier, J., Bushnell, P. G., Christiansen, J. S., Olsen, J., Ramsey, C. B., Brill, R. W., Simon, M., Steffensen, K. F., & Steffensen, J. F. (2016). Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (*Somniosus microcephalus*). *Science*, *353*(6300), 702–704. https://doi.org/10.1126/science.aaf1703

O'Brien, C. J., Vogt, M., & Gruber, N. (2016). Global coccolithophore diversity: Drivers and future change. *Progress in Oceanography*, *140*, 27–42. https://doi.org/10.1016/j.pocean.2015.10.003

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2024). *vegan: Community Ecology Package* (Version 2.6-6.1) [Computer software]. https://cran.r-project.org/web/packages/vegan/index.html

Olguín Salinas, H. F., Alder, V. A., Puig, A., & Boltovskoy, D. (2015). Latitudinal diversity patterns of diatoms in the Southwestern Atlantic and Antarctic waters. *Journal of Plankton Research*, *37*(4), 659–665. https://doi.org/10.1093/plankt/fbv042

Panzeri, D., Russo, T., Arneri, E., Carlucci, R., Cossarini, G., Isajlović, I., Krstulović Šifner, S., Manfredi, C., Masnadi, F., Reale, M., Scarcella, G., Solidoro, C., Spedicato, M. T., Vrgoč, N., Zupa, W., & Libralato, S. (2024). Identifying priority areas for spatial management of mixed fisheries using ensemble of multi-species distribution models. *Fish and Fisheries*, *25*(2), 187–204. https://doi.org/10.1111/faf.12802

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate Change and Distribution Shifts in Marine Fishes. *Science*, *308*(5730), 1912–1915. https://doi.org/10.1126/science.1111322

Pianka, E. R. (1966). Latitudinal Gradients in Species Diversity: A Review of Concepts. *The American Naturalist*, *100*(910), 33–46.

Pielou, E. C. (1996). The Measurement of Diversity in Different Types of Biological Colledions. *Journal of Theoretical Biology*, *13*, 131–114.

Pierella Karlusich, J. J., Ibarbalz, F. M., & Bowler, C. (2020). Phytoplankton in the Tara Ocean. *Annual Review of Marine Science*, *12*(1), 233–265. https://doi.org/10.1146/annurev-marine-010419-010706

Preston, F. W. (1948). The Commonness, And Rarity, of Species. *Ecology*, *29*(3), 254–283. https://doi.org/10.2307/1930989

Price, C. A., Weitz, J. S., Savage, V. M., Stegen, J., Clarke, A., Coomes, D. A., Dodds, P. S., Etienne, R. S., Kerkhoff, A. J., McCulloh, K., Niklas, K. J., Olff, H., & Swenson, N. G. (2012). Testing the metabolic theory of ecology. *Ecology Letters*, *15*(12), 1465–1474. https://doi.org/10.1111/j.1461-0248.2012.01860.x

Procheş, Ş., Watkeys, M. K., Ramsay, L. F., & Cowling, R. M. (2023). Why we should be looking for longitudinal patterns in biodiversity. *Frontiers in Ecology and Evolution*, *11*, 1032827. https://doi.org/10.3389/fevo.2023.1032827

Purvis, A., & Hector, A. (2000). *Getting the measure of biodiversity*.

Raffaelli, D., Solan, M., & Webb, T. J. (2005). Do marine and terrestrial ecologists do it differently? *Marine Ecology Progress Series*, *304*, 283–289.

Rapoport, E. (1982). *Areography: Geographical strategies of species*. Pergamon Press. https://doi.org/10.1016/C2013-0-03639-0

Record, S., Charney, N. D., Zakaria, R. M., & Ellison, A. M. (2013). Projecting global mangrove species and community distributions under climate change. *Ecosphere*, *4*(3), art34. https://doi.org/10.1890/ES12-00296.1

Reygondeau, G., Guidi, L., Beaugrand, G., Henson, S. A., Koubbi, P., MacKenzie, B. R., Sutton, T. T., Fioroni, M., & Maury, O. (2018). Global biogeochemical provinces of the mesopelagic zone. *Journal of Biogeography*, *45*(2), 500–514. https://doi.org/10.1111/jbi.13149

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013). Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles*, *27*(4), 1046–1058. https://doi.org/10.1002/gbc.20089

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L., & Melo, S. (2002). Towards a functional classification of the freshwater phytoplankton. *Journal of Plankton Research*, *24*(5), 417–428. https://doi.org/10.1093/plankt/24.5.417

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., Stevens, D., & Witt, M. (2006). Using continuous plankton recorder data. *Progress in Oceanography*, *68*(1), 27–74. https://doi.org/10.1016/j.pocean.2005.09.011

Ricotta, C., & Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity*, *31*, 201–205. https://doi.org/10.1016/j.ecocom.2017.07.003

Rieseberg, L. H. (1997). Hybrid Origins of Plant Species. *Annual Review of Ecology and Systematics*, *28*, 359–389.

Righetti, D., Vogt, M., Gruber, N., Psomas, A., & Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, *5*(5), eaau6253. https://doi.org/10.1126/sciadv.aau6253

Righetti, D., Vogt, M., Gruber, N., & Zimmermann, N. E. (2023). *Mapping global marine biodiversity under sparse data conditions* [Preprint]. Ecology. https://doi.org/10.1101/2023.02.28.530497

Righetti, D., Vogt, M., Zimmermann, N. E., Guiry, M. D., & Gruber, N. (2020). PhytoBase: A global synthesis of open-ocean phytoplankton occurrences. *Earth System Science Data*, *12*(2), 907. Gale Academic OneFile.

Rodríguez-Ramos, T., Dornelas, M., Marañón, E., & Cermeño, P. (2014). Conventional sampling methods severely underestimate phytoplankton species richness. *Journal of Plankton Research*, *36*(2), 334–343. https://doi.org/10.1093/plankt/fbt115

Rodríguez-Ramos, T., Marañón, E., & Cermeño, P. (2015). Marine nano- and microphytoplankton diversity: Redrawing global patterns from sampling-standardized data. *Global Ecology and Biogeography*, *24*(5), 527–538. https://doi.org/10.1111/geb.12274

Rohde, K. (1992). Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. *Oikos*, *65*(3), 514–527. https://doi.org/10.2307/3545569

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press.

RStudio. (2020). *RStudio: Integrated Development Environment for R* [Computer software]. RStudio, PBC. http://www.rstudio.com/

Sanches, P. F., Pellizzari, F., & Horta, P. A. (2016). Multivariate analyses of Antarctic and sub-Antarctic seaweed distribution patterns: An evaluation of the role of the Antarctic Circumpolar Current. *Journal of Sea Research*, *110*, 29–38. https://doi.org/10.1016/j.seares.2016.02.002

Saupe, E. E. (2023). Explanations for latitudinal diversity gradients must invoke rate variation. *Proceedings of the National Academy of Sciences*, *120*(33), e2306220120. https://doi.org/10.1073/pnas.2306220120

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., & Nickless, D. (2004). Terra and Aqua MODIS products available from NASA GES DAAC. *Advances in Space Research*, *34*(4), 710–714. https://doi.org/10.1016/j.asr.2004.03.012

Sellner, K. G., Doucette, G. J., & Kirkpatrick, G. J. (2003). Harmful algal blooms: Causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, *30*(7), 383–406. https://doi.org/10.1007/s10295-003-0074-9

Stephens, T. (2023). The Kunming–Montreal Global Biodiversity Framework. *International Legal Materials*, *62*(5), 868–887. https://doi.org/10.1017/ilm.2023.16

Stirling, G., & Wilsey, B. (2001). Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. *The American Naturalist*, *158*(3), 286–299. https://doi.org/10.1086/321317

Strauss, J., Choi, C. J., Grone, J., Wittmers, F., Jimenez, V., Makareviciute-Fichtner, K., Bachy, C., Jaeger, G. S., Poirier, C., Eckmann, C., Spezzano, R., Löscher, C. R., Sarma, V. V. S. S., Mahadevan, A., & Worden, A. Z. (2023). The Bay of Bengal exposes abundant photosynthetic picoplankton and newfound diversity along salinity-driven gradients. *Environmental Microbiology*, *25*(11), 2118–2141. https://doi.org/10.1111/1462-2920.16431

Sukhanova, I. N., Flint, M. V., Pautova, L. A., Stockwell, D. A., Grebmeier, J. M., & Sergeeva, V. M. (2009). Phytoplankton of the western Arctic in the spring and summer of 2002: Structure and seasonal changes. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(17), 1223–1236. https://doi.org/10.1016/j.dsr2.2008.12.030

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

Thomas, M. K., Kremer, C. T., Klausmeier, C. A., & Litchman, E. (2012). A Global Pattern of Thermal Adaptation in Marine Phytoplankton. *Science*, *338*(6110), 1085–1088. https://doi.org/10.1126/science.1224836

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Timmermans, M.-L., & Marshall, J. (2020). Understanding Arctic Ocean Circulation: A Review of Ocean Dynamics in a Changing Climate. *Journal of Geophysical Research: Oceans*, *125*(4), e2018JC014378. https://doi.org/10.1029/2018JC014378

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010a). Global patterns and predictors of marine biodiversity across taxa. *Nature*, *466*(7310), 1098–1101. https://doi.org/10.1038/nature09329

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010b). Global patterns and predictors of marine biodiversity across taxa. *Nature*, *466*(7310), 1098–1101. https://doi.org/10.1038/nature09329

Tittensor, D. P., Novaglio, C., Harrison, C. S., Heneghan, R. F., Barrier, N., Bianchi, D., Bopp, L., Bryndum-Buchholz, A., Britten, G. L., Büchner, M., Cheung, W. W. L., Christensen, V., Coll, M., Dunne, J. P., Eddy, T. D., Everett, J. D., Fernandes-Salvador, J. A., Fulton, E. A., Galbraith, E. D., … Blanchard, J. L. (2021). Next-generation ensemble projections reveal higher climate risks for marine ecosystems. *Nature Climate Change*, *11*(11), 973–981. https://doi.org/10.1038/s41558-021-01173-9

Tittensor, D. P., & Worm, B. (2016). A neutral-metabolic theory of latitudinal biodiversity. *Global Ecology and Biogeography*, *25*(6), 630–641. https://doi.org/10.1111/geb.12451

Trew, B. T., & Maclean, I. M. D. (2021). Vulnerability of global biodiversity hotspots to climate change. *Global Ecology and Biogeography*, *30*(4), 768–783. https://doi.org/10.1111/geb.13272

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), 9132. https://doi.org/10.1038/s41598-017-09084-6

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*(2), 272–281. https://doi.org/10.1111/j.1466-8238.2011.00656.x

Van Daele, F., Honnay, O., & De Kort, H. (2021). The role of dispersal limitation and reforestation in shaping the distributional shift of a forest herb under climate change. *Diversity and Distributions*, *27*(9), 1775–1791. https://doi.org/10.1111/ddi.13367

