Biome partitioning of the global ocean based on phytoplankton biogeography

Author(s):
Hofmann Elizondo, Urs; Righetti, Damiano; Benedetti, Fabio; Vogt, Meike

Publication date:
2021-06

Permanent link:
https://doi.org/10.3929/ethz-b-000479179

Originally published in:
Progress in Oceanography 194, https://doi.org/10.1016/j.pocean.2021.102530

Progress in Oceanography 194 (2021) 102530

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Biome partitioning of the global ocean based on phytoplankton biogeography

Urs Hofmann Elizondo *, Damiano Righetti, Fabio Benedetti, Meike Vogt

Environmental Physics Group, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland

ARTICLE INFO

Keywords:
Biomes
Phytoplankton
Self-organizing map
Indicator species
Species network
Species distribution models

ABSTRACT

Biomes are geographical units that can be defined based on biological communities sharing specific environmental and climatic requirements. Contemporary ocean biomes have been constructed based on various approaches. These included the biogeographic patterns of higher trophic level organisms, physical and biogeochemical properties, or bulk biological properties such as chlorophyll-a, but none considered the biogeographic patterns of the first trophic level explicitly, i.e. phytoplankton biogeography. A global description of marine biomes based on phytoplankton and defined in analogy to terrestrial vegetation biomes is still lacking. A bioregionalization based on phytoplankton appears particularly timely, as phytoplankton have a high sensitivity to climatic changes and fuel marine productivity. Here, we partition the global ocean into biomes by using self-organizing maps and hierarchical clustering, drawing on the biogeographic patterns of 536 phytoplankton species predicted from empirical evidence. Our approach reveals eight different biomes at the seasonal scale, and seven at the annual scale. The biomes host characteristic phytoplankton species compositions, and differ in their prevailing environmental conditions. The largest differences in phytoplankton composition are found between a Pacific equatorial biome and other tropical biomes, and between subtropical and high latitude biomes. The Pacific equatorial biome is characterized by species with narrower ecological niches, the tropical and subtropical biomes by cosmopolitan generalists, and the high latitudes by species with a heterogeneous biogeography. The strongest differences between biomes are found along gradients of temperature and macronutrient availability, associated with latitude. We test whether our biomes can be reproduced based on indicator species, or potential co-occurrence networks of species determined from the predicted species distributions that are wide-spread in some but rare in other biomes. We find that our biomes can be reproduced by the 51 species identified, which together form significant species co-occurrences. This suggests that species co-occurrences, rather than individual indicator species drive oceanic biome partitioning at the first trophic level. Our biome partitioning may be especially useful for comparative analyses on the functional implications of phytoplankton organization, and impacts on zoogeographical partitionings. Furthermore, it provides a framework for predicting large-scale changes in phytoplankton community structure due to anthropogenic climate and environmental change.

1. Introduction

Biomes classify global natural biodiversity, including species assemblages and countless ecosystem types into geographic realms with distinct life forms, which provide similar ecosystem services. Classically, terrestrial biomes have been delimited by rather sharp transitions in vegetation type driven by global climatic gradients of temperature and precipitation (Bailey, 1998; Townsend et al., 2008). These biomes have emerged from characteristic life-forms and plant functional traits that represent adaptations to the climatic zones (Ringelberg et al., 2020). Originally focusing on vegetation (e.g. Clements, 1917; Whittaker, 1970), the biome concept was later expanded to include differences in ecosystem services and functions (e.g. Higgins et al., 2016; Silva de Miranda et al., 2018), with results being compared to terrestrial zoogeographic partitionings (Ficetola et al., 2017; Holt et al., 2012; Wallace, 1876). Over the past three decades, biomes have been increasingly delineated in the ocean system, yet the biotic and environmental factors used to define such biomes vary strongly between

* Corresponding author.
E-mail addresses: urs.hofmann@usys.ethz.ch (U. Hofmann Elizondo), damiano.righetti@env.ethz.ch (D. Righetti), fabio.benedetti@usys.ethz.ch (F. Benedetti), meike.vogt@env.ethz.ch (M. Vogt).

https://doi.org/10.1016/j.pocean.2021.102530
Received 31 January 2020; Received in revised form 15 January 2021; Accepted 15 February 2021
Available online 24 March 2021

studies. While some studies relied on environmental and biogeochemical characteristics (Sarmiento et al., 2004; Oliver and Irwin, 2008; Fay and McKinley, 2014; Zhao et al., 2019), others implemented biological criteria. Recent studies have drawn on the taxonomic biogeography of zooplankton and higher trophic level organisms (Reygondeau et al., 2011; Sutton et al., 2017; Costello et al., 2017), both biological and physical factors (Longhurst, 1995; Spalding et al., 2012), or they have contested the applicability of the biome concept to the ocean (van der Spoel, 1994; Ekman, 1953).

Currently, the most widely-used marine classification separates the ocean into 57 provinces using physical properties (temperature, salinity, turbulence), chlorophyll concentration, primary production, and in-situ zooplankton observations (Longhurst, 1995, 2007). In similarity to the terrestrial biomes concept, this partitioning aimed to establish a link between ecosystem structure and function, and strived to characterize the ocean's carbon budget and its seasonality (Longhurst, 1995). The approach by Longhurst (1995) assumed a bottom-up control of abiotic factors over marine ecosystems structure and species composition, in which physical processes shaped plankton communities via nutrient supply, ambient temperature, and light availability. Similar studies have expanded on Longhurst's work and objectively distinguish ecoregions from satellite observables such as sea surface temperature, nutrient concentration, and chlorophyll-a concentration (Vichi et al., 2011; Hardman-Mountford et al., 2008). Nevertheless, the definition of these marine provinces has been debated, given the lack of inclusion of more detailed biogeographic information on taxa, their static nature (e.g. Vichi et al., 2011; Reygondeau et al., 2013), and given their possible lack of significance for biological communities and ecosystem functions (Vichi et al., 2011).

Marine bioregionalizations based on taxonomic biogeography have often been of regional character (Waters et al., 2010; Kulbicki et al., 2013; Griffiths et al., 2009; Briggs and Bowen, 2011), have focused on individual taxa (e.g. Radiolaria Polycistina; Boltovskoy and Correa, 2016), taxa of coastal and shallow seas (Briggs and Bowen, 2011; Spalding et al., 2007), or higher trophic level taxa, including top predators such as tuna and billfish (Reygondeau et al., 2011). Although global biodiversity distributions in the ocean have been recently characterized for phytoplankton (Righetti et al., 2019a) as well as zooplankton and marine mammals (Tittensor et al., 2010), Costello et al. (2017) defined the only generalized, global-scale, multi-taxa marine bioregionalization (Zhao et al., 2019). Costello et al. (2017) partitioned the ocean into 30 marine regions based on the occurrence observations of 65′000 pelagic and benthic species spanning eleven different phyla (of which 96% are heterotrophic, e.g. Arthropoda, Mollusca, Chordata). This partitioning resulted in a set of unique ecological provinces with a high degree of species endemicity (Costello et al., 2017). Similar zoogeographic studies in terrestrial and marine systems have suggested that biomes may be strongly dependent on the trophic level addressed and the climatic or evolutionary drivers of individual taxa, within a hierarchically nested structure (Ficetola et al., 2017; Spalding et al., 2012). Hence, Costello's regionalization of the open ocean may be characteristic for fauna, but does not consider the degree of geographic variability in lower trophic level organisms, such as autotrophs, that inhabit the niche space of the entire surface ocean, with many abundant cosmopolitan species (Brun et al., 2015).

In summary, current ocean partitionings have either been based on environmental, biogeochemical, and bulk biological properties such as chlorophyll-a (e.g. Oliver and Irwin, 2008; Fay and McKinley, 2014; Vichi et al., 2011; Kavanaugh et al., 2014) or the distribution and biogeographic patterns of higher trophic level organisms (Reygondeau et al., 2011; Costello et al., 2017). However, none of them has considered the biogeographic patterns of the ocean's main primary producers, the phytoplankton. Partitionings based on environmental and biogeochemical properties (e.g. Longhurst, 2007) have revealed distinct differences in environmental conditions between biomes, but biomes have tended to overlap in their biological bulk properties (e.g. chlorophyll-a concentration or bacterial and heterotrophic biomass; Vichi et al., 2011). The first zoogeographic partitionings lead to a large number of biologically unique biomes with a high degree of endemicity, yet these biomes are sensitive to specific taxa and trophic levels considered (Costello et al., 2017; Ficetola et al., 2017), and the links between ecosystem structure and ecosystem function have not yet been established.

Owing to the recent increase in global plankton distribution data based on satellites (e.g. Moisan et al., 2017; Navarro et al., 2014; Bracher et al., 2015; Sathyendranath et al., 2014), research cruises (see e.g. Villar et al., 2018), and global species projections (Righetti et al., 2019a), a classification of the ocean into biomes using phytoplankton biogeography has now become possible. Phytoplankton constitute the basis of marine food-webs (Litchman, 2007), and changes in matter and energy input from phytoplankton communities therefore cascade towards higher trophic levels (Voigt et al., 2003; Sarmento et al., 2010). Phytoplankton have been shown to influence global biogeochemical cycles (e.g. Falkowski, 1998; Bopp et al., 2005; Morán et al., 2010; Litchman et al., 2015; Guidi et al., 2016), and to respond sensitively to environmental change (e.g. Reid et al., 2007; Beaugrand, 2009). Hence, analogous to terrestrial primary producers counterparts, major marine primary producers (Field, 1998) are likely to partition the ocean into geographically constrained and ecologically distinct units with relevance for ecosystem function and global biogeochemical cycling. The wide environmental niches found in phytoplankton (Brun et al., 2015), spanning on average 16 °C in the surface ocean (Righetti et al., 2019a), suggest that an ocean partitioning based on phytoplankton species distribution will result in a lower number of biomes compared to partitionings based on higher trophic level taxa, with the latter partitionings potentially being nested within phytoplankton-based biomes (Costello et al., 2017). Moreover, differences in phytoplankton composition from one region to another have been shown to reflect seasonal variability of environmental and biogeochemical patterns (Mojica et al., 2015; Righetti et al., 2019a). Biomes based on phytoplankton biogeography may integrate top-down and bottom-up factors controlling ecosystem structure and function, since both biotic and abiotic factors have been shown to govern phytoplankton composition (Lima-Mendez et al., 2015; Boyd et al., 2010; Brun et al., 2015; Endo et al., 2018), and phytoplankton species communities display local associations with distinct physicochemical conditions (Logares et al., 2014; Endo et al., 2018), as well as processes related to global biogeochemical cycling and the biological carbon pump (Guidi et al., 2016).