Van Dam, A. (2019). Diversity and its decomposition into variety, balance and disparity. *Royal Society Open Science*, *6*(7), 190452. https://doi.org/10.1098/rsos.190452

Velasco, J. A., & Pinto-Ledezma, J. N. (2022). Mapping species diversification metrics in macroecology: Prospects and challenges. *Frontiers in Ecology and Evolution*, *10*, 951271. https://doi.org/10.3389/fevo.2022.951271

Villamaña, M., Marañón, E., Cermeño, P., Estrada, M., Fernández-Castro, B., Figueiras, F. G., Latasa, M., Otero-Ferrer, J. L., Reguera, B., & Mouriño-Carballido, B. (2019). The role of mixing in controlling resource availability and phytoplankton community composition. *Progress in Oceanography*, *178*, 102181. https://doi.org/10.1016/j.pocean.2019.102181

Visintini, N., Martiny, A. C., & Flombaum, P. (2021). Prochlorococcus, Synechococcus, and picoeukaryotic phytoplankton abundances in the global ocean. *Limnology and Oceanography Letters*, *6*(4), 207–215. https://doi.org/10.1002/lol2.10188

Visser, V., Clayton, W. D., Simpson, D. A., Freckleton, R. P., & Osborne, C. P. (2014). Mechanisms driving an unusual latitudinal diversity gradient for grasses. *Global Ecology and Biogeography*, *23*(1), 61–75. https://doi.org/10.1111/geb.12107

Wadgymar, S. M., DeMarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence. *Annual*

Review of Ecology, Evolution, and Systematics, *53*(1), 87–111. https://doi.org/10.1146/annurev-ecolsys-012722-035231

Wang, J., Meier, S., Soininen, J., Casamayor, E. O., Pan, F., Tang, X., Yang, X., Zhang, Y., Wu, Q., Zhou, J., & Shen, J. (2017). Regional and global elevational patterns of microbial species richness and evenness. *Ecography*, *40*(3), 393–402. https://doi.org/10.1111/ecog.02216

Wang, Y.-R., Hessen, D. O., Samset, B. H., & Stordal, F. (2022). Evaluating global and regional land warming trends in the past decades with both MODIS and ERA5-Land land surface temperature data. *Remote Sensing of Environment*, *280*, 113181. https://doi.org/10.1016/j.rse.2022.113181

Ward, B. A., Cael, B. B., Collins, S., & Young, C. R. (2021). Selective constraints on global plankton dispersal. *Proceedings of the National Academy of Sciences*, *118*(10), e2007388118. https://doi.org/10.1073/pnas.2007388118

Webb, T. J. (2012). Marine and terrestrial ecology: Unifying concepts, revealing differences. *Trends in Ecology & Evolution*, *27*(10), 535–541. https://doi.org/10.1016/j.tree.2012.06.002

Whittaker. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338. https://doi.org/10.2307/1943563

Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, *21*(2/3), 213–251. https://doi.org/10.2307/1218190

Whittaker, R. J., Willis, K. J., & Field, R. (2001). Scale and species richness: Towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, *28*(4), 453–470. https://doi.org/https://doi.org/10.1046/j.1365-2699.2001.00563.x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilson, L. J., Fulton, C. J., Hogg, A. M., Joyce, K. E., Radford, B. T. M., & Fraser, C. I. (2016). Climate-driven changes to ocean circulation and their inferred impacts on marine dispersal patterns. *Global Ecology and Biogeography*, *25*(8), 923–939. https://doi.org/10.1111/geb.12456

Woerheide, W., & Persson, D. (1992). An Index of Portfolio Diversification. *Financial Services Review*, *2*(2), 73–85. https://doi.org/10.1016/1057-0810(92)90003-U

World Meteorological Organization. (2017). *WMO Guidelines on the Calculation of Climate Normals* (WMO-No. 1203). https://library.wmo.int/viewer/55797/download?file=1203_en.pdf&type=pdf&navigator=1

Worm, B., & Tittensor, D. P. (2018). *A Theory of Global Biodiversity (MPB-60)*. Princeton University Press.

Wright, D. H. (1983). Species-Energy Theory: An Extension of Species-Area Theory. *Oikos*, *41*(3), 496–506. https://doi.org/10.2307/3544109

Xu, Z., Cheung, S., Endo, H., Xia, X., Wu, W., Chen, B., Ho, N. H. E., Suzuki, K., Li, M., & Liu, H. (2022). Disentangling the Ecological Processes Shaping the Latitudinal Pattern of Phytoplankton Communities in the Pacific Ocean. *mSystems*, *7*(1), e01203-21. https://doi.org/10.1128/msystems.01203-21

Yasuhara, M., Wei, C.-L., Kucera, M., Costello, M. J., Tittensor, D. P., Kiessling, W., Bonebrake, T. C., Tabor, C. R., Feng, R., Baselga, A., Kretschmer, K., Kusumoto, B., & Kubota, Y. (2020). Past and future decline of tropical pelagic biodiversity.

Proceedings of the National Academy of Sciences, *117*(23), 12891–12896. https://doi.org/10.1073/pnas.1916923117

Zacaï, A., Monnet, C., Pohl, A., Beaugrand, G., Mullins, G., Kroeck, D. M., & Servais, T. (2021a). Truncated bimodal latitudinal diversity gradient in early Paleozoic phytoplankton. *Science Advances*, *7*(15), eabd6709. https://doi.org/10.1126/sciadv.abd6709

Zacaï, A., Monnet, C., Pohl, A., Beaugrand, G., Mullins, G., Kroeck, D. M., & Servais, T. (2021b). Truncated bimodal latitudinal diversity gradient in early Paleozoic phytoplankton. *Science Advances*, *7*(15), eabd6709. https://doi.org/10.1126/sciadv.abd6709

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins* (pp. 97–166). Elsevier. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6

7 Supplementary material

7.1 Taxonomic information obtained from AlgaeBase

Tabulated is all genera information obtained from PhytoBase which was then cross-referenced with AlgaeBase.

Table 7 - PhytoBase taxonomic information obtained from AlgaeBase for analysis.

Genus	Family	Order	Class	Phylum	Kingdom	Domain
Carteria	Chlamydomonadaceae	Chlamydomonadales	Chlorophyceae	Chlorophyta	Plantae	Eukaryota
Coccopterum	Pyramimonadaceae	Pyramimonadales	Pyramimonadophyceae	Chlorophyta	Plantae	Eukaryota
Dunaliella	Dunaliellaceae	Chlamydomonadales	Chlorophyceae	Chlorophyta	Plantae	Eukaryota
Halosphaera	Pyramimonadaceae	Pyramimonadales	Pyramimonadophyceae	Chlorophyta	Plantae	Eukaryota
Pachysphaera	Halosphaeraceae	Pyramimonadales	Pyramimonadophyceae	Chlorophyta	Plantae	Eukaryota
Polyasterias	Prorodontidae	Prorodontida	Prostomatea	Ciliophora	Chromista	Eukaryota
Poropila	NA	NA	NA	NA	NA	NA
Pterosperma	Pterospermataceae	Pyramimonadales	Pyramimonadophyceae	Chlorophyta	Plantae	Eukaryota
Pyramimonas	Pyramimonadaceae	Pyramimonadales	Pyramimonadophyceae	Chlorophyta	Plantae	Eukaryota
Chroomonas	Hemiselmidaceae	Cryptomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Cryptomonas	Cryptomonadaceae	Cryptomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Geminigera	Geminigeraceae	Pyrenomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Hillea	Hilleaceae	Cryptomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Isoselmis	Cryptomonadaceae	Cryptomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Leucocryptos	Katablepharidaceae	Katablepharidales	Katablepharidophyceae	Katablepharidophyta	Chromista	Eukaryota
Plagioselmis	Geminigeraceae	Pyrenomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Teleaulax	Geminigeraceae	Pyrenomonadales	Cryptophyceae	Cryptophyta	Chromista	Eukaryota
Telonema	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA
Richelia	Nostocaceae	Nostocales	Cyanophyceae	Cyanobacteria	Eubacteria	Prokaryota
Trichodesmium	Microcoleaceae	Oscillatoriales	Cyanophyceae	Cyanobacteria	Eubacteria	Prokaryota
Eutreptia	Eutreptiidae	Eutreptiiida	Euglenophyceae	Euglenozoa	Protozoa	Eukaryota

Eutreptiella	Eutreptiidae	Eutreptiiida	Euglenophyceae	Euglenozoa	Protozoa	Eukaryota
Acanthoica	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Algirosphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Alisphaera	NA	NA	NA	Haptophyta	Chromista	Eukaryota
Anacanthoica	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Anthosphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Aspidolithus	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Assipetra	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Braarudosphaera	Braarudosphaeraceae	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calcidiscus	Calcidiscaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calciopappus	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calciosolenia	Calciosoleniaceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calicasphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calyptrolithina	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calyptrolithophora	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Calyptrosphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Canistrolithus	Alisphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Ceratolithus	Ceratolithaceae	NA	NA	Haptophyta	Chromista	Eukaryota
Chrysochromulina	Chrysochromulinaceae	Prymnesiales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Coccolithus	Coccolithaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Corisphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Coronosphaera	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Corymbellus	Prymnesiaceae	Prymnesiales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Cyrtosphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Dicrateria	Isochrysidaceae	Isochrysidales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Discosphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Emiliania	Noelaerhabdaceae	Isochrysidales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Florisphaera	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Gephyrocapsa	Noelaerhabdaceae	Isochrysidales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Gladiolithus	Ceratolithaceae	NA	NA	Haptophyta	Chromista	Eukaryota
Glaukolithus	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Haptolina	Prymnesiaceae	Prymnesiales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Hayaster	Calcidiscaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Helicosphaera	Helicosphaeraceae	Zygodiscales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Helladosphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Holococcolithophora	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Homozygosphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Hymenomonas	Hymenomonadaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Isthmolithus	Zygodiscaceae	Zygodiscales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Lohmannosphaera	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Michaelsarsia	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Micrantholithus	Braarudosphaeraceae	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota

Neosphaera	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Oolithotus	Calcidiscaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Ophiaster	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Palusphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Pappomonas	Papposphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Papposphaera	Papposphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Periphyllophora	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Phaeocystis	Phaeocystaceae	Phaeocystales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Picarola	Papposphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Polycrater	Alisphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Pontosphaera	Pontosphaeraceae	Zygodiscales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Poricalyptra	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Poritectolithus	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Reticulofenestra	Noelaerhabdaceae	Isochrysidales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Rhabdolithes	NA	NA	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Rhabdosphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Scyphosphaera	Pontosphaeraceae	Zygodiscales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Solisphaera	Rhabdosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Sphaerocalyptra	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Syracolithus	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Syracosphaera	Syracosphaeraceae	Syracosphaerales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Tetralithoides	NA	NA	NA	Haptophyta	Chromista	Eukaryota
Turrilithus	NA	NA	NA	Haptophyta	Chromista	Eukaryota
Umbellosphaera	Umbellosphaeraceae	NA	NA	Haptophyta	Chromista	Eukaryota
Umbilicosphaera	Calcidiscaceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Zygosphaera	Calyptrosphaeraceae	Coccolithales	Coccolithophyceae	Haptophyta	Chromista	Eukaryota
Acanthodinium	Cladopyxidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Achradina	Amphitholaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Actiniscus	Actiniscaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Adenoides	Peridiniales	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Adinimonas	Desmocapsaceae	Desmocapsales	Dinophyceae	Miozoa	Chromista	Eukaryota
Adnatosphaeridium	Adnatosphaeridiaceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Akashiwo	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Alexandrium	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Alisocysta	NA	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Amphidinium	Amphidiniaceae	Amphidiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Amphidoma	Amphidomataceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Amphisolenia	Amphisoleniaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Amylax	Lingulodiniaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Archaeperidinium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Asterodinium	Kareniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Aureodinium	Suessiaceae	Suessiales	Dinophyceae	Miozoa	Chromista	Eukaryota

Azadinium	Amphidomataceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Barrufeta	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Bitectatodinium	Gonyaulacaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Blepharocysta	Podolampadaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Brachidinium	Brachidiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Cachonina	Peridiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Centrodinium	Ostreopsidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ceratium	Ceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ceratocorys	Protoceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ceratoperidinium	Ceratoperidiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Citharistes	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Cladopyxis	Cladopyxidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Cochlodinium	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Coolia	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Corythodinium	Oxytoxaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Cucumeridinium	NA	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Dinophysis	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Dinothrix	Kryptoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Diplopelta	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Diplopsalis	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Diplopsalopsis	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Dissodinium	NA	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Dissodium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Dolichodinium	NA	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Durinskia	Kryptoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Echinidinium	NA	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Ensiculifera	Ensiculiferaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Fragilidium	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Gambierdiscus	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Goniodoma	Thoracosphaeraceae	Thoracosphaerales	Dinophyceae	Miozoa	Chromista	Eukaryota
Gonyaulax	Gonyaulacaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Gymnodinium	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Gyrodinium	Gyrodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Hemidinium	Hemidiniaceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Heterocapsa	Heterocapsaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Heterodinium	Hemidiniaceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Hexagonifera	Peridiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Histioneis	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Impagidinium	Gonyaulacaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Islandinium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Kapelodinium	Kapelodiniaceae	Torodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Karenia	Kareniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota

Karlodinium	Kareniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Katodinium	Tovelliaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Kofoidinium	Kofoidiniaceae	Noctilucales	Noctilucophyceae	Miozoa	Chromista	Eukaryota
Kryptoperidinium	Kryptoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Lebessphaera	Thoracosphaeraceae	Thoracosphaerales	Dinophyceae	Miozoa	Chromista	Eukaryota
Lebouridinium	NA	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Leonella	Thoracosphaeraceae	Thoracosphaerales	Dinophyceae	Miozoa	Chromista	Eukaryota
Leptodinium	NA	NA	NA	NA	NA	NA
Lingulodinium	Lingulodiniaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Margalefidinium	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Melitasphaeridium	Gonyaulacaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Mesoporos	Prorocentraceae	Prorocentrales	Dinophyceae	Miozoa	Chromista	Eukaryota
Metaphalacroma	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Micracanthodinium	Cladopyxidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Nematodinium	Warnowiaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Neoceratium	Ceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Nusuttodinium	Gymnodiniaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Oblea	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ornithocercus	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ostreopsis	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Oxyrrhis	Oxyrrhinaceae	Oxyrrhinales	NA	Miozoa	Chromista	Eukaryota
Oxytoxum	Oxytoxaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Pachydinium	Ostreopsidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Palaeophalacroma	Cladopyxidaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Parahistioneis	Dinophysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Pentapharsodinium	Ensiculiferaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Peridiniella	NA	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Peridinium	Peridiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Phalacroma	Oxyphysaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Podolampas	Podolampadaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Polykrikos	Polykrikaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Preperidinium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Pronoctiluca	Protodiniferaceae	Noctilucales	Noctilucophyceae	Miozoa	Chromista	Eukaryota
Prorocentrum	Prorocentraceae	Prorocentrales	Dinophyceae	Miozoa	Chromista	Eukaryota
Protoceratium	Protoceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Protoperidinium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Protopsis	Warnowiaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Ptychodiscus	Ptychodiscaceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Pyrocystis	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Pyrodinium	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Pyrophacus	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Quinquecuspis	Peridiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota

Scaphodinium	Leptodiscaceae	Noctilucales	Noctilucophyceae	Miozoa	Chromista	Eukaryota
Schuettiella	Protoceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Sclerodinium	Ptychodiscaceae	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Scrippsiella	Thoracosphaeraceae	Thoracosphaerales	Dinophyceae	Miozoa	Chromista	Eukaryota
Sinophysis	NA	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Spatulodinium	Noctilucaceae	Noctilucales	Noctilucophyceae	Miozoa	Chromista	Eukaryota
Spiniferites	Protodiniferaceae	Noctilucales	Noctilucophyceae	Miozoa	Chromista	Eukaryota
Spiraulax	Gonyaulacaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Svalbardella	Deflandreaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Thoracosphaera	Thoracosphaeraceae	Thoracosphaerales	Dinophyceae	Miozoa	Chromista	Eukaryota
Torodinium	Torodiniaceae	Torodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Triadinium	Pyrocystaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Trinovantidinium	NA	NA	Dinophyceae	Miozoa	Chromista	Eukaryota
Tripos	Ceratiaceae	Gonyaulacales	Dinophyceae	Miozoa	Chromista	Eukaryota
Triposolenia	Amphisoleniaceae	Dinophysales	Dinophyceae	Miozoa	Chromista	Eukaryota
Warnowia	Warnowiaceae	Gymnodiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Zygabikodinium	Protoperidiniaceae	Peridiniales	Dinophyceae	Miozoa	Chromista	Eukaryota
Achnanthes	Achnanthaceae	Mastogloiales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Actinocyclus	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Actinoptychus	Heliopeltaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Alveus	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Amphicocconeis	Cocconeidaceae	Cocconeidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Amphiprora	Amphipleuraceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Amphora	NA	NA	NA	NA	NA	NA
Ardissonea	NA	NA	NA	NA	NA	NA
Asterionella	Tabellariaceae	Tabellariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Asterionellopsis	Asterionellopsidaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Asterolampra	Asterolampraceae	Asterolamprales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Asteromphalus	Asterolampraceae	Asterolamprales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Asteroplanus	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Attheya	Attheyaceae	Biddulphiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Aulacodiscus	Aulacodiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Aulacoseira	Aulacoseiraceae	Aulacoseirales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Auliscus	Eupodiscaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Azpeitia	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Bacillaria	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Bacteriastrum	Chaetocerotaceae	Chaetocerotales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Bacterosira	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Banquisia	Plagiotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Bellerochea	Bellerocheaceae	Biddulphiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Biddulphia	Biddulphiaceae	Biddulphiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Bleakeleya	Asterionellopsidaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota

Caloneis	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Campylodiscus	Surirellaceae	Surirellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Carinasigma	Pleurosigmataceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Catacombas	Ulnariaceae	Licmophorales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Catenula	Catenulaceae	Thalassiophysales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Cerataulina	Hemiaulaceae	Hemiaulales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Cerataulus	Eupodiscaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Ceratoneis	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Chaetoceros	Chaetocerotaceae	Chaetocerotales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Charcotiella	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Climacodium	Bellerocheaceae	Biddulphiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Climacosphenia	Climacospheniaceae	Toxariales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Cocconeis	Cocconeidaceae	Cocconeidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Corethron	Corethraceae	Corethrales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Coscinodiscopsis	Coscinodiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Coscinodiscus	Coscinodiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Craticula	Stauroneidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Cuneolus	Rhoicospheniaceae	Cymbellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Cyclotella	Stephanodiscaceae	Stephanodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Cylindrotheca	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Cymatonitzschia	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Cymatosira	Cymatosiraceae	Cymatosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Dactyliosolen	Rhizosoleniaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Delphineis	Rhaphoneidaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Denticella	Triceratiaceae	Triceratiales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Detonula	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Diatoma	Tabellariaceae	Tabellariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Dimeregramma	Plagiogrammaceae	Plagiogrammales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Diploneis	Diploneidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Ditylum	Lithodesmiaceae	Lithodesmiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Ellerbeckia	Radialiplicataceae	Paraliales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Entomoneis	Entomoneidaceae	Surirellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Ephemera	Plagiotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Epithemia	Rhopalodiaceae	Rhopalodiales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Ethmodiscus	Ethmodiscaceae	Ethmodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Eucampia	Hemiaulaceae	Hemiaulales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Eunotogramma	Anaulaceae	Anaulales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Euodia	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Eupodiscus	Eupodiscaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Eupyxidicula	Stephanopyxidaceae	Stephanopyxales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Falcula	Ulnariaceae	Licmophorales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Fallacia	Sellaphoraceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota

Fragilaria	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Fragilariopsis	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Glyphodesmis	Plagiogrammaceae	Plagiogrammales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Gossleriella	Gossleriellaceae	Stellarimales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Grammatophora	Grammatophoraceae	Rhabdonematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Guinardia	Rhizosoleniaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Gyrosigma	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Halamphora	Catenulaceae	Thalassiophysales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Haslea	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Helicotheca	Streptothecaceae	Briggerales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Hemiaulus	Hemiaulaceae	Hemiaulales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Hemidiscus	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Isthmia	Isthmiaceae	Hemiaulales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Lauderia	Lauderiaceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Leptocylindrus	Leptocylindraceae	Chaetocerotales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Licmophora	Licmophoraceae	Licmophorales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Lioloma	Thalassionemataceae	Thalassionematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Lithodesmium	Lithodesmiaceae	Lithodesmiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Lyrella	Lyrellaceae	Lyrellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Manguinea	Plagiotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Mastogloia	Mastogloiaceae	Mastogloiales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Melosira	Melosiraceae	Melosirales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Membraneis	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Meuniera	Plagiotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Microtabella	Grammatophoraceae	Rhabdonematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Minidiscus	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Minutocellus	Cymatosiraceae	Cymatosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Nanoneis	NA	NA	NA	Bacillariophyta	Chromista	Eukaryota
Navicula	Naviculaceae	Naviculales	Bacillariophyceae	NA	Chromista	Eukaryota
Neocalyptrella	Rhizosoleniaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Neodelphineis	Rhaphoneidaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Neostreptotheca	Streptothecaceae	Briggerales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Nitzschia	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Odontella	Odontellaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Oestrupia	Pinnulariaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Opephora	Staurosiraceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Orthoneis	NA	NA	NA	Bacillariophyta	Chromista	Eukaryota
Palmerina	Coscinodiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Paralia	Paraliaceae	Paraliales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Pauliella	Anomoeoneidaceae	Cymbellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Petrodictyon	Surirellaceae	Surirellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Petroneis	Lyrellaceae	Lyrellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota

Phaeodactylum	Phaeodactylaceae	NA	NA	Bacillariophyta	Chromista	Eukaryota
Pinnularia	Pinnulariaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Plagiogramma	Plagiogrammaceae	Plagiogrammales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Plagiogrammopsis	Cymatosiraceae	Cymatosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Plagiotropis	Plagiotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Planktoniella	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Pleurosigma	Pleurosigmataceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Pleurosira	Eupodiscaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Podocystis	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Podosira	Hyalodiscaceae	Melosirales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Porosira	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Prestauroneis	Stauroneidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Proboscia	Probosciaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Progonoia	Scoliotropidaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Psammodictyon	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Psammodiscus	Psammodiscaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Pseudo	NA	NA	NA	NA	NA	NA
Pseudogomphonema	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Pseudoguinardia	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Pseudosolenia	Rhizosoleniaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Rhabdonema	Rhabdonemataceae	Rhabdonematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Rhaphoneis	Rhaphoneidaceae	Rhaphoneidales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Rhizosolenia	Rhizosoleniaceae	Rhizosoleniales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Rhopalodia	Rhopalodiaceae	Rhopalodiales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Roperia	Hemidiscaceae	Coscinodiscales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Shionodiscus	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Skeletonema	Skeletonemataceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Stellarima	Stellarimaceae	Stellarimales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Stephanopyxis	Stephanopyxidaceae	Stephanopyxales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Streptotheca	Streptothecaceae	Briggerales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Striatella	Striatellaceae	Striatellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Surirella	Surirellaceae	Surirellales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Synedra	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Synedropsis	Fragilariaceae	Fragilariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Tabularia	Ulnariaceae	Licmophorales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Tetramphora	Mastogloiaceae	Mastogloiales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Thalassionema	Thalassionemataceae	Thalassionematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Thalassiosira	Thalassiosiraceae	Thalassiosirales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Thalassiothrix	Thalassionemataceae	Thalassionematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Toxarium	Toxariaceae	Toxariales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Toxonidea	Pleurosigmataceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Trachyneis	Naviculaceae	Naviculales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota

Triceratium	Triceratiaceae	Triceratiales	Coscinodiscophyceae	Bacillariophyta	Chromista	Eukaryota
Trichotoxon	Thalassionemataceae	Thalassionematales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Trieres	Parodontellaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Tropidoneis	Lithodesmiaceae	Lithodesmiales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Tryblionella	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota
Zygoceros	Eupodiscaceae	Eupodiscales	Mediophyceae	Bacillariophyta	Chromista	Eukaryota
Calycomonas	NA	NA	Chrysophyceae	Ochrophyta	Chromista	Eukaryota
Dinobryon	Dinobryaceae	Chromulinales	Chrysophyceae	Ochrophyta	Chromista	Eukaryota
Tetraparma	Triparmaceae	Parmales	Bolidophyceae	Ochrophyta	Chromista	Eukaryota
Triparma	Triparmaceae	Parmales	Bolidophyceae	Ochrophyta	Chromista	Eukaryota
Apedinella	Pedinellaceae	Pedinellales	Dictyochophyceae	Ochrophyta	Chromista	Eukaryota
Dictyocha	Dictyochaceae	Dictyochales	Dictyochophyceae	Ochrophyta	Chromista	Eukaryota
Octactis	Dictyochaceae	Dictyochales	Dictyochophyceae	Ochrophyta	Chromista	Eukaryota
Pelagococcus	Pelagomonadaceae	Pelagomonadales	Pelagophyceae	Ochrophyta	Chromista	Eukaryota
Chattonella	Chattonellaceae	Chattonellales	Raphidophyceae	Ochrophyta	Chromista	Eukaryota
Fibrocapsa	Fibrocapsaceae	Chattonellales	Raphidophyceae	Ochrophyta	Chromista	Eukaryota
Olisthodiscus	NA	NA	Raphidophyceae	Ochrophyta	Chromista	Eukaryota
Pseudonitzschia	Bacillariaceae	Bacillariales	Bacillariophyceae	Bacillariophyta	Chromista	Eukaryota

7.2 Species removed in PhytoBase taxonomic cleaning process

Table 8 - Species removed in taxanomic filtering process.

	Scientific name
1	*Poropila dubia*
2	*Telonema subtile*
3	*Prochlorococcus*
4	*Richelia*
5	*Synechococcus*
6	*Trichodesmium*
7	*Alisphaera ordinata*
8	*Alisphaera pinnigera*
9	*Alisphaera spatula*
10	*Alisphaera unicornis*
11	*Aspidolithus signatus*
12	*Assipetra infracretacea*
13	*Braarudosphaera bigelowii*
14	*Canistrolithus valliformis*
15	*Ceratolithus cristatus*
16	*Ceratolithus tricorniculatus*
17	*Florisphaera profunda*
18	*Gladiolithus flabellatus*

19	*Glaukolithus diplogrammus*
20	*Micrantholithus hoschulzii*
21	*Neosphaera coccolithomorpha*
22	*Pappomonas flabellifera*
23	*Papposphaera borealis*
24	*Papposphaera lepida*
25	*Phaeocystis*
26	*Picarola margalefii*
27	*Polycrater galapagensis*
28	*Rhabdolithes claviger*
29	*Tetralithoides quadrilaminata*
30	*Turrilithus latericioides*
31	*Umbellosphaera irregularis*
32	*Umbellosphaera tenuis*
33	*Adenoides eludens*
34	*Adenoides kofoidii*
35	*Adnatosphaeridium robustum*
36	*Adnatosphaeridium vittatum*
37	*Alisocysta circumtabulata*
38	*Alisocysta ornata*
39	*Amphidoma nucula*
40	*Azadinium caudatum*
41	*Cucumeridinium coeruleum*
42	*Dissodinium gerbaultii*
43	*Dissodinium pseudolunula*
44	*Dolichodinium lineatum*
45	*Echinidinium delicatum*
46	*Echinidinium granulatum*
47	*Hemidinium nasutum*
48	*Heterodinium australiae*
49	*Heterodinium blackmanii*
50	*Heterodinium crassipes*
51	*Heterodinium curvatum*
52	*Heterodinium doma*
53	*Heterodinium fides*
54	*Heterodinium hindmarchii*
55	*Heterodinium leiorhynchum*
56	*Heterodinium mediocre*

57	*Heterodinium mediterraneum*
58	*Heterodinium milneri*
59	*Heterodinium minutum*
60	*Heterodinium rigdeniae*
61	*Heterodinium varicator*
62	*Lebouridinium glaucum*
63	*Leptodinium dispertitum*
64	*Leptodinium victorianum*
65	*Oxyrrhis marina*
66	*Peridiniella catenata*
67	*Peridiniella danica*
68	*Peridiniella globosa*
69	*Peridiniella sphaeroidea*
70	*Protoperidinium marielebouriae*
71	*Ptychodiscus noctiluca*
72	*Sclerodinium calyptroglyphe*
73	*Sinophysis microcephala*
74	*Trinovantidinium applanatum*
75	*Amphora angusta*
76	*Amphora arenaria*
77	*Amphora arenicola*
78	*Amphora crassa*
79	*Amphora grevilleana*
80	*Amphora hyalina*
81	*Amphora laevis*
82	*Amphora lineolata*
83	*Amphora marina*
84	*Amphora proteus*
85	*Ardissonea crystallina*
86	*Cocconeis caribensis*
87	*Nanoneis hasleae*
88	*Navicula abunda*
89	*Navicula besarensis*
90	*Navicula cancellata*
91	*Navicula capitatoradiata*
92	*Navicula cincta*
93	*Navicula criophila*
94	*Navicula cryptocephala*
95	*Navicula directa*
96	*Navicula distans*
97	*Navicula fundata*
98	*Navicula gelida*
99	*Navicula incarum*

100	*Navicula kariana*
101	*Navicula longa*
102	*Navicula menisculus*
103	*Navicula normalis*
104	*Navicula oahuensis*
105	*Navicula pagophila*
106	*Navicula palpebralis*
107	*Navicula pelagica*
108	*Navicula pennata*
109	*Navicula pinnata*
110	*Navicula platyventris*
111	*Navicula radiosa*
112	*Navicula septentrionalis*
113	*Navicula transitans*
114	*Navicula vanhoeffenii*
115	*Orthoneis binotata*
116	*Phaeodactylum tricornutum*
117	*Calycomonas gracilis*
118	*Olisthodiscus luteus*
119	*Picoeukaryotes*

7.3 Species distribution models

7.3.1 Supplementary SDM figures

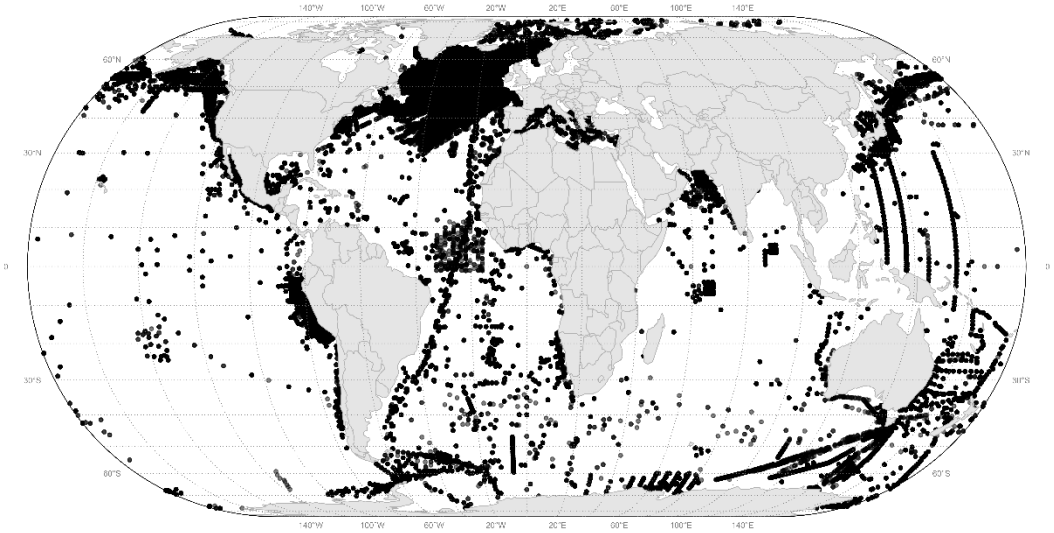


Figure 7.1 - Observational sampling distribution of all modelled species in species distribution modelling method.