A major quest associated with the biogeochemical and ecological relevance of marine biomes is their link to specific ecosystem functions (Hattam et al., 2015; Townsend et al., 2018) such as the maintenance of biodiversity, productivity or essential biogeochemical processes (Gamfeldt and Roger, 2017; Manning et al., 2018). To this end, the link between biodiversity, ecosystem structure, or species composition, and essential ecological or biogeochemical functions needs to be investigated. At present, it is unclear to what degree planktonic ecosystem functions may be quantified based on the abundance or activity of individual predominant species (Schwartz et al., 2000), species networks (Guidi et al., 2016), or the full spectrum of species diversity (Goebel et al., 2014; Vallina et al., 2014; Cermeño et al., 2013). Phytoplankton species diversity has been shown to be positively correlated with ecosystem services such as primary productivity (Ptacnik et al., 2008; Goebel et al., 2014). Primary productivity may be predominantly carried out by the most abundant species rather than the full spectrum of diversity (see Schwartz et al., 2000; Cermeño et al., 2013), and certain ecosystem functions tend to increase rapidly with increasing species richness and saturate asymptotically at higher richness, (see Schwartz et al., 2000; Goebel et al., 2014; Vallina et al., 2014). While a few dominant species may essentially drive ecosystem processes at any particular snapshot in time (Lyons et al., 2005; Cermeño et al., 2016), a vast number of rare species present in the plankton (Ser-Giacomi et al., 2018) may stabilize these processes over time, as species may quickly

assemble and raise to dominance in a variable environment. As a consequence, phytoplankton networks rather than individual dominant species may maintain ecosystem function and services in the ocean, including the biological carbon pump (Strom, 2008; Decelle et al., 2012; Lima-Mendez et al., 2015; Guidi et al., 2016). Hence, a key task in plankton ecology is to determine ecosystem constituents (e.g. dominant species or species networks) that are essential to ecosystem functions. Delineating biomes that emerge from these constituents that emerge from the most common and wide-spread species may offer a valid alternative to traditional biomes and become essential tools for global marine ecosystem monitoring and management.

In this study, we partitioned the open ocean into biomes using newly available, monthly biogeographic patterns of 536 phytoplankton species (Righetti et al., 2019b), and self-organizing maps (Fendereski et al., 2014; Kavanaugh et al., 2014). Our partitioning of the ocean into biomes allowed us to test the hypotheses that (i) the open ocean can be partitioned into ecologically relevant biomes based on biogeographic patterns of phytoplankton species, (ii) biome boundaries are dynamic in nature and vary throughout the year, (iii) biomes are characterized by distinct phytoplankton species compositions, and (iv) by characteristic indicator species or phytoplankton species co-occurrence networks, and (v) biomes differ in their prevailing environmental and biogeochemical conditions, as well as in their productivity.

2. Data and methods

2.1. Phytoplankton species occurrence projections

Biomes were constructed based on monthly distribution maps of species' presence derived from statistical species distribution model (SDM) ensembles developed by Righetti et al. (2019a). These models were adapted to the relatively sparse data available for many species and addressed the spatio-temporal bias present in sampling density through pseudo-absence selection techniques (i.e. target group selection at the level of phytoplankton groups; Phillips et al., 2009). Adaptations to the sparse data also included the use of SDM ensembles with varying predictor choice and simplicity in the tuning of statistical algorithms fitted on the empirical data. Generalized additive model (GAM; Hastie and Tibshirani, 2017) ensembles with multiple predictors served as the standard in Righetti et al. (2019a), along with Generalized Linear Models (GLMs; Nelder and Wedderburn, 1972) and Random Forests (RFs; Breiman, 2004). We stick to the GAM-based species projections as the basis for our analysis of species' monthly biogeographic patterns.

Original data used to train successful GAM ensembles included 541′926 phytoplankton presence records, spanning 536 open ocean species, and ten phyla or classes: *Cyanobacteria, Chlorophyta, Cryptophyta, Dinophyceae, Haptophyta, Bacillariophyceae, Chrysophyceae, Pelagophyceae, Raphidophyceae*, and *Euglenoidea*. Additionally, observations of *Prochlorococcus*, *Synechococcus*, and *Phaeocystis* were included at the genus level, but treated as species herein. Data spanned all major ocean basins, latitudes, and most seasons. Data were cleaned according to multiple criteria, and taxonomically harmonized between sources following the backbone taxonomy of the World Register of Marine Species and Algaebase (Guiry and Guiry, 2017). Sources included the Global Biodiversity Information Facility, the Ocean Biogeographic Information System, Villar et al. (2015), Sal et al. (2013), and MAREDAT (Buitenhuis et al., 2013). Data collected at locations where the water depth was shallower than 200 m and/or at locations with surface salinities below 20 PSU were excluded to address open ocean conditions and to avoid coastal effects. The final occurrence data considered were sampled mostly near the surface of the open ocean (median observational depth of 7.5 m), and predominantly during the period from 1950 to 2000 (1984 ± 17; mean ± standard deviation; Righetti et al., 2019a).

Prior to their inclusion in SDMs, the phytoplankton occurrence data were binned to monthly (climatological) 1° latitude × 1° longitude resolution. For each monthly 1°-pixel, multiple presence records of a species were thereby counted as single presence, and abundances larger than zero were transformed to presence-only data (Righetti et al., 2019a). The latter served to minimize methodological heterogeneity in the data, as presence-absence data are typically less sensitive to the original sampling volumes. Second, the pooling to 1° spatial and monthly climatological resolution considered the presence information of potentially several cruises or sampling methods in an aggregate manner. Five individual GAMs, used to build the GAM ensemble, were fitted on the pooled presences of each individual phytoplankton species. These statistical fits of species' realized ecological niches, obtained from the five GAMs were projected on the global ocean and averaged to obtain a monthly ensemble mean presence-absence projection of the species (Righetti et al., 2019a).

Predictor variables in individual GAMs used to form the ensembles, were selected in a species-specific manner from a pool of 25 environmental variables that are proxies for ocean productivity and biotic interactions (e.g. chlorophyll-a concentration; O'Brien et al., 2016) or affect phytoplankton physiology and ecology (Brun et al., 2015; Rivero-Calle et al., 2015). The selection of variables was guided by an initial ranking procedure of the importance of the 25 potential predictors for the species in question, and a subsequent randomized and balanced inclusion of higher-ranked predictors into the five GAMs of the species (Righetti et al., 2019a). This procedure served to capture predictors of high importance for the particular species, while avoiding allocation bias during predictor selection. Variables included but were not limited to: sea surface temperature from Locarnini et al. (2013), (excess) nitrate, (excess) phosphate (NO_3, PO_4, N*, P*), and (excess) silicic acid ($SiOH_4$, Si*) from Garcia et al. (2013), salinity from Zweng et al. (2013), mixed layer depth from Montégut (2004), photosynthetically active radiation from NASA (2018b), chlorophyll-a concentration from NASA (2018a), sea surface wind stress from Atlas et al. (2011), and sea surface CO_2 partial pressure from Landschützer et al. (2014).

To obtain valid species distribution patterns in the presence of limited or uneven observational data (e.g. 45% of data stemmed from the north Atlantic; Righetti et al., 2019b), a critical choice in the GAMs was the selection of pseudo-absences through a 'target-group' approach (Phillips et al., 2009). This approach assumes that the total sampling sites of an entire study group (e.g. phylum or class) reflect the sampling effort applied to each individual species within the group. Pseudo-absences were selected from the target-group (i.e. from all sites where measurements of species of the same taxonomic group were made; Righetti et al., 2019a), in a randomized and stratified manner. Sampling the pseudo-absences from target-groups, thus served two purposes: pseudo-absences of species received a similar bias as the species' presence observations, thus balancing the presence-data distribution bias (Righetti et al., 2019a). Second, none of the pseudo-absences were selected from areas devoid of observations, thus avoiding the problem of classifying vast areas with sampling gaps as areas of species absence (Righetti et al., 2019a). For major taxa, namely the *Bacillariophyceae*, *Dinoflagellata* and *Haptophyta*, which showed different global sampling distribution patterns (Righetti et al., 2019b), target-groups were defined as all species of the respective taxon. For the remaining taxa, too few species were available to define a reasonable target-group (Righetti et al., 2019a). Thus, a broader set of phytoplankton species served as the target-group, reflecting the global phytoplankton sampling effort in general (Righetti et al., 2019a).

Original species projections have been subjected to SDM quality criteria (Righetti et al., 2019a), using a true skill statistic (Allouche et al., 2006) threshold of at least 0.35 during cross validations, and sensitivity tests to methodological choices affecting projections (e.g. pseudo-absence selection, complexity, or predictor choice). The final GAM-based projections used in subsequent analyses include a total of 536 phytoplankton species, spanning 166 genera and nine higher-rank taxa (Table 1). Monthly ensemble mean projections were transformed to presence-absence, with values larger than 0.5 counting as present. The twelve months of the year provide collectively 341′647 1° latitude × 1°

Table 1
Number of genera and species per phylum/class found in the monthly phytoplankton species presence dataset.

Phylum/Class	Number of genera	Number of species
Dinoflagellata	61	258
Bacillariophyceae	72	232
Haptophyta	21	32
Chlorophyta	3	4
Cyanobacteria	3	4
Cryptophyta	2	2
Dictyochophyceae	2	2
Chrysophyceae	1	1
Euglenozoa	1	1

longitude cells with presence-absence predictions of species community composition, where each $1^\circ \times 1^\circ$ grid cell in each month is characterized by a 536-dimensional (number of species) presence-absence vector. For each $1^\circ \times 1^\circ$ grid cell, the presence-absence projections of the 536 species are referred to as the (projected) phytoplankton community. The phytoplankton communities were the basis to determine open ocean biomes using a self-organizing map algorithm (SOM; Kohonen, 1990) and hierarchical agglomerative clustering (Jain et al., 1999).

2.2. Self-organizing map

The projected phytoplankton community information (at global monthly $1^\circ \times 1^\circ$ resolution) was used to train a self-organizing map (SOM; Kohonen, 1990). We used the Neural Network Toolbox® of Matlab R2017b with batch learning and hexagonal topology for the neuron lattice (Beale et al., 2017). The SOM is an unsupervised clustering tool, which projects high-dimensional input data onto a two-dimensional lattice of neurons (Kohonen, 1990, 2001), i.e. for this project it groups input data into subsets of similar patterns of presence-absence observations. All observations in a subset are thus assigned to a neuron, which represents the mean presence-absence pattern of the subset (Vesanto and Alhoniemi, 2000). These neurons are represented as a 536-dimensional vector of the mean presence-absence pattern, and their ID/label indicates their position in the two-dimensional lattice of neurons. We chose batch learning to ensure that all observations are shown to the neurons prior to updating their values (Beale et al., 2017). The Manhattan distance was chosen to assess the similarity between neurons and species projection patterns, because it is more sensitive to small differences than the Euclidean distance for high-dimensional data (Aggarwal et al., 2001).

The training of the SOM required the selection of an optimal number of neurons, and epochs, i.e. the number of times the SOM is trained with the input data (Sinha et al., 2009). In a first step, we approximated the optimal number of neurons M of the two dimensional neuron lattice based on the total number of data points n in the training dataset ($n = 341'647$), following the approach by Vesanto and Alhoniemi (2000):

$$M = 5 \times \sqrt{n} = \frac{1}{12} \sum_{i=1}^{12} (5 \times \sqrt{n_i}) = 834, \quad (1)$$

where the number of data points (n_i) denotes the number of $1^\circ \times 1^\circ$ pixels with phytoplankton species community information resolved to monthly resolution (average $n_i = 28'471$).