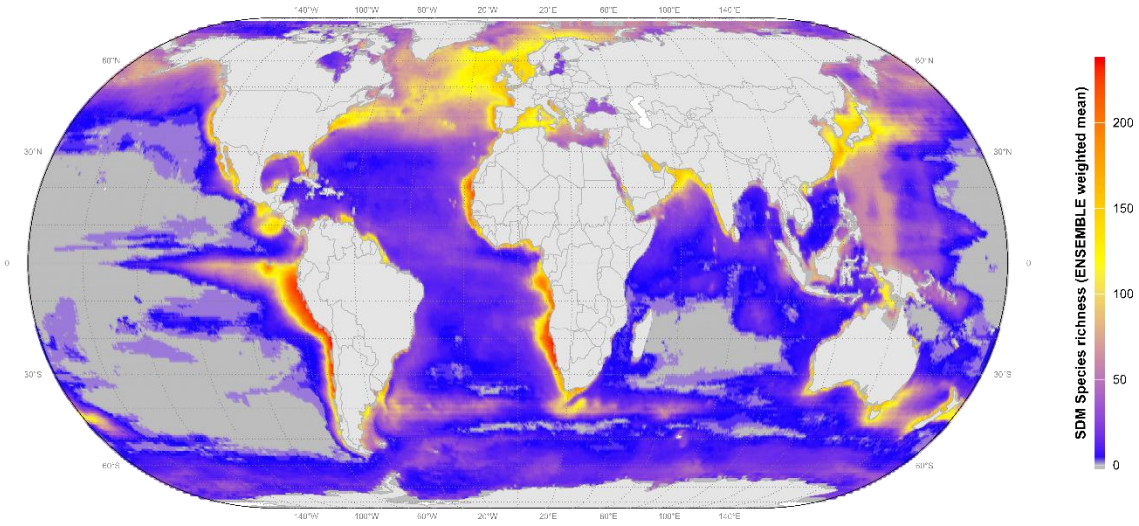


Figure 7.2 – Modelled species richness with greyscale added for locations with no species projected to occur.

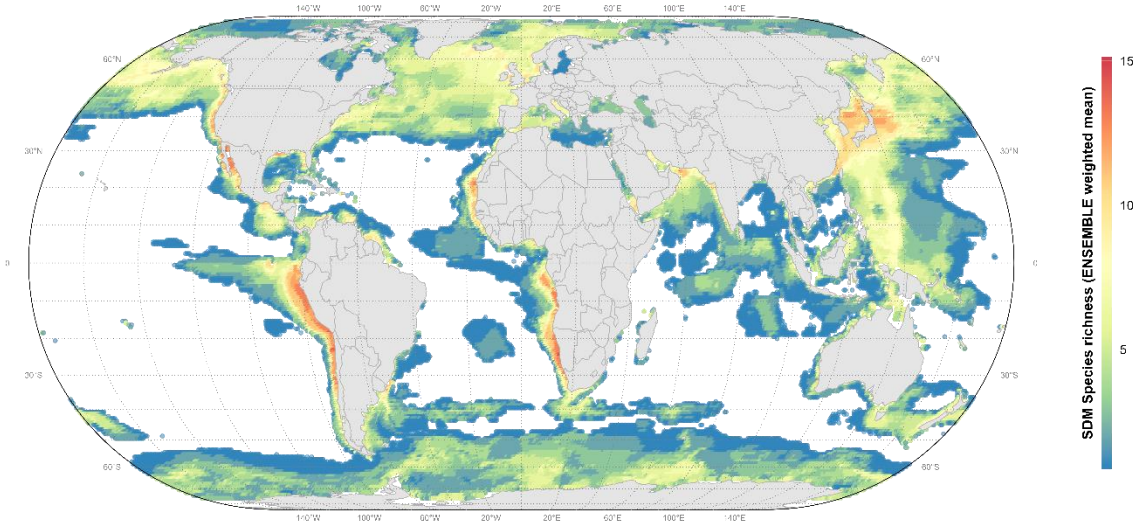


Figure 7.3 – Modelled ensemble projection of all diatoms as modelled species richness within the species distribution modelling framework.

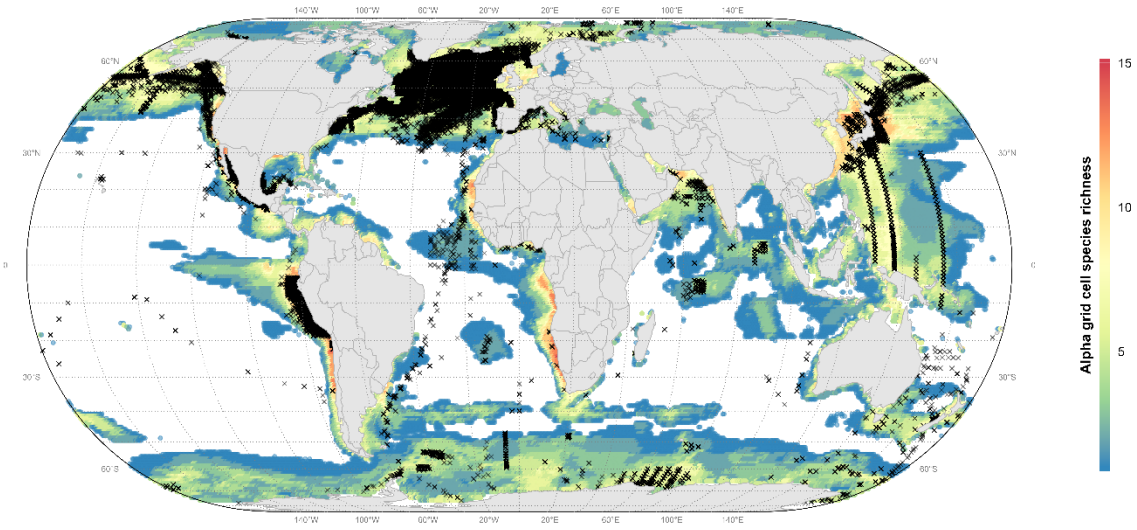


Figure 7.4 - Modelled ensemble projection of all diatoms and their locations of observations (black crossmarks).

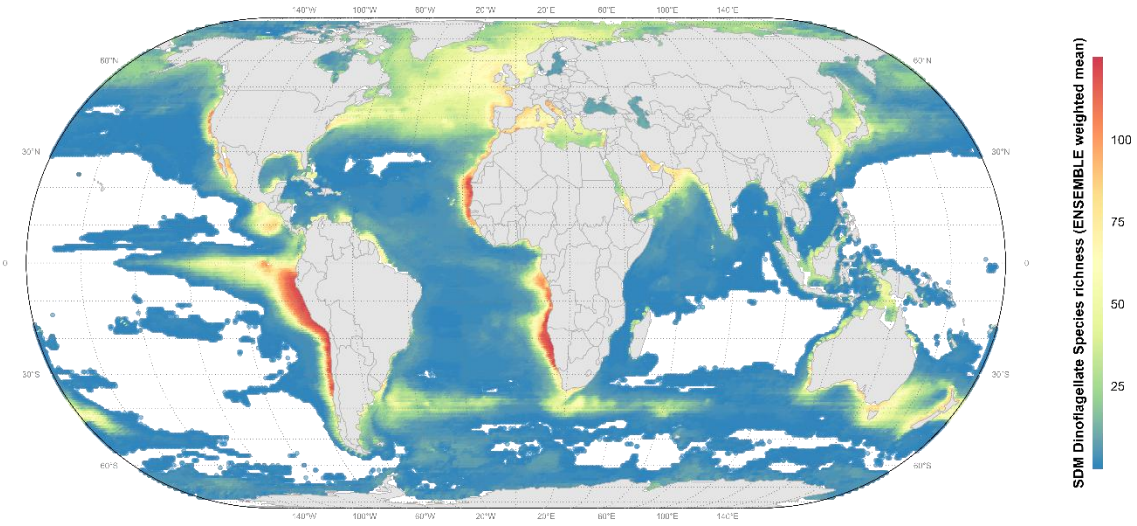


Figure 7.5 – Modelled ensemble projection of all dinoflagellates as modelled species richness within the species distribution modelling framework.

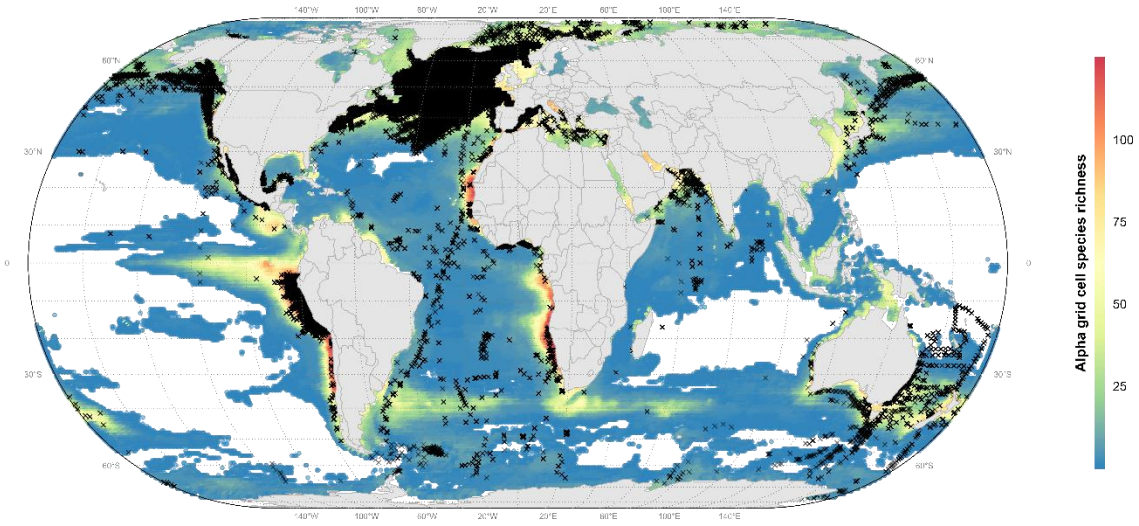


Figure 7.6 - Modelled ensemble projection of all dinoflagellates and their locations of observations (black crossmarks).

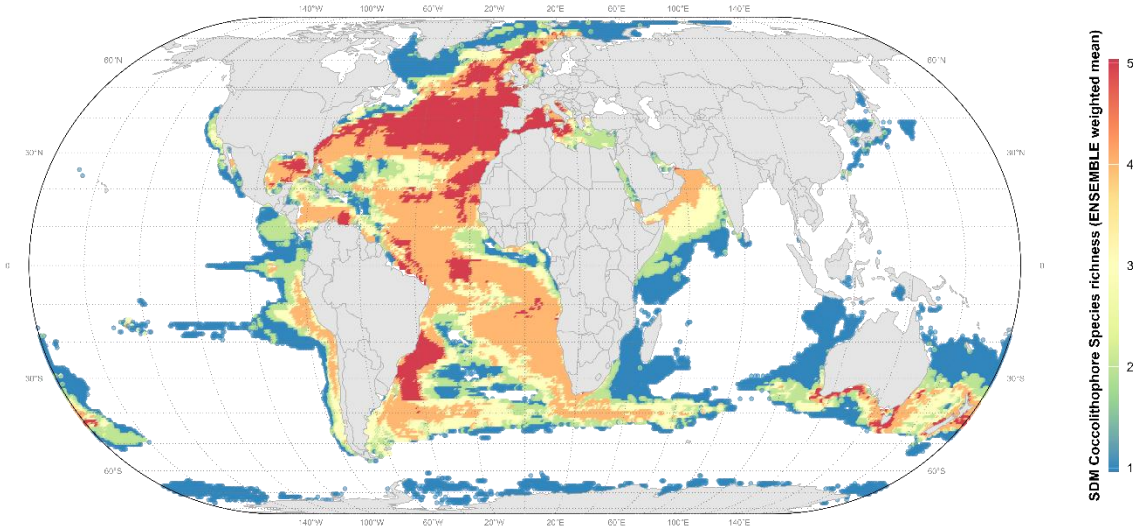


Figure 7.7 - Modelled ensemble projection of all coccolithophores as modelled species richness within the species distribution modelling framework.