The approximation $M = 843$ corresponds to a 29×29 neuron lattice, which formed the starting point for our optimization. The optimal number of neurons corresponds to the first neuron lattice size where the total error does not decrease significantly when further neurons are added, with the threshold for improvement in the total error being 5% of the initial performance gradient (see supplement in Fendereski et al., 2014). To identify the optimal lattice size, we trained SOMs with different values for M ranging between 25 (5×5) and $4'900$ (70×70) neurons (Fig. A.9), and compared them to one another in terms of their total error (Fendereski et al., 2014), defined as the sum of the topological error (TE; Kiviluoto, 1996) and the Manhattan distance based quantization error (QE; Kohonen, 2001; Section A.1). A lattice size of 31×31 neurons ($M = 961$) was identified as the optimal lattice size (Fig. A.9a).

The optimal number of epochs was determined by evaluating the total error of the standard 31×31 lattice as a function of the number of epochs, i.e. where the number of data presentations ranged between [1, 1000] (Fig. A.9b). The optimal number of epochs was determined to be 200, since a further increase in the number of epochs did not reduce the total error (Fig. A.9b). As these appear to be the optimal settings for our SOM in relation to the phytoplankton presence projections, for the remainder of the study we only used SOMs with a lattice size of 31×31 neurons, 200 epochs, and the Manhattan distance as its similarity metric. Using this SOM, we performed a first dimensionality reduction of the phytoplankton presence projections, which formed the basis of our subsequent analysis.

2.3. Clustering and identification of the optimal number of clusters

A major challenge in cluster analysis is the objective determination of an optimal number of clusters (Xu and II, 2005), which best summarizes the dataset in question (Milligan and Cooper, 1985; Vesanto and Alhoniemi, 2000; Salvador and Chan, 2004; Oliver, 2004; Fendereski et al., 2014). We used a K-fold cross validation approach to determine the optimal number of clusters, to avoid both over-generalization and overfitting of data (Sugar and James, 2003; Zaki et al., 2014; Aggarwal, 2015).

The optimal number of clusters was identified through four sets of K-fold cross-validation experiments. K was set to ten, five, three, and two, i.e. the fraction of data left out in the training process was increased from 10% to 20%, 33%, and 50% (Fig. 1; De'ath and Fabricius, 2000; Zaki et al., 2014; Fendereski et al., 2014; Aggarwal, 2015). The cross-validation experiments consisted of three steps: (i) splitting of the data into K equally sized, randomly chosen, non-overlapping subsets prior to training one SOM on K-1 of these subsets, (ii) clustering of the trained neurons of the SOM into 2 to 100 clusters, and (iii) predicting the data structure of the remaining subset of the data (i.e. of the data left out). This procedure was repeated until all K subsets were used K-1 times for training, and once for validation.

In step (i), we trained ten, five, three, and two optimized SOMs (Section 2.2) with 90%, 80%, 66%, and 50% of the $1^\circ \times 1^\circ$ pixels with phytoplankton species composition projections, respectively. For each SOM, this resulted in 961 trained neurons with a specific biogeographic phytoplankton species composition presence-absence pattern (consisting of 536 species). This step group the phytoplankton species presence data into a 961×536 matrix. This means that each of the 961 neurons is representative for a subset of the global monthly grid cells with phytoplankton community projections.

Following step (i), we performed an additional dimensionality reduction of the 961×536 matrices through a principal component analysis (PCA; Abdi and Williams, 2010). This step was necessary, since the performance of clustering algorithms generally decreases with an increasing number of features (here defined by the total number of projected species; $N = 536$ Bellman, 2015), and the number of neurons ($M = 961$) should significantly exceed the number of projected species that characterize each neuron (Cunningham, 2008). Thus, we projected our 961 trained neurons onto their principal components with an eigenvalue above 1 (Kaiser's rule; Wilks, 2011).

After reducing the 961×536 matrix through PCA, we applied a hierarchical agglomeration clustering (HAC; Jain et al., 1999) with the Manhattan distance as the similarity metric in step (ii). We used the weighted average-linkage (Hastie et al., 2009), because this linkage is statistically more robust than the complete- or single-linkage clustering (Hastie et al., 2009). Through HAC, we grouped the PCA-transformed neurons into 2 to 100 clusters, which is equal to 99 levels of

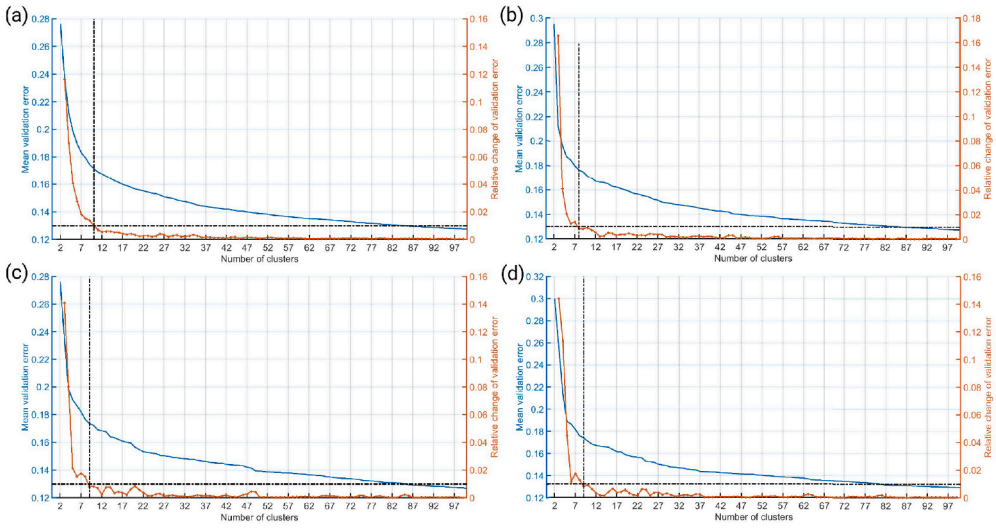


Fig. 1. Results of the four K-fold cross-validation experiments with increasing increasing leave-out data, i.e. fractions of the data that are left out during training and spared for validation. The leave-out data amounted to: (a) 10%, (b) 20%, (c) 33%, and (d) 50%. Blue lines show the mean validation error as a function of the number of clusters. Red lines show the mean absolute change of the validation error (i.e. the loss in validation error by adding one more cluster). The mean validation error is calculated as the average difference between observations in the validation dataset and the centroid of their corresponding clusters emerging from the training dataset.

granularity in the clustering. This chosen range in the amount of clusters (2 to 100), includes the amount of biomes and provinces reported previously (e.g. 57 in Longhurst, 2007), while allowing for unanticipated higher—or lower—levels of structural detail than reported in Longhurst's scheme or other biome ocean partitionings (e.g. Oliver and Irwin, 2008; Fay and McKinley, 2014; Kavanaugh et al., 2014; Reygondeau et al., 2011; Costello et al., 2017). For each level of granularity, each of the 961 neurons was dynamically associated with one of the available clusters, e.g. for the first level (two clusters) each neuron was either dynamically associated with the first or the second cluster, and the second level (three clusters) each neuron was dynamically associated with one of the three clusters, and so forth.

Following Fendereski et al. (2014), we calculated the cluster centroids, to evaluate the goodness of fit of each individual level of granularity tested during the clustering procedure. For each individual cluster at each level of granularity, we calculated the mean of the trained neurons associated with the cluster (without PCA transformation). This mean value was weighted by the number of times a cluster occurred throughout the year (cluster centroids). To this end, we associated each 1°-pixel in the validation set to a neuron based on the minimum Manhattan distance between the community presence of a pixel and the trained neurons. Using the association between neurons and clusters, we then assigned a cluster centroid to each 1°-pixel.

During the final step (iii), we calculated the goodness of fit for each level of granularity based on the difference between each 1°-pixel in the validation set and its corresponding cluster centroid. This difference was calculated as the mean validation error D_{xy} for each of the K cross-validation experiments $x = [1, K]$ ($K = 10, 5, 3, 2$), at each level of granularity $y = [2, 100]$ (Fendereski et al., 2014). D_{xy} represents the difference between phytoplankton species presence projections v_{ijxy} contained in the validation dataset and the corresponding predicted cluster centroids w_{ijxy}:

$$D_{xy} = \frac{1}{536}\sum_{j=1}^{536}\left(\frac{1}{N}\sum_{i=1}^{N}|v_{ijxy} - w_{ijxy}|\right), \quad (2)$$

where i denotes a specific 1° × 1° pixel in a specific month, j denotes one of the 536 phytoplankton species, and N stands for the number of observations in the validation set (34′165).

After calculating D_{xy} for each of our four validation experiments, we calculated an average validation error (D_y) at any given K as follows:

$$D_y = \frac{1}{K}\sum_{x=1}^{K} D_{xy}. \quad (3)$$

We analyzed the evolution of D_y as a function of the increasing number of clusters, and for each fraction of data left out during training (Fig. 1). We identified the minimum number of clusters beyond which D_y did not decrease by more than 1% relative to the maximum D_y for at least three consecutive increases in the number of clusters (Oliver, 2004). The stagnation in the decline of the mean validation error was reached at ten clusters in the 10-fold cross-validation (Fig. 1a), at eight clusters in the 5-fold cross-validation (Fig. 1b), and at nine clusters in the 3-, and 2-fold cross-validations (Fig. 1c, d). We chose nine clusters as our decisive number of clusters beyond which the mean validation error stagnated, as this value was the mean value of our four cross-validation experiments, and also the result of the most stringent cross-validation tests (i.e. at the largest fractions of data left out; 33%, and 50%). These insights from the cross-validation experiments were used to calculate the final clusters used in all further analyses as follows: first, we trained a new SOM using the entire dataset, which resulted in a 961 × 536 matrix. Second, we reduced the dimensionality of the 961 × 536 matrix to a 961 × 8 matrix using the principal components found based on Kaiser's rule ($n = 8$; associated eigenvalue > 1; Table A.6). The resulting principal components explained 75.19% of the variance in the trained neurons across all 536 species, where the first three principal components (PCs) captured the majority of the variance (40.89% explained by PC1; 12.36% by PC2; 7.87% by PC3, 4.69% by PC4, and the remaining 9.38% by PC5 to PC8). Third, we clustered the 961 × 8 matrix into the optimal number of nine clusters using HAC, and we calculated the cluster centroids for these nine clusters.

2.4. Transforming clusters into biomes and smoothing procedure

We assessed the spatial distribution of the nine clusters (Section 2.3) at the monthly, seasonal mean, and annual mean scale. At the monthly scale, we defined spatially coherent biomes that covered at least 0.5% of the surface ocean area. Based on these monthly biomes, we derived biomes at the seasonal and annual scale using the biome that was most frequently predicted at monthly scale, at any particular location, as the seasonal or annual biome, respectively.

To obtain monthly biomes, we eliminated small heterogeneities in the distribution of our nine monthly clusters. Heterogeneities were defined as small patches of a particular cluster that were fully or partly contained within one of the other eight clusters, and with an area below 0.5% of the global surface ocean area. We reassigned these small patches to the cluster with the most similar cluster centroid (as visualized in the dendrogram of Fig. 4a). In cases where reassignment was not possible because patches were surrounded by land or missing

values (e.g. in the Sea of Japan), we assigned a missing value. This reassignment was repeated until each patch was assimilated by another cluster, and the combination covered an area above 0.5% of the available surface ocean area.

To visualize seasonal or annual biomes (integrating three months, or twelve months, respectively), an aggregation of the monthly biomes was needed. To each 1°-pixel, we assigned a seasonal biome by selecting the monthly biome that occurred more often than 50% of the months in that season, or over the full year, respectively. If monthly biomes occurred at even frequency in a pixel and season or year, we assigned the phytoplankton community of that pixel to the biome with the most similar cluster centroid. The similarity was defined as the distance between the phytoplankton community matrix and the biome centroids. Potential emerging patches (with area coverage below 0.5% of the global ocean area) were reassigned to the next most similar cluster as described above (smoothing procedure). The aggregation and post-processing of monthly biomes resulted in eight seasonal biomes, and seven annual biomes. We note that one out of the nine monthly biomes only occurred in four months. This biome was not frequent enough to be reflected in seasonal or annual biomes and had overall the least mean spatial coverage (1.2%; A.8). We thus excluded it from further analysis, leading to a final distinction of eight monthly biomes.

2.5. Robustness assessment of the optimal number of clusters

Next, we assessed the robustness of our classification in terms of its sensitivity to information loss, i.e. we tested whether the optimal number of clusters and their spatio-temporal distribution depended strongly on subsets of the initial training data. To this end, we tested the robustness of our eight biomes to (i) spatial/temporal information loss and (ii) feature (species) loss during training.

First, we assessed the robustness of our classification to loss of spatio-temporal phytoplankton community information. To this end, we retained all 536 species, but trained five optimized SOMs (Section 2.2) on a decreasing fraction of 99%, 95%, 90%, 80%, and 70% of the original monthly $1^\circ \times 1^\circ$ phytoplankton community projections, chosen at random from the original data set (spatial/temporal data loss experiment). Second, we trained five optimized SOMs (Section 2.2) with all monthly $1^\circ \times 1^\circ$ phytoplankton community projections, but excluded 1%, 5%, 10%, 20%, and 30% of the 536 species chosen at random (species data loss experiment).

The training of the SOMs in both tests resulted in 961 neurons, which we grouped into nine clusters. We calculated the cluster's spatial monthly distribution, and transformed them into seasonally corrected monthly biomes (Section 2.4). We quantified the robustness of the method with respect to data loss, using the Kappa index of agreement (Cohen, 1960; Landis and Koch, 1977). The Kappa index ranges between one and zero, where a Kappa index above 0.6 indicates a substantial/good agreement between maps (Landis and Koch, 1977; Monserud and Leemans, 1992; Fendereski et al., 2014). This index has previously been used to assess the robustness of the biogeographic classification of the Caspian Sea (Fendereski et al., 2014), and for the comparison of global vegetation maps (Monserud and Leemans, 1992). Here we used the surface area of the individual 1°-pixels to weight the Kappa index. We calculated the average monthly Kappa index between our monthly biomes and those produced using test data (Table 2).

Overall, the average monthly Kappa index ranged between 0.66 ± 0.06 and 0.85 ± 0.10 in the spatial/temporal data loss experiment, and between 0.72 ± 0.08 and 0.94 ± 0.03 in the species data loss experiment (Table 2). This indicates a substantial (Landis and Koch, 1977; Monserud and Leemans, 1992) agreement between our biomes and the biomes produced using test data. However, in both robustness experiments, the average monthly Kappa index does not decrease monotonically with increasing fraction of missing data. This suggests that the general patterns remain consistent across experiments, but the exact spatial distribution of our biomes in the robustness experiments depended on the data subsets. Additionally, the Kappa index is generally higher in the species loss experiment compared to the spatial/temporal data loss experiment (except for the 70% data loss level; Table 2). Thus our biomes are more sensitive to spatial/temporal data loss than species loss, which is likely due to the decreasing performance of our clustering algorithms with increasing number of features (species) in a given dataset (Bellman, 2015), as randomly removing $1^\circ \times 1^\circ$-pixels effectively results in an increase of species compared to the available observations.

2.6. Testing our hypotheses

Our hypotheses were tested using the monthly biomes (Section 2.4). For each biome, we calculated the median number of species found across the months as a proxy for the projected species composition. We visualized the difference between biomes with regard to their phytoplankton species composition using non-parametric multidimensional scaling (Section 2.6.1; NMDS; Clarke, 1993). Indicator species were identified for each biomes using the monthly surface area occupied by the 536 phytoplankton species (Section 2.6.2). The role of species networks was assessed by identifying phytoplankton species that frequently occurred together in the monthly biomes (Section 2.6.3). Finally, we characterized the environmental conditions of the monthly biomes (Section 2.6.4).

2.6.1. Comparison of phytoplankton species composition between biomes

To test whether our biomes harbor unique phytoplankton compositions (Townsend et al., 2008), we quantified and inter-compared four aspects of phytoplankton diversity: (i) biome-specific species richness, (ii) taxon-specific diversity, which takes into account the number of species occurring within each phylum and acts as a proxy for dominance of certain taxa, (iii) species composition, i.e. the presence-absence patterns of all species within each biome, and (iv) differences in dominant species across biomes.

To calculate the biome-specific temporal mean diversity, we (i) integrated monthly species richness at the biome scale, i.e. we counted the total number of phytoplankton species across all 1°-pixels within each monthly biome (Table A.9). We then calculated the median and interquartile range of the monthly numbers of phytoplankton species per biome across the twelve months. To examine aspect (ii), we calculated the median number and interquartile range of different species per phyla/class separately, across the twelve months (Tables A.10,A.11,A.12).

To quantify differences between biomes in terms of species composition (iii), we applied a NMDS (Cardoso et al., 2017; Fendereski et al., 2014) to the species presence-absence patterns contained within each of the trained neurons in our optimal SOM (Section 2.3). The NMDS

Table 2
The average and standard deviation of the monthly kappa index of agreement between our biomes and biomes produced using a decreasing fraction of 99%, 95%, 90%, 80%, and 70% of the original monthly $1^\circ \times 1^\circ$ phytoplankton community projections, chosen at random from the original dataset (*Spatial/temporal data loss experiment*), and the monthly kappa index of agreement for the comparisons between our biomes and monthly biomes produced after excluding 1%, 5%, 10%, 20%, and 30% of the 536 species chosen at random (*Species data loss experiment*).

Biome	99%	95%	90%	80%	70%
Spatial/temporal data loss	0.77 ± 0.16	0.78 ± 0.12	0.73 ± 0.17	0.66 ± 0.06	0.85 ± 0.10
Species data loss	0.94 ± 0.03	0.78 ± 0.07	0.76 ± 0.12	0.79 ± 0.11	0.72 ± 0.08

projects the multidimensional data onto a two dimensional space, where the distances between the projected points correspond to the (dis) similarity in the multidimensional data (Clarke, 1993). We used the Manhattan distance as our dissimilarity metric of species composition. The NMDS was used to visualize both between-biome differences in species community compositions, and within-biome differences in species community composition. To examine differences between biomes in community composition, we first depicted the core region of each biome in the two-dimensional scaling space. To this end we drew an envelope embracing the 10th to 90th percentile ranges of the species communities of each biome (i.e. we used the median-centered 80% confidence interval of species communities, in the two-dimensional scaling space). Between-biome differences in the range of community compositions were quantified using the envelope overlap between two biomes (Table A.13). Within-differences in projected species communities were calculated using the spread of the envelope in both dimensions (Table A.14).

To denote dominant species in each biome we determined the distribution and taxonomy of the most frequently found wide-spread species within each biome, i.e. we assumed that species with a large area coverage also represent dominant species. This assumption stems from the observation that the majority of species in our analysis (490 out of 536) are representatives of the comparatively larger-bodied marine micro-plankton (Table 1). The relative representation of species included in our analyses, relative to the total marine phytoplankton species known in the literature, is roughly 7% to 14% (depending on the exact phylum or class). By contrast, smaller-sized plankton groups, containing also less species, are represented by only about 3% of the species known (Righetti et al., 2019a). Thus, our species projections reflect the occurrence patterns of the larger micro-phytoplankton species groups—including functionally important, silicifying Bacillariophyceae, calcifying haptophytes, and the large group of dinoflagellates—to a better degree compared to the nano- and pico-pythoplankton.

The original phytoplankton occurrence data remain incomplete with regards to species detection (Cermeño et al., 2014), and does not allow us to include particularly rare species into our biome definitions, despite the fact that they account for a large fraction of the local species richness (Ser-Giacomi et al., 2018). Thus, the species for which enough occurrence data were available to train SDMs, are species that have frequently been observed, possibly due higher abundances and commonness. Frequent species have been shown to account for the majority of phytoplankton biomass or abundance within different phytoplankton groups, with 16 species of coccolithophore out of 195, or 43 diatom species out of 552, accounted for 75% of coccolithophore abundance (O'Brien et al., 2013), or 90% of total diatom biomass (Leblanc et al., 2012), in marine analyses. In addition, common species, unlike rare ones (Ser-Giacomi et al., 2018), have been shown to carry most biogeographical information and ecosystem function (Lyons et al., 2005; Cermeño et al., 2013), suggesting that their distribution patterns reflect the ecologically and biogeochemically important processes.

For each biome, wide-spread species were defined as those which rank in the top 100 based on the species' biome-specific area coverage. To this end, we calculated the area-weighted fraction of pixels covered by each species within a biome, relative to the total biome area (Table A.19).

2.6.2. Identification of indicator species

In analogy to the flora and fauna characteristic of certain terrestrial biomes (e.g. Picea abies as species typical for the boreal biome; Yang et al., 2020), we explored whether our biomes can be discerned based on indicator species. This analysis served to test the fourth hypothesis, i.e. that biomes can be characterized through indicator species or species composition (Section 2.6.3).

To assess whether our biomes can be unequivocally identified based on a common, easily identifiable subset of species, we categorized the phytoplankton species found in each biome using the core-satellite species hypothesis (Hanski, 1982). This hypothesis differentiates between two types of species (Gibson et al., 1999; Ellison, 2019): *core species*, which are abundant and common (i.e. found in $>$ 90% of a region), and *satellite species*, which are sparse and uncommon (i.e. found in $\leqslant$10% of a region). While the species found between these thresholds are not specified in the core-satellite hypothesis, an extension of this hypothesis recognizes them as *subordinate species* (Gibson et al., 1999), adopted herein.

To classify a species as a core, subordinate, or satellite species, we calculated the relative presence area for each of the 536 phytoplankton species, based on the $1^\circ \times 1^\circ$ phytoplankton species presence data for each biome in each month (normalized area coverage). The normalized area coverage of each phytoplankton species (ranging between 0 and 1) was used as a proxy for the *ubiquity* of a species within a biome, i.e. the generality of its spatial presence within a biome. The normalized area coverage of all species per biome was used as a proxy for habitat heterogeneity, where biomes harboring many species with a minor normalized area coverage are more heterogeneous than biomes harboring the same number of species, with each species covering a large relative area. For each biome and month we calculated a vector of 536 species with a number between zero and one specifying the normalized area where the species was present. For each biome, we then calculated the average normalized area from the twelve monthly vectors to visualize the habitat heterogeneity of our biomes.