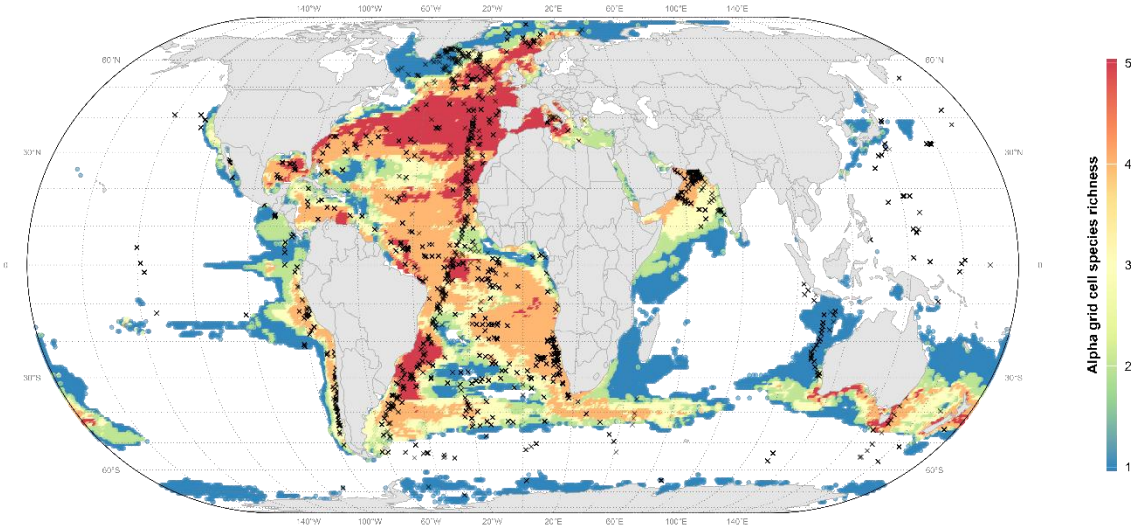


Figure 7.8 - Modelled ensemble projection of all coccolithophores and their locations of observations (black crossmarks).

7.3.2 Species modelled using species distribution modelling

Table 9 - Modelled species in SDM framework

AphiaID	AcceptedSciName
232125	Amphisolenia bispinosa
624607	Archaeperidinium minutum
248070	Asterolampra marylandica
418467	Asteromphalus brookei
957381	Asteromphalus cleveanus
162255	Asteromphalus flabellatus
162258	Asteromphalus heptactis
162259	Asteromphalus hookeri
254442	Asteromphalus hyalinus
577365	Asteromphalus robustus
248066	Bacteriastrum comosum
418486	Bacteriastrum elegans
162916	Bacteriastrum elongatum
149119	Bacteriastrum hyalinum
418487	Bacteriastrum mediterraneum
623773	Bitectatodinium spongium
235923	Calcidiscus leptoporus
555889	Calciosolenia brasiliensis
235926	Calciosolenia murrayi
156509	Ceratium arcticum
109984	Ceratocorys armata
109986	Ceratocorys horrida
233340	Ceratocorys reticulata
163932	Ceratoneis closterium
254448	Chaetoceros aequatorialis
149241	Chaetoceros affinis

149288	Chaetoceros atlanticus
149124	Chaetoceros borealis
149291	Chaetoceros brevis
341472	Chaetoceros bulbosus
178180	Chaetoceros coarctatus
149129	Chaetoceros compressus
156607	Chaetoceros concavicornis
156609	Chaetoceros constrictus
156611	Chaetoceros convolutus
149289	Chaetoceros costatus
465389	Chaetoceros criophilus
149221	Chaetoceros curvisetus
465390	Chaetoceros dadayi
149120	Chaetoceros danicus
149219	Chaetoceros debilis
149126	Chaetoceros decipiens
149121	Chaetoceros densus
178215	Chaetoceros dichaeta
149122	Chaetoceros didymus
157431	Chaetoceros diversus
149624	Chaetoceros furcellatus
640251	Chaetoceros indicus
160523	Chaetoceros lauderi
156617	Chaetoceros lorenzianus
341479	Chaetoceros neglectus
593877	Chaetoceros pacificus
611570	Chaetoceros paradoxus
248065	Chaetoceros pendulus
178185	Chaetoceros peruvianus
163055	Chaetoceros protuberans
163112	Chaetoceros radicans
178201	Chaetoceros rostratus
577634	Chaetoceros saltans
149294	Chaetoceros simplex
149123	Chaetoceros socialis
418516	Chaetoceros vanheurckii
160524	Chaetoceros wighamii
196804	Climacodium frauenfeldianum
179596	Corethron hystrix
341496	Corethron pennatum
235944	Coronosphaera mediterranea
149274	Coscinodiscus asteromphalus
149159	Coscinodiscus centralis
148992	Coscinodiscus concinnus
149266	Coscinodiscus curvatulus
149271	Coscinodiscus granii

156629	Coscinodiscus marginatus
149158	Coscinodiscus radiatus
156632	Coscinodiscus wailesii
179785	Dactyliosolen antarcticus
573457	Diatoma rhombica
157463	Dictyocha fibula
109603	Dinophysis acuminata
109604	Dinophysis acuta
109609	Dinophysis argus
109612	Dinophysis caudata
109627	Dinophysis hastata
109637	Dinophysis norvegica
109662	Dinophysis tripos
110131	Diplopelta asymmetrica
233760	Diplopelta steinii
110001	Diplopsalis lenticula
149023	Ditylum brightwellii
149131	Eucampia zodiacus
616809	Eucampia zoodiacus
110652	Eutreptiella gymnastica
341554	Fragilariopsis curta
172418	Fragilariopsis cylindrus
366281	Fragilariopsis doliolus
341557	Fragilariopsis obliquecostata
235953	Gephyrocapsa caribbeanica
235954	Gephyrocapsa ericsonii
341588	Gephyrocapsa muellerae
233522	Gonyaulax elongata
110033	Gonyaulax pacifica
110035	Gonyaulax polygramma
110041	Gonyaulax spinifera
149338	Grammatophora marina
109811	Gymnodinium marinum
576265	Gymnodinium wulffii
109856	Gyrodinium fusiforme
134546	Halosphaera viridis
248063	Haslea wawrikae
615435	Helicosphaera carteri
632650	Hemiaulus chinensis
163249	Hemiaulus hauckii
418540	Hemiaulus membranaceus
180367	Hemidiscus cuneiformis
110152	Heterocapsa rotundata
623772	Impagidinium patulum
1368615	Kapelodinium vestifici
109920	Kofoidinium velleloides

149135	Lauderia annulata
149106	Leptocylindrus danicus
149039	Leptocylindrus minimus
157062	Licmophora abbreviata
233592	Lingulodinium polyedra
292728	Lioloma delicatulum
418646	Lioloma pacificum
418547	Melosira moniliformis
232516	Mesoporos perforatus
235961	Michaelsarsia elegans
149150	Nitzschia longissima
418208	Nitzschia pacifica
1310442	Octactis speculum
235962	Oolithotus fragilis
115099	Ophiaster hydroideus
109693	Ornithocercus magnificus
109694	Ornithocercus quadratus
109696	Ornithocercus steinii
109697	Ornithocercus thumii
110097	Oxytoxum laticeps
110099	Oxytoxum longiceps
110109	Oxytoxum parvum
110113	Oxytoxum reticulatum
110115	Oxytoxum scolopax
233868	Oxytoxum variabile
149055	Paralia sulcata
109925	Pentapharsodinium dalei
341585	Phaeocystis antarctica
115106	Phaeocystis pouchetii
624358	Phalacroma oxytoxoides
232492	Phalacroma rapa
156505	Phalacroma rotundatum
149517	Plagiotropis lepidoptera
248088	Pleurosigma directum
110199	Podolampas bipes
232625	Podolampas palmipes
110202	Podolampas spinifera
109901	Polykrikos schwartzii
156689	Porosira glacialis
614618	Preperidinium meunieri
149168	Proboscia alata
341501	Proboscia inermis
109903	Pronoctiluca pelagica
110293	Prorocentrum balticum
110298	Prorocentrum dentatum
110300	Prorocentrum gracile

110301	Prorocentrum lima
110303	Prorocentrum micans
110316	Prorocentrum triestinum
110321	Protoceratium reticulatum
233521	Protoperidinium americanum
110208	Protoperidinium bipes
110210	Protoperidinium brevipes
110211	Protoperidinium brochii
162749	Protoperidinium cerasus
163862	Protoperidinium claudicans
110212	Protoperidinium conicoides
110214	Protoperidinium crassipes
110217	Protoperidinium depressum
172460	Protoperidinium diabolus
110219	Protoperidinium divergens
110220	Protoperidinium excentricum
418150	Protoperidinium fatulipes
233296	Protoperidinium grahamii
110222	Protoperidinium grande
110223	Protoperidinium granii
110229	Protoperidinium leonis
233198	Protoperidinium longipes
233195	Protoperidinium longispinum
233168	Protoperidinium mendiolae
110238	Protoperidinium oblongum
110239	Protoperidinium obtusum
110240	Protoperidinium oceanicum
110241	Protoperidinium ovatum
110244	Protoperidinium pallidum
233244	Protoperidinium pedunculatum
110245	Protoperidinium pellucidum
110247	Protoperidinium pentagonum
233013	Protoperidinium peruvianum
232989	Protoperidinium pyrum
110250	Protoperidinium quarnerense
110257	Protoperidinium steinii
232882	Protoperidinium tenuissimum
110263	Protoperidinium tristylum
345882	Pterosperma moebii
345881	Pterosperma vanhoeffenii
110327	Pyrocystis elegans
110328	Pyrocystis fusiformis
232598	Pyrophacus horologium
110267	Pyrophacus steinii
232595	Pyrophacus vancampoae
149066	Rhaphoneis amphiceros