To identify core, subordinate, and satellite species we used the monthly vectors of the normalized area coverage of the 536 species for each biome and the thresholds mentioned above (Gibson et al., 1999). Thus, the core, subordinate, and satellite species were identified for each month and each biome, and indicator species were defined as those that are core species in one specific biome (i.e. found in $>$ 90% of a biome), and satellite species in all other biomes in a given month (i.e. found in $\leqslant$ 10% of a biome).

2.6.3. Identification of phytoplankton species co-occurrences

We further tested whether the species compositions of our biomes were determined by the potential species co-occurrences derived from the species presence-absence projections, and whether biomes can be identified solely based on a subset of species that characterize biome-specific phytoplankton compositions. For this, we used a text analysis algorithm that identified pairs of species which co-occurred more frequently than expected by chance, given their individual areal coverage (Dunning, 1993; Wahl and Gries, 2018).

The text analysis algorithm assigned an association score to pairs of phytoplankton species found in all $1^\circ \times 1^\circ$ pixels of a biome during all months (Dunning, 1993), reflecting the intensity of links between species. The first step served to construct a contingency table, which contained the number of times both species were present (k_{11}; co-occurrence), the number of times only the second species was present (k_{12}; alternating occurrence), the number of times only the first species was present (k_{21}; alternating occurrence), and the number of times both species were absent (k_{22}; co-absence) at the 1° spatial and monthly resolution per biome. Based on this contingency table, the algorithm used Shannon's entropy (H), to compare the presence probability of the two species found together against the presence probability of the two species found individually. H is a measure for the uncertainty in the observed probability distribution (Ricotta, 2002; Ricotta, 2005), and is then used to define the following likelihood ratio (LLR; Dunning, 1993):

$$LLR = 2 \cdot (\sum k) \cdot [H(k) - H(k_r) - H(k_c)], \tag{4}$$

$$H(k) = \frac{k_{ij}}{\sum k} log\left(\frac{k_{ij}}{\sum k}\right), \tag{5}$$

$$k = \begin{pmatrix} k_{11} \\ k_{12} \\ k_{21} \\ k_{22} \end{pmatrix}, \tag{6}$$

$$k_r = \begin{pmatrix} k_{11} + k_{12} \\ k_{21} + k_{22} \end{pmatrix}, \tag{7}$$

$$k_c = \begin{pmatrix} k_{11} + k_{21} \\ k_{12} + k_{22} \end{pmatrix}. \tag{8}$$

The values of *LLR* are non-negative, and the magnitude of the value indicates the level of significance of the co-occurrence (Dunning, 1993; Evert, 2009). The differentiation between positive (k_{11}) and negative (k_{12}, k_{21}, k_{22}) associations is done by multiplying the *LLR* by -1, if the observed occurrence frequency of the species pair is lower than the product of occurrence frequencies of the individual species divided by the sample size (i.e. the total amount of pixels in a biome across all months; Evert, 2009). Thus, negative values indicate pairs of species that co-occur less than the product of occurrence frequency of the individual species (i.e. the expected co-occurrence).

This procedure resulted in a 536×536 matrix with the *LLR*-values for all possible species pairs per biome and month. As our goal was to find species co-occurrences characteristic to individual biomes, we were interested in significant species pairs, i.e. species pairs with positive co-occurrences in one biome and alternating occurrences or co-absences in all other biomes. Additionally, the positive co-occurrences should be common for a biome (i.e. the pairs should be present in most of the biome area). For each month, common positive co-occurrences were those with a positive *LLR*-value and a minimum area coverage (cf. Section 2.6.2) of the involved species above the 90*th* percentile of the area coverage of individual species in that biome. Alternatively, uncommon or rare co-occurrences were those with a negative *LLR*-value or a minimum area coverage of the involved species below the 10*th* percentile of the area coverage of individual species in all other biomes. We used the area coverage approach to exclude species pair that had a large *LLR* value, yet were only present in a few 1°-pixels of a biome.

Next, we defined species co-occurrences that consisted of at least two significant species pairs with strong co-occurrences. For instance, a strong co-occurrence of species *A* and *B*, and species *A* and *C* represented a co-occurrence consisting of *A*, *B*, and *C*.

Finally, we used the presence distribution of all species that formed significant pairs in this analysis to train a SOM and produce biomes. The monthly spatial patterns of these biomes were compared to those of the original biomes based on the area overlap between the biomes of both clusterings. This resulted in a measure of how well the biomes were represented using only the selected species co-occurrences.

2.6.4. Biome-specific differences in environmental conditions

To identify variables that could facilitate biome monitoring through easily measurable environmental (e.g. using satellites or Argo-floats), rather than biological properties, we inter-compared prevailing environmental conditions at the biome scale. Any statistically significant biome differences in measurable environmental parameters such as sea surface temperature, chlorophyll-a concentration, or light could serve to compare our ocean partitioning to previous biogeochemical ocean classifications (e.g. Fay and McKinley, 2014; Gregor et al., 2019; Zhao et al., 2019).

We consider the ten core environmental predictors used to model the species' presence-absence distribution (Righetti et al., 2019a), and net primary production at monthly climatological 1°-resolution (Table A.5): sea surface temperature (SST; Locarnini et al., 2013), surface nitrate (N; Garcia et al., 2013), phosphate (P; Garcia et al., 2013), silicic acid (Si; Garcia et al., 2013) concentration, sea surface salinity (SSS; Zweng et al., 2013), photosynthetic active radiation (PAR; NASA, 2018b), surface chlorophyll-a concentration (chl; NASA, 2018a), mixed layer depth (MLD; Montégut, 2004), net primary production (NPP; Behrenfeld and Falkowski, 1997, 2008, 2017), sea surface wind stress (wind; Atlas et al., 2011), sea surface CO_2 partial pressure (pCO_2; Landschützer et al., 2014). In addition to the eleven variables listed above, we considered excess phosphate (P*; Deutsch et al., 2007) using the surface nitrate and phosphate concentrations.

To identify a set of environmental variables suitable for the distinction of phytoplankton-based biomes, we tested for differences in individual environmental parameters between our biomes, using a Kruskal-Wallis test (Wilks, 2011). We tested whether biomes differed in their distribution of these environmental parameters at the 1% significance level using all 1°-pixels of each biome across all months (Fig. A.18). For those environmental parameters in which at least one biome differed from all others, we compared the environmental conditions between each pair of biomes ($p \leqslant 0.01$) with a Tukey-Kramer multi-comparison test (Kramer, 1956). For each month, and each pairwise comparison between biomes, we counted the number of months where we could not differentiate between the two biomes based on each environmental factor. Lastly, for each biome we determined a set of two environmental parameters that were best suited to separate this biome from all others. Thus, for each combination of two environmental factors we counted the number of months where we could not differentiate between two biomes using either one or both environmental factors individually.

3. Results

3.1. Can the open ocean be partitioned into ecologically relevant biomes based on phytoplankton species biogeography?

We predict a partitioning of the open ocean into seven statistically separable annual biomes (Fig. 2), which are derived from eight seasonal and monthly biomes. This partitioning shows a clear differentiation between the high latitudes and the low latitudes. Overall, we found three tropical biomes, one high-latitude biome, one transitional biome between the high-latitudes and lower latitudes, two seasonally alternating subtropical biomes and a tropical/temperate biome that is absent on the annual scale, but present during spring, summer, and winter (Fig. 3). The dynamic nature of the seasonal biomes is mostly seen around the subtropics (Fig. A.15), and evident for all seasons, where between 15.8% (spring) and 37.4% (winter) of the 1°-pixels have a different biome association relative to the annual distribution.

The TRoPical biome (TRP) is found equatorward of 30° in both hemispheres of the Indian, Pacific, and North Atlantic Ocean (Fig. 2). TRP covers 26.8% of the surface open ocean in the annual analysis (Table 3). It is largest in summer (39.0%), where it extends between 42°N and 40°S (Fig. 3b), and smallest in winter (8.3%), where it stretches from 21°N to 18°S (Fig. 3d).

The HIgh Latitude biome (HIL) extends from the poles to around 40° in both hemispheres (Fig. 2) and covers 25.4% of the surface open ocean on average (Table 3). The area covered by the HIL biome shows little seasonality (26.3% to 30.1%; Fig. 3). Differences between seasons are linked to the loss and gain of $1° \times 1°$ projections available poleward from 60° in the Southern Hemisphere and the North Atlantic Ocean.

The WInter Subtropical biome (WIS) is located in the Pacific and Atlantic Oceans (Fig. 2), and covers on average 19.9% of the surface open ocean (Table 3). This biome is largest in winter (31.1%) and is found between 30° and 5° in both hemispheres in all ocean basin (Fig. 3d). The WIS biome is absent during all summer months (Table A.8).

The SUmmer Subtropical biome (SUS) is located in the Pacific and Atlantic Oceans (Fig. 2), and covers 14.6% of the surface open ocean in the annual analysis (Table 3). Similarly to the WIS biome, it varies in size between seasons. SUS is absent during winter months (Table A.8) and is largest in summer (18.8%), where it is located between 15° and 45° in the Atlantic Ocean and South Pacific, and between 30° and 45° in the Indian and North Pacific Ocean (Fig. 3b).

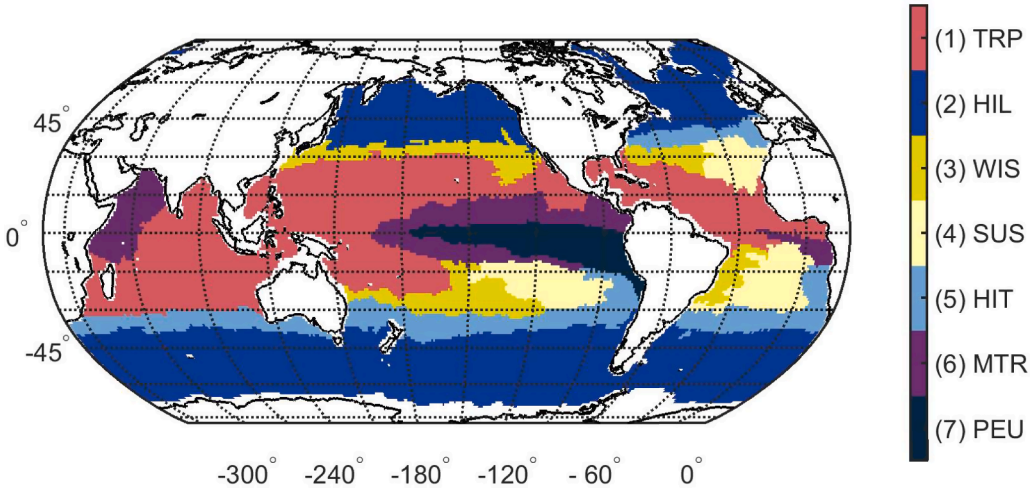


Fig. 2. Annual distribution of phytoplankton-based biomes. The annual distribution of each biome was based on monthly distributions of biomes, assigning the most frequent biome to each 1°-pixel. The final annual biome boundaries were smoothed by reducing the patchiness of initial biomes (Section 2.4). Biomes are numbered according to their mean geographic coverage across months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon biome. Note that the SMN biome is absent on the annual scale.

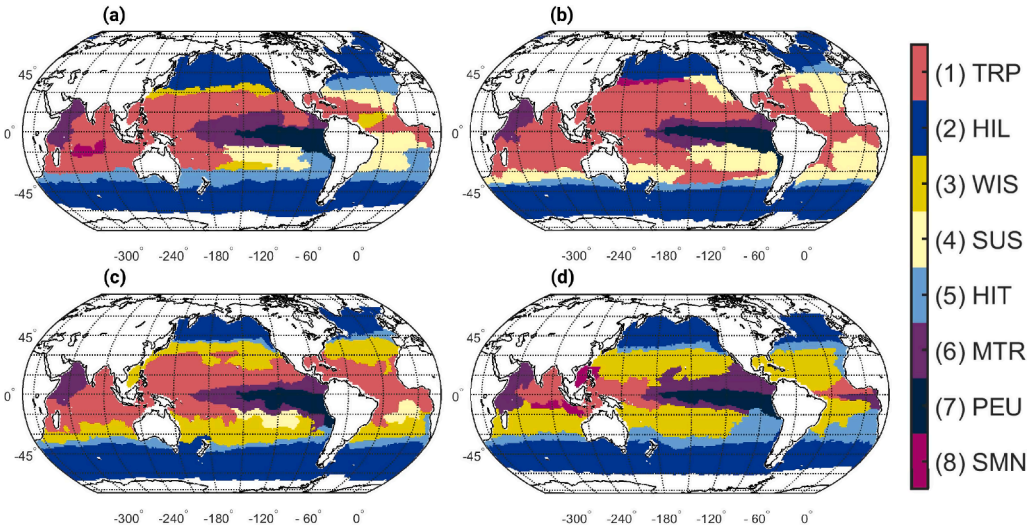


Fig. 3. Seasonal distribution of our phytoplankton-based biomes in (a) spring, (b) summer, (c) fall, and (d) winter. The Southern Hemisphere was shifted by six months for the calculation of the seasonally integrated biomes based on the monthly scale distribution of our biomes. The seasonal distribution of each biome was based on the respective monthly distribution of the biome, assigning the most frequent biome to each 1°-pixel. The final seasonal biome boundaries were smoothed by reducing the patchiness of initial biomes (Section 2.4). Biomes are numbered according to their mean annual geographic coverage across months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

The HIgh latitude Transition biome (HIT) is located between 30° and 40° in both hemispheres of all ocean basins, except for the North Pacific Ocean (Fig. 2). It covers on average 9.9% of the surface open ocean (Table 3), with little seasonality. HIT is largest in winter (14.0%; Fig. 3d) and smallest in fall (4.7%; Fig. 3c).

The Monsoon and TRopical biome (MTR) is located in the Arabian Sea and around the Pacific and Atlantic Eastern equatorial region between 15°N and 15°S (Fig. 2). MTR covers on average 8.8% of the global surface ocean area (Table 3). The location and extent of the MTR biome remains relatively constant throughout the seasons with maximum area coverage in winter (10.5%; Fig. 3d) and minimum area coverage in summer (6.1%; Fig. 3b).

The Pacific Equatorial Upwelling biome (PEU) is located in the Pacific Eastern equatorial region (Fig. 2). PEU covers on average 4.4% of the surface ocean area, and its area displays little seasonality (Fig. 3). PEU is largest in summer (4.4%; Fig. 3b) and smallest in fall (3.8%; Fig. 3). In spring and fall (Fig. 3a and c), PEU extends to 150°W, and in summer and winter (Fig. 3b and d) it extends to 180°W.

The Seasonal MoNsoon biome (SMN) does not appear on the annual scale, but it is present during spring, summer and winter in the Indian and Pacific Oceans (see Fig. 3a,b,c). SMN covers on average 1.2% of the surface ocean area. SMN is largest in winter (1.7%) where it is located in the Central Indian Ocean around 10°S, and the South China Sea (Fig. 3d). In spring, it is located in the Central Indian Ocean at 15°S as a small patch (1.0%; Fig. 3a). In summer, the SMN biome is located at the boundary between TRP and HIL in the North Pacific Ocean (0.6%; Fig. 3b).

In summary, phytoplankton-based biomes divide the ocean into tropical and high latitude biomes, with more complexity found in the tropical latitudes. Four biomes are characterized by a rather constant extent throughout the year (HIL, HIT, MTR, and PEU), and four biomes are variable in their spatio-temporal extent (TRP, WIS, SUS, and SMN). WIS and SUS transcend ocean basins, or vanish for an entire season. SMN is the smallest and temporally most variable biome as its location changes in each season.

3.2. How do biomes differ in terms of their overall and taxon-specific diversity?

In a first step we assessed differences in species richness between

Table 3
Names of the annual and seasonal open ocean biomes and their respective maximum and minimum relative area across all twelve months. Brackets indicate the month or season in which this value was reached. The mean monthly area corresponds to the biome area relative to the available open ocean area on a monthly scale, calculated using only months where the biome was present. The biomes are sorted by geographic coverage across months.

Abbreviation	Name	Max. monthly area [%]	Min. monthly area [%]	Mean monthly area [%]
TRP	TRoPical biome	43.0 (September)	6.5 (January)	26.8
HIL	HIgh Latitude biome	30.9 (March)	18.2 (December)	25.4
WIS	WInter Subtropical biome	38.6 (December)	0 (Summer)	19.9
SUS	SUmmer Subtropical biome	23.4 (June)	0 (Winter)	14.6
HIT	HIgh Latitude Transition biome	16.4 (March)	4.5 (August)	9.9
MTR	Monsoon and TRopical biome	12.7 (December)	6.4 (July)	8.8
PEU	Pacific Equatorial Upwelling biome	5.0 (December)	3.8 (October)	4.4
SMN	Seasonal MoNsoon biome	2.5 (January)	0 (Fall)	1.2

biomes. The monthly species richness (median of months ± interquartile range) of our biomes ranges from 395 ± 42.25 in SMN to 510.5 ± 12.5 in TRP (Table 4), suggesting that the overall diversity of species differs between biomes. However, this finding is not supported statistically in pairwise biome comparisons, since the interquartile range of the monthly median species richness in any biome overlaps with that of at least one other biome (Table 4). Based on the monthly median richness, our biomes can be separated into two groups. The HIL, WIS, HIT, PEU, and SMN biomes are characterized by a generally lower overall median species richness (< 500) compared to the TRP, SUS, and MTR biomes (> 500; Table 4). This separation between higher species richness in TRP, SUS, and MTR relative to the other biomes is also consistently seen for taxon-specific species diversity of the dinoflagellata, bacillariophyceae, and haptophyta.

Table 4
Monthly median numbers and interquartile ranges of overall and taxon-specific species richness found in our biomes. Taxon-specific richness was calculated for *Dinoflagellata* (Din), *Bacillariophyceae* (Bac), and *Haptophyta* (Hap). Numbers refer to the monthly number of species found in each biome as denoted in A.8. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	Total	Din	Bac	Hap
TRP	510.5 ± 12.5	249 ± 6.5	217 ± 4	31 ± 1
SUS	507 ± 22.5	249 ± 7.5	215 ± 12.75	31 ± 1
MTR	506 ± 9.5	248.5 ± 4.5	215 ± 6	30 ± 0
HIT	484 ± 19	237.5 ± 8	212 ± 13	30 ± 5.5
PEU	480 ± 13.5	239 ± 5.5	201 ± 7.5	29 ± 1
WIS	459 ± 49.5	218 ± 30.5	201 ± 15.75	27 ± 2.5
HIL	445 ± 108.5	192.5 ± 73.5	218 ± 22	23.5 ± 14
SMN	395 ± 42.25	178 ± 20.75	183 ± 23	23 ± 4.5

3.3. How do biomes differ in terms of their phytoplankton species compositions?

Since we found significant overlaps in overall and taxon-specific species richness in certain pairs of biomes, we assessed the between-, and within-biome differences in species compositions, and species communities, respectively. With regard to between-biome differences in phytoplankton community compositions, as shown by their level of dissimilarity in Fig. 4a, we find that the first partitioning of the ocean divides biomes according to latitude, i.e. into three clusters consisting of the equatorial PEU biome, the higher latitudes (HIL and HIT), and the tropics and subtropics (TRP, WIS, SUS, MTR, SMN; Fig. 4a). Thus, the projected species compositions of these three regions differ most strongly between one another. This latitudinal differentiation of biomes is also displayed in the two-dimensional projection of the phytoplankton species compositions (Fig. 4b; NMDS Section 2.6.1). On this two-dimensional projection, biomes located at higher latitudes (HIL and HIT) are associated with higher values on the first dimension, unlike biomes found at lower latitudes (Fig. 4b). The second dimension in NMDS space appears to relate to the surface nutrient concentrations, and/or the efficiency of the biological pump (see Section 3.8). Here, the PEU biome is associated with higher values on the second dimension compared to all other biomes. The two-dimensional projection confirms a clear separation of the phytoplankton species compositions of the HIL, WIS, HIT, PEU, and SMN biomes, but overlaps in the species compositions of the TRP, SUS, and MTR biomes (Table A.13). Additionally, the within-biome differences in projected phytoplankton communities are largest for biomes located near the equator, and decreasing with increasing latitude (Table A.14).

3.4. Do biomes differ in their most wide-spread, dominant species?

We examined whether the dissimilarities identified between biomes were also reflected in the biogeographic patterns of dominant species (top 100 most wide-spread species per biome; Section 2.6.1).

The dominant species identified for the TRP, WIS, SUS, MTR, PEU, and SMN biomes cover a large fractions (between 72.2% and 100.0%; Table A.19) of the respective biome areas, and are homogeneously distributed in space and time at the biome level. The dominant species identified for our biomes belong to relatively few genera. For instance, for the *Dinoflagellata* within the top 100 species of all biomes, the genera *Tripos, Prorocentrum and Oxytoxum* are found most prevalently. For the *Bacillariophyceae*, the genus *Chaetoceros* is most prevalent (Table A.19). Moreover, the cyanobacterial genera *Prochlorococcus*, *Synechococcus* (treated as species in our analysis), are among the top 100 most wide-spread taxa in all biomes, except in the PEU biome.

Some of the most dominant species occur in multiple biomes, such as the dinoflagellate *Tripos extensus*, which is found in 76.7% of the global ocean area on the monthly average, and covers at least 95.7% of the monthly area of each individual biome (except for the HIL biome; Table A.19). However, among these dominant species there are characteristic biome-specific patterns in the occurrence of dominant phytoplankton species, including diatom species known to form endosymbiotic associations with cyanobacteria in TRP, the haptophyte *Phaeocystis antarctica* in HIL, the red-tide forming dinoflagellate *Lingulodinium polyedra* (Moorthi et al., 2006) in MTR and PEU, and the haptophyte *Umbilicosphaera sibogae* (Table A.19), which is usually found in the subtropics in SMN, WIS (Baumann et al., 2016). This finding suggests that certain biomes can be defined based on the biogeographic patterns of some of their dominant species, and further opens the question of whether a characterization of our biomes based on a subset of its overall diversity, i.e. the presence of biome-specific, easily identifiable indicator species (see Section 2.6.2), were possible.