196805	Rhizosolenia acuminata
196811	Rhizosolenia bergonii
248062	Rhizosolenia castracanei
341502	Rhizosolenia chunii
418569	Rhizosolenia formosa
149070	Rhizosolenia hebetata
149116	Rhizosolenia imbricata
149629	Rhizosolenia styliformis
248071	Rhizosolenia temperei
149105	Roperia tesselata
233856	Schuettiella mitra
1321853	Scrippsiella acuminata
233786	Scrippsiella regalis
573446	Shionodiscus oestrupii
149074	Skeletonema costatum
110048	Spiraulax kofoidii
236039	Syracosphaera molischii
577881	Syracosphaera nodosa
235978	Syracosphaera prolongata
235979	Syracosphaera pulchra
418647	Thalassionema bacillare
555052	Thalassionema frauenfeldii
149093	Thalassionema nitzschioides
666449	Thalassiosira antarctica
148919	Thalassiosira decipiens
149102	Thalassiosira gravida
149100	Thalassiosira hyalina
593823	Thalassiosira mendiolana
341519	Thalassiosira partheneia
149101	Thalassiosira subtilis
573662	Thalassiothrix heteromorpha
157083	Thalassiothrix longissima
110277	Thoracosphaera heimii
196828	Triadinium polyedricum
839996	Trieres regia
841182	Tripos arietinus
837310	Tripos azoricus
841184	Tripos belone
841190	Tripos bucephalus
841191	Tripos buceros
841193	Tripos candelabrum
841194	Tripos carnegiei
837448	Tripos carriensis
841199	Tripos compressus
841200	Tripos contortus
841202	Tripos declinatus

840628	Tripos dens
841206	Tripos digitatus
841210	Tripos euarcuatus
841211	Tripos eugrammus
837460	Tripos extensus
837221	Tripos falcatiformis
840627	Tripos furca
840626	Tripos fusus
837222	Tripos gibberus
841246	Tripos gravidus
841248	Tripos hexacanthus
841251	Tripos incisus
837455	Tripos inflatus
841256	Tripos karstenii
837456	Tripos kofoidii
837457	Tripos lamellicornis
841258	Tripos limulus
837459	Tripos lineatus
841259	Tripos longipes
837229	Tripos lunula
1391716	Tripos macroceros
841261	Tripos massiliensis
841263	Tripos minutus
495363	Tripos muelleri
841741	Tripos paradoxides
841746	Tripos pentagonus
841751	Tripos pulchellus

7.4 World Ocean Atlas Environmental Data

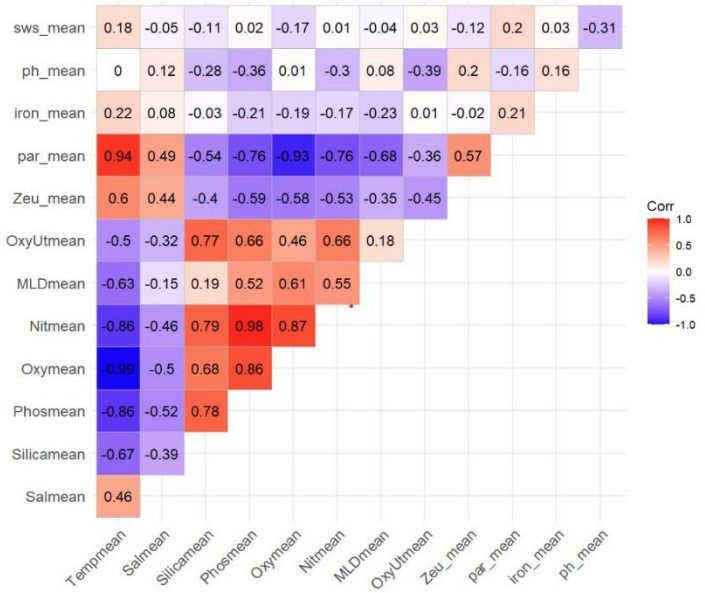


Figure 7.9 – SDM environmental predictor pearson correlation plot showing pairwise correlations among environmental predictors. Positive correlations are indicated by red cells, negative correlations by blue cells, with color intensity reflecting the strength of the correlation (ranging from –1 to +1).

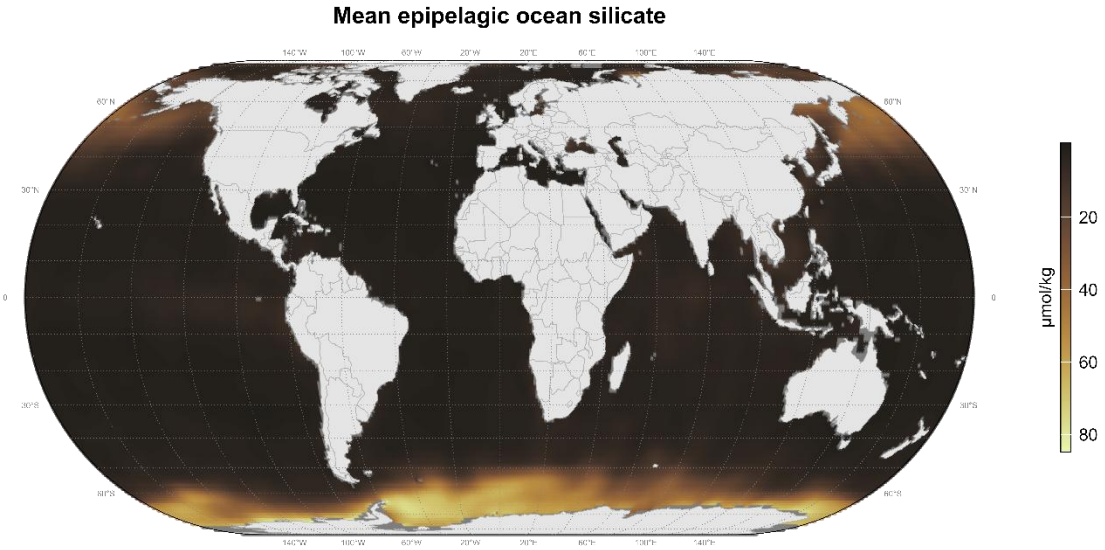


Figure 7.10 – Mean epipelagic ocean silicate.

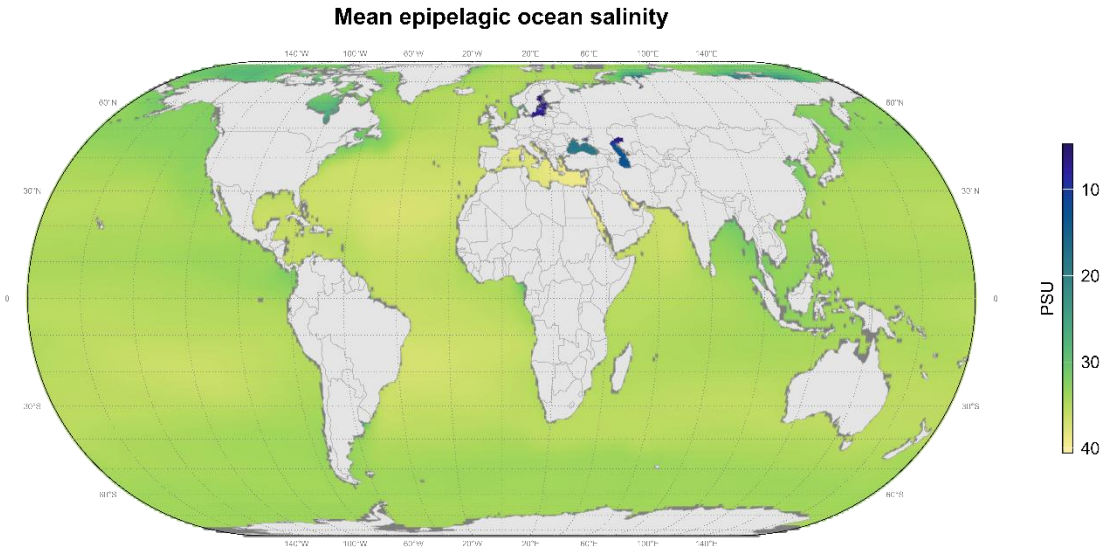


Figure 7.11 – Mean epipelagic ocean salinity.

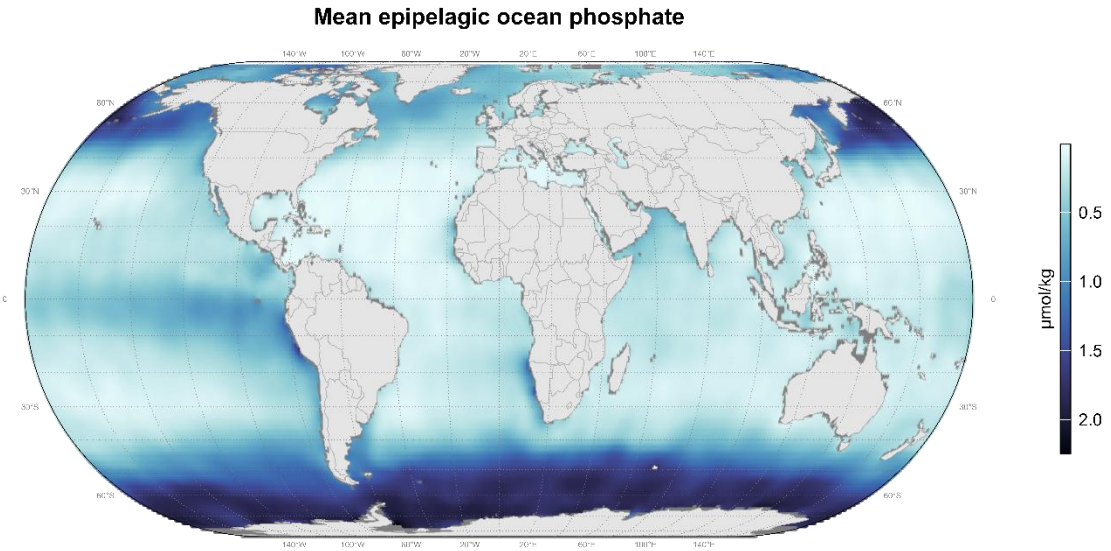


Figure 7.12 – Mean epipelagic ocean phosphate.