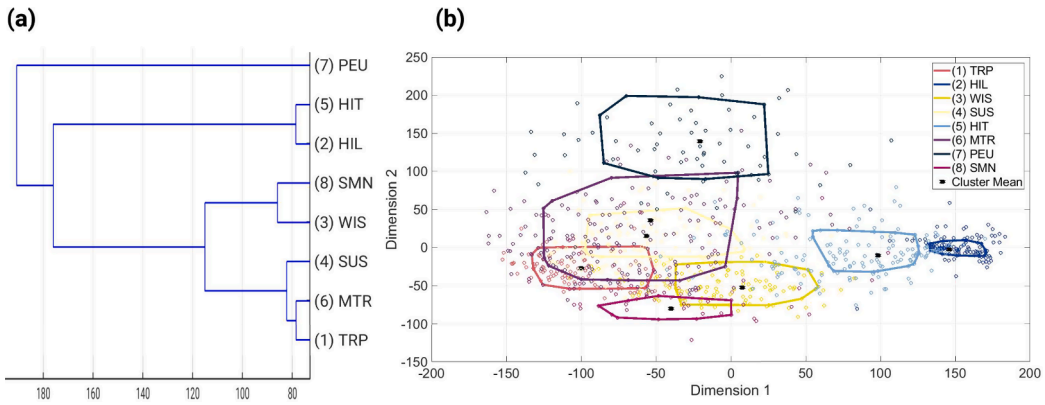


Fig. 4. (a) Similarity between the eight biomes based on the similarity of the respective centroids. The similarity was measured using the Manhattan distance as similarity metric, and the weighted average-linkage (see Section 2.3). (b) Non-metric multidimensional scaling ordination plot of the trained 961 neurons, associated with the eight biomes. The trained neurons associated with each $1° \times 1°$ pixel in each month were projected onto two dimensions, and are displayed as open dots, with different colors indicating the association to the different biomes. The black dot shows the respective median biome value in these two dimensions. Envelopes show the convex hull defined by projected points between the 10th and 90th percentile of all projected points of each biome. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

3.5. Are biomes characterized by indicator species?

To identify biome-specific indicator species, i.e. species that are wide-spread in one, but rare in other biomes, we ranked all phytoplankton species in descending order of their global average normalized area coverage (see Section 2.6.2), i.e. the normalized area coverage averaged across all biomes. This resulting list (Table A.18) exhibits a gradient from wide-spread species with a high normalized area coverage across all months towards localized species with a smaller global, normalized area coverage during only a few months. For instance the biogeographic distribution of the species with the largest average global normalized area coverage (rank 1), dinoflagellate *Tripos extensus*, extends from 30°N and 30°S throughout the year (Fig. A.16a). By contrast, the dinoflagellate *Pyrophacus vancampoae* (rank 179) is always present near the equator and its geographic habitat extends from 45°N to 45°S in at least one month (Fig. A.16b). The dinoflagellate *Heterodinium blackmanii* (rank 358) occurs only in the Pacific equatorial region and the West Indian Ocean (Fig. A.16c). Lastly, the species with the lowest average global normalized area coverage *Fragilaria striatula* (*Bacillariophyceae*) occurs mainly in the North Pacific Ocean (Fig. A.16d).

The average normalized area coverage of all phytoplankton species is shown in Fig. 5 at global and at biome scale. The gradient from wide-spread species to more localized species is apparent in the global perspective. Using the same sequence in species at a biome scale results in biome specific bar-code-like patterns, which highlight differences in the within-biome heterogeneity of species' coverage (Section 2.6.2) using their 100 most wide-spread species, each (Section 3.4). These biome specific patterns do not show a continuous decrease in average relative area coverage by the species as is the case for the global perspective. Notable patterns are seen in the HIL, HIT, and PEU biomes. The HIL and HIT biomes show higher levels of heterogeneity, and species that are wide-spread in all other biomes have a relatively low area coverage in these two biomes. Several wide-spread species of the PEU biome generally have a low average area coverage in other biomes.

Using the biome-specific patterns of relative area coverage, we categorized each of the 536 phytoplankton species into a core, subordinate, or satellite species, for each biome and month (Section 2.6.2). Although we found core species for all biomes during at least one month (Table A.16), we were not able to identify indicator species for any of our biomes, since the core species of a biome were also identified as

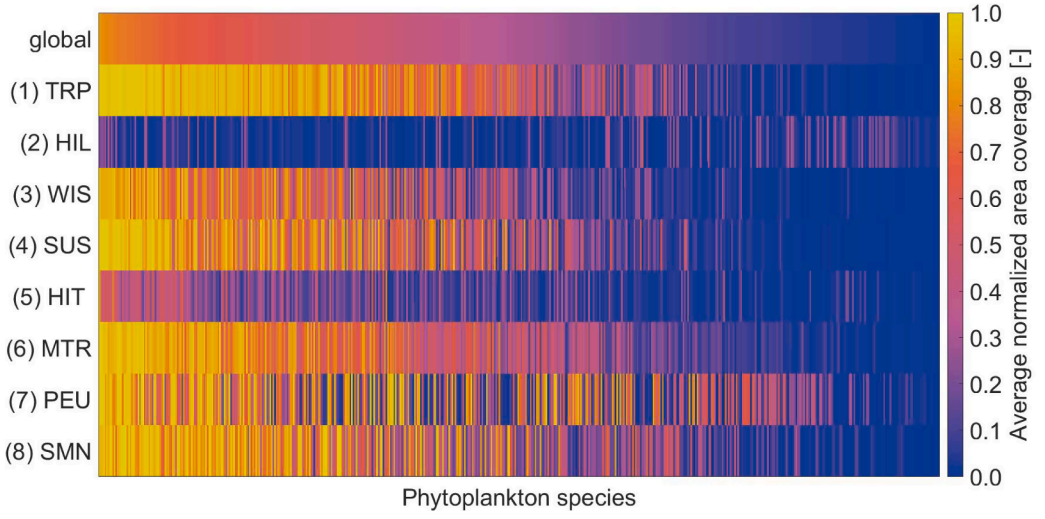


Fig. 5. Average normalized area coverage of each of the 536 phytoplankton species at a global scale and for each individual biome. The average normalized area coverage of each species was calculated as the mean of the twelve months. The colorbar denotes the average normalized area coverage of each phytoplankton species in each biome with values ranging between zero (absent everywhere throughout the year) and one (present everywhere throughout the year). The 536 species are sorted in descending order according to their global average normalized area coverage, shown in the first row (labelled "*global*"). The corresponding sorted list of the phytoplankton species is found in Table A.18. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

subordinate or core species in at least one other monthly biome (Fig. A.17). This means that we cannot distinguish our biomes based on a unique set of biome-specific phytoplankton species alone.

3.6. Do biomes harbor characteristic species co-occurrences?

Since we could not identify indicator species for specific biomes (Section 3.5), we tested whether there are characteristic species co-occurrences that would allow for the discrimination of our biomes based only on a subset of their diversity.

To this end, we explored the co-occurrence patterns of the 536 phytoplankton species, and identified significant species pairs, which we defined as the smallest building blocks of species co-occurrences of our biomes (Section 2.6.3). We visualized the species co-occurrences using a schemaball (Fig. 6; Layden, 2020), whereby two species involved in a species pair are linked by a dark line if the species pair can be found in the biome, and significant pairs are highlighted in red to orange colors. The brightness of the line indicates the magnitude of co-occurrence of the linked species. We identify 72 significant species pairs (Section 2.6.3; Table A.20) across biomes. We found at least one significant species pair for all of our biomes (red lines in Fig. 6), except for the SMN biome (depicted by the lack of red lines in Fig. 6). The number of significant pairs identified for each biome ranges from one for the TRP, WIS, and SUS biomes to 58 in the HIL biome. We found at least two significant species co-occurrences for the HIL, HIT, MTR, and PEU biomes, each.

All 72 significant species pairs together involve 51 different species (15 *Dinoflagellata*, 33 *Bacillariophyceae*, 1 *Haptophyta*, *Dictyochophyceae*, and *Chlorophyta*). Species involved in significant species pairs of the TRP, and HIT biomes include the diatom *Eucampia cornuta (125)* for the TRP biome, or *Nitzschia longissima (168)* for the HIT biome (Fig. 6). These two species are not involved in the significant species pairs of any other biome. Significant species pairs of the HIL, WIS, SUS, MTR, and PEU biomes share at least one species. For instance the HIL and WIS biomes share the species *Impagidinium patulum (330)*, the SUS, and MTR biomes share the dinoflagellate *Tripos cariiensis (471)*, and the HIL, MTR, and PEU biomes share the species *Pentapharsodinium dalei* (367; Fig. 6).

3.7. What is the role of phytoplankton species co-occurrences in the definition of biomes?

We measured how well our biomes were reproduced based only on the species involved in significant species pairs mentioned in Section 3.6 (Table A.20). In the following section we refer to the biomes based on significant species pairs as *co-occurrence biomes*.

The area overlap between co-occurrence biomes and the corresponding original biomes (analyzed at the monthly 1°-pixel with a one-to-one correspondence between a co-occurrence biome and an original biome) is 63.0% on average (Fig. 7) and ranges between 46.9% in December to 72.9% in July. The TRP, HIL and SUS biomes present the highest area overlap across months (77.7%, 94.8%, and 50.2%, respectively). The area overlap of the remaining biomes (WIS, HIT, MTR, PEU, SMN) ranges between zero for the MTR and SMN biomes and 48.7% for the PEU biome, suggesting that biomes such as the MTR and SMN biomes are not well represented by the species involved in significant species pairs. Overall, we were able to reproduce roughly 50% of the temporal and spatial features of our biomes using a relatively small subset of the entire species diversity.