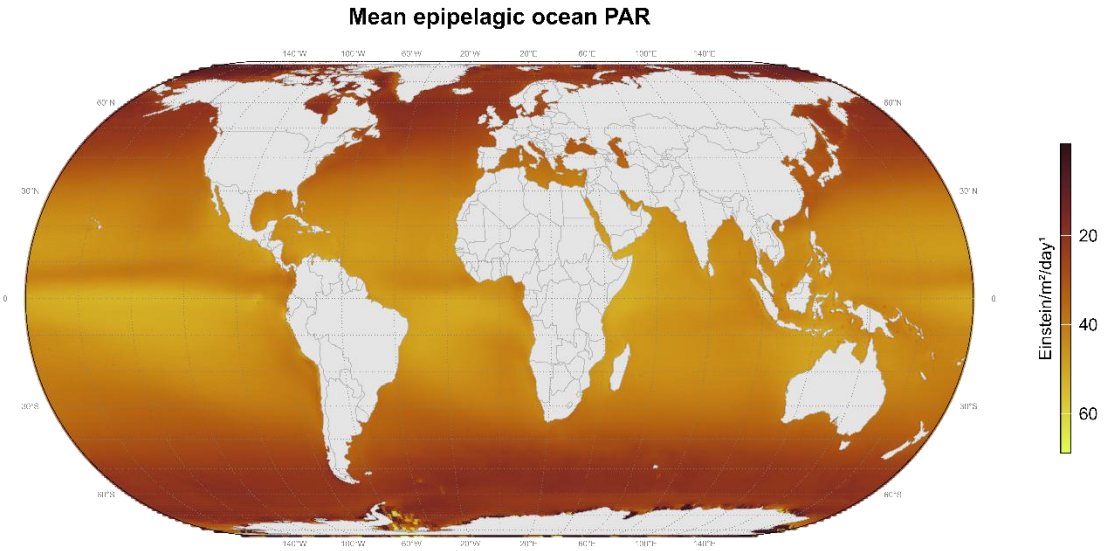


Figure 7.13 - Mean epipelagic ocean photosynthetically active radiation (PAR)

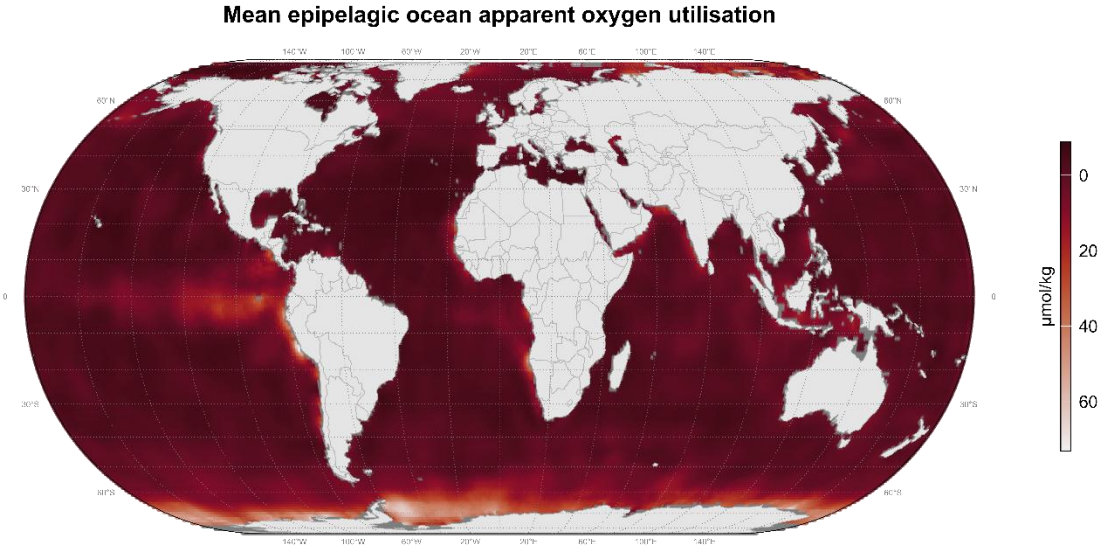


Figure 7.14 - Mean epipelagic ocean apparent oxygen utilisation.

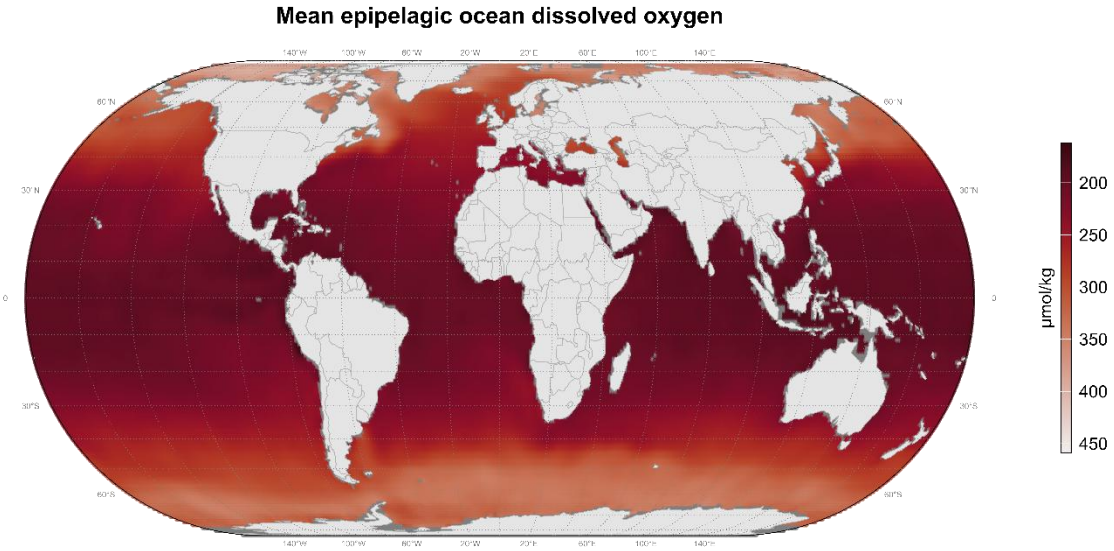


Figure 7.15 - Mean epipelagic ocean dissolved oxygen.

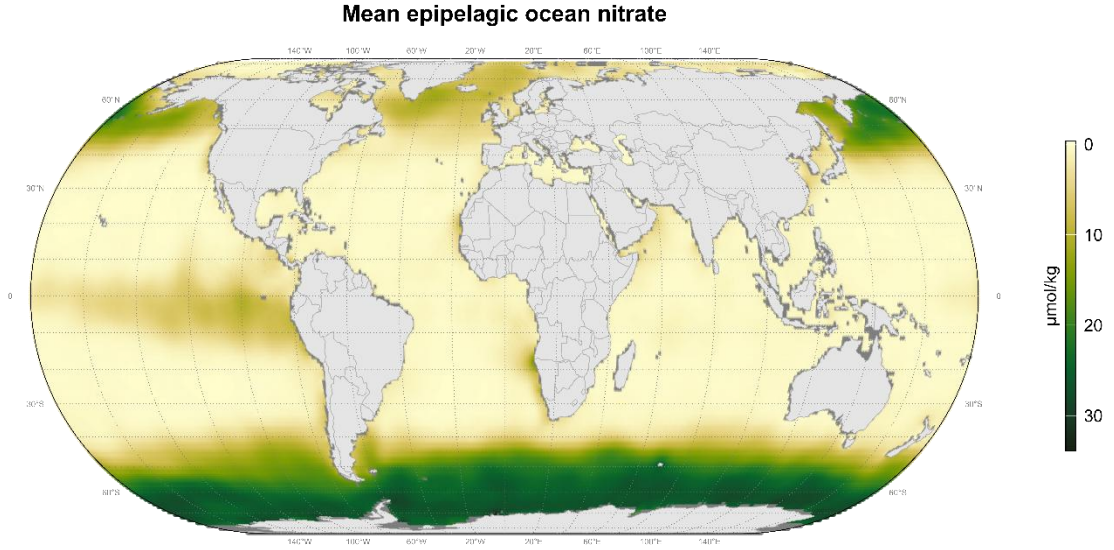


Figure 7.16 - Mean epipelagic ocean nitrate.

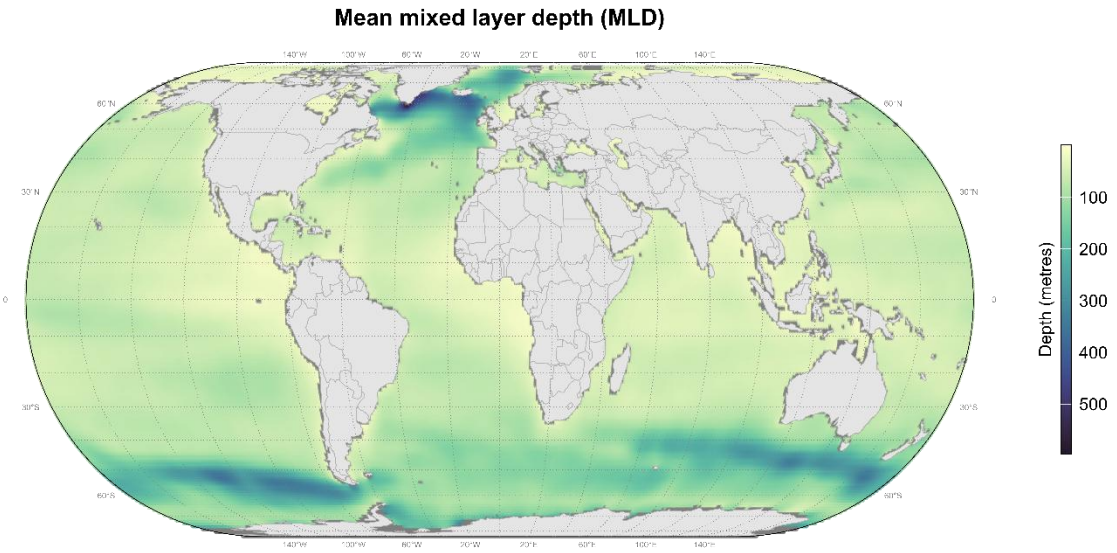


Figure 7.17 - Mean epipelagic ocean mixed layer depth (MLD)

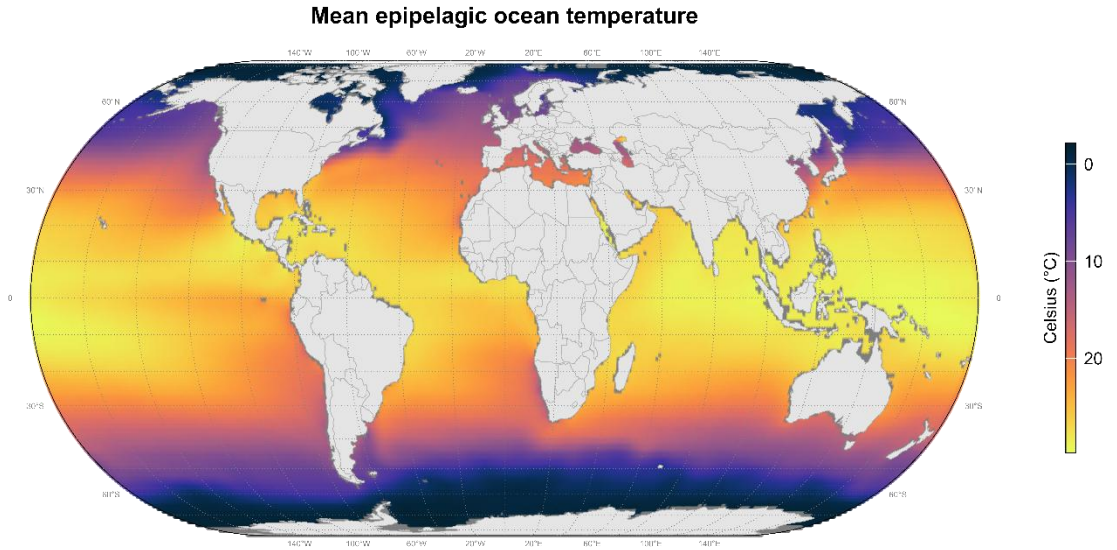


Figure 7.18 - Mean epipelagic ocean temperature.