3.8. Are biomes characterised by distinct environmental and biogeochemical conditions?

To determine the environmental factors that may allow for a

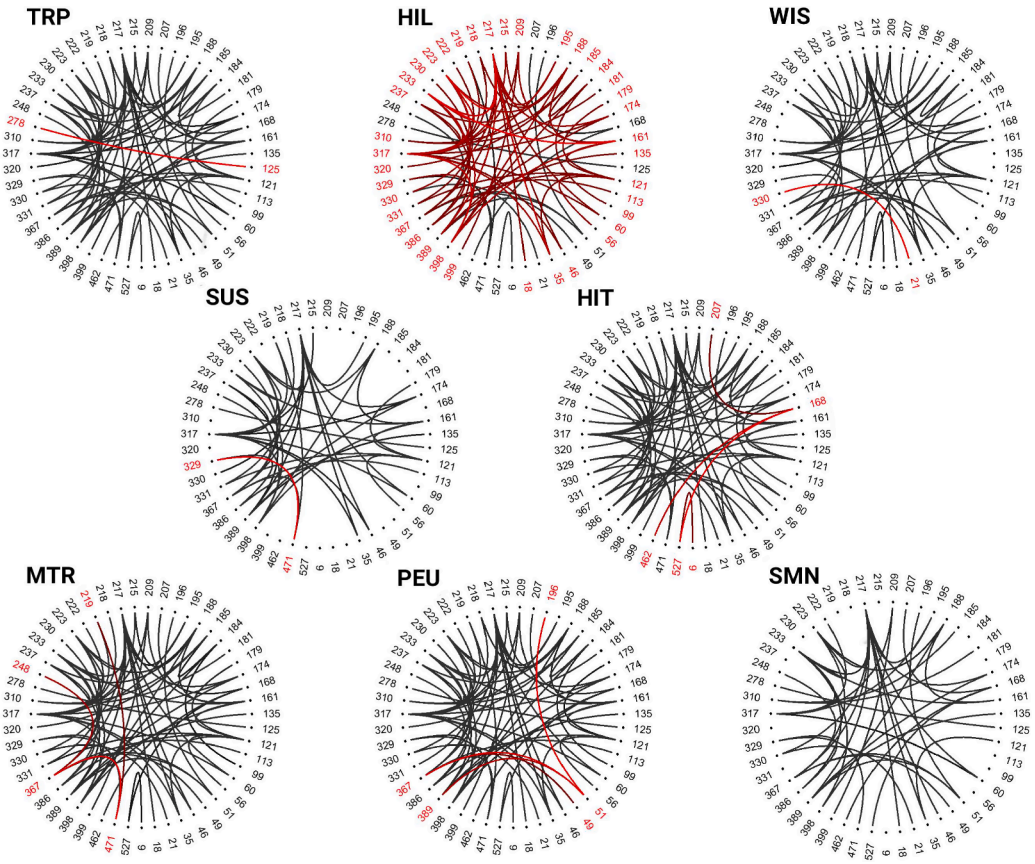


Fig. 6. Co-occurrence of significant species pairs in biomes. Each line links two species involved in a species pair. Red lines highlight significant species pairs of the respective biome. The numbers rank phytoplankton species as listed in Table A.20. The brightness of the red lines indicates the magnitude of the co-occurrence of the linked species (Section 2.6.3). The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

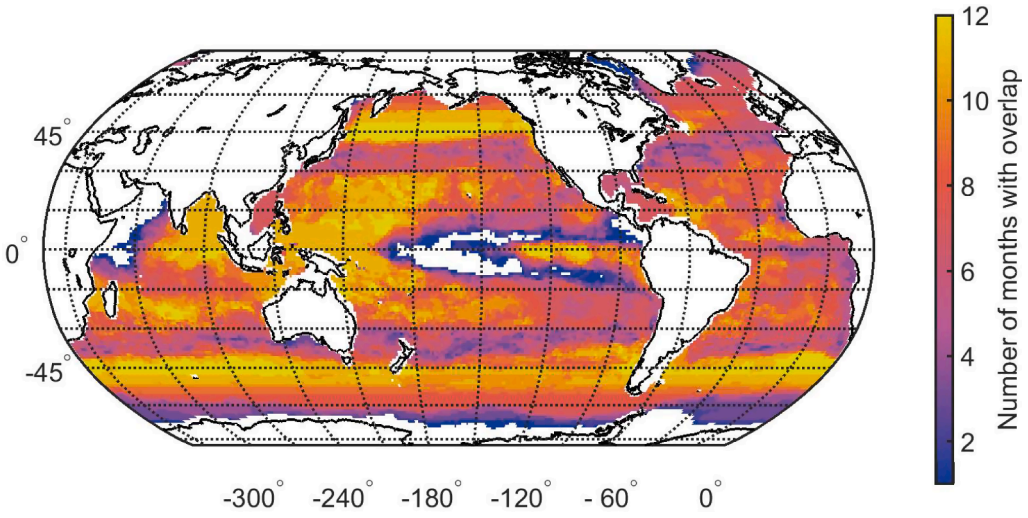


Fig. 7. The pixel-wise overlap between biomes that were produced based on a set of species with significant co-occurrences (Section 2.6.3) and biomes derived from the full set of 536 phytoplankton species. The colorbar shows the number of months in which there is an overlap between biomes. White spaces denote pixels where no overlap was found.

distinction between biomes for monitoring purposes, we examined environmental and biogeochemical differences at monthly resolution.

Biomes differ in SST, PAR, sea surface wind speed, and pCO_2 in accordance with their latitudinal distribution. P*, N, P, Si, NPP, and chlorophyll-a are highest in biomes located near the poles, decrease towards the tropics, and increase again near the equator (Fig. A.18). We performed pairwise comparisons of the environmental conditions between biomes for each individual month using a Tukey-Kramer multi-comparison test (Section 2.6.4; Kramer, 1956). The multi-comparison between biomes is summarised in Fig. 8, where each upper and lower triangular matrix, separated by the main diagonal (in gray) shows the comparison between biomes for a different environmental factor, and values denote the number of months where the monthly environmental conditions of two biomes do not differ significantly ($p \leqslant 0.01$). We rank the environmental factors according to the number of times an overlap occurred between any biome pair (nc) with regard to each factor (sum of all matrix entries in Fig. 8 for each triangular section). We determined environmental factors best suited to differentiate between biomes as those parameters with the lowest number of months with overlaps between biomes.

We find that monthly pCO_2 ($nc = 9$ comparisons with overlap between multiple biomes; Fig. 8f), SST ($nc = 16$; Fig. 8c), and Si ($nc = 20$; Fig. 8b) are the top three environmental variables that best differentiate our biomes. For the other environmental variables, biomes did not differ from one another in between 22 (monthly chl) or 37 (MDL) pairwise comparisons, with biological variables (chl, $nc = 22$; NPP, $n = 24$) differing more frequently between biomes than biogeochemical (N, $nc = 24$; P, $nc = 25$; P^*, $nc = 26$) and physical parameters (SSS, $nc = 27$; wind stress, $nc = 36$; MLD, $nc = 37$).

Moreover, we determined the environmental variable that was best suited to distinguish one specific biome from all other biomes. For instance, the TRP, MTR, PEU, and SMN biomes were best differentiated from all other biomes using pCO_2 (Fig. 8f), the WIS and HIT biomes using SST (Fig. 8c), the HIL biome using SST or wind speed (Fig. 8f) and the SUS biome using Si, pCO_2 and chl (Fig. 8b, e, f). Thus, the combination of pCO_2 and SST appears to be the best environmental variable combination to discern individual biomes. Projecting biomes into the two-dimensional pCO_2 and SST space (Fig. A.19), shows a similar pattern as depicted in phytoplankton-based NMDS projections of Fig. 4 (Section 2.6.1). Moreover, we find that the separation between the high latitudes (HIL and HIT), the Pacific equatorial region (PEU), and the remaining low latitude biomes (TRP, WIS, SUS, MTR, SMN) is well reflected by the differences in pCO_2 and SST values found between the biomes. As is the case in Section 3.3 (cf. Fig. 4b), the overlap between the TRP, SUS, and MTR biomes and the overlap between the WIS and SMN biomes is readily observable in the two-dimensional pCO_2 and SST space (Fig. A.19).

In summary, the combination of pCO_2 and SST can be used to clearly distinguish between the HIL, HIT, and PEU biomes from all other biomes, and to a lesser extent to discern between the MTR and SUS biome. Additionally, the similarity between phytoplankton-based NMDS projections and projections in the pCO_2-SST space indicates that phytoplankton-based biomes reflect patterns in some of the abiotic drivers underlying species' distributions.

4. Discussion

Using biogeographic projections rooted in an extensive set of species observations, together with a clustering method, reveals global biome boundaries for the ocean's primary producers, which represent the most basal component of marine ecosystem structure. We show that the ocean's autotrophic species composition self-organize into high-latitude biomes (HIL, HIT), and low latitude biomes, such as the Pacific equatorial region (PEU) and two tropical biomes (TRP, and MTR). We also find dynamic ocean biome components, such as the seasonally alternating subtropical biomes (WIS and SUS). Our estimates are objective and data driven, as they omit the geographic location and time of phytoplankton presence projections during both the training and clustering phase. Furthermore, the spatial coherence of the individual biomes indicates that our approach generates valid global bioregionalizations (Fendereski et al., 2014).

4.1. Are phytoplankton-based ocean biomes consistent with recent ocean partitionings?

Recent studies have partitioned the ocean based on biogeochemical and physical criteria (Sarmiento et al., 2004; Hardman-Mountford et al., 2008; Fay and McKinley, 2014; Reygondeau et al., 2013), individual taxa (Reygondeau et al., 2011; Kulbicki et al., 2013; Keith et al., 2013), or plankton functional groups detected by remote sensing (De Monte et al., 2013). In general, our open-ocean clustering leads to biomes that are consistent with many features of previous partitionings, such as the clear separation between high and low latitudes (e.g. Reygondeau et al., 2011; Zhao et al., 2019). We also confirm the prominent biogeographic unit near the equator in the Pacific Ocean reported previously (e.g.

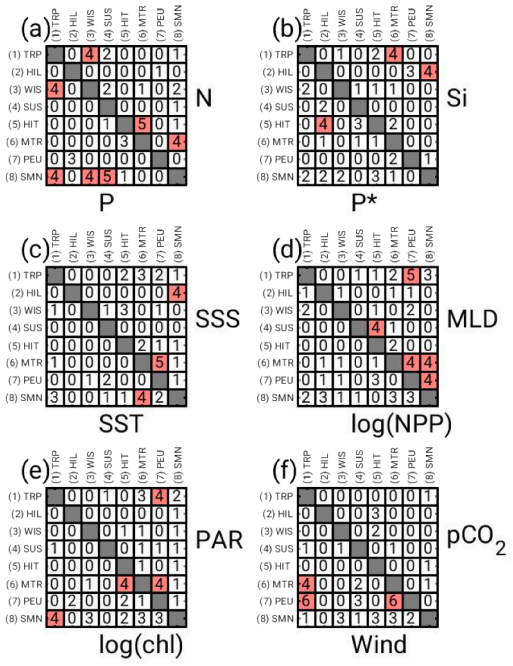


Fig. 8. Results of the Tukey-Kramer multiple comparison test (Kramer, 1956), summarizing the overall significant pairwise differences between biomes with respect to monthly environmental conditions. Each upper and lower triangular matrix, separated by a diagonal (gray) denotes a different environmental factor. Values denote the number of months where the monthly environmental conditions of two biomes were not significantly different from each other at $p \leqslant 0.01$. Biomes that are not significantly different from each other in a specific environmental conditions for more than three months are highlighted in red. Subfigure (a) contains the results for nitrate (N) and phosphate concentration (P), (b) for silicate (Si) and excess phosphate/nitrate (P*/N*), (c) for sea surface salinity (SSS) and sea surface temperature (SST), (d) for mixed layer depth (MLD) and net primary production (NPP), (e) for photosynthetic active radiation (PAR) and chlorophyll-a concentration (chl), and (f) for sea surface CO_2 partial pressure (pCO_2) and wind speed (wind). The units of the environmental variables are: N, P, Si and P*/N* in $\frac{\mu mol}{L}$; SSS in *PSU*; SST in $^{\circ}C$; MLD in m; NPP in $\frac{mg\,C}{m^2 d}$; PAR in $\frac{Einstein}{m^2 d}$; chl in $\frac{mg}{m^3}$; pCO_2 in μatm; wind in $\frac{m}{s}$. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Oliver and Irwin, 2008; Spalding et al., 2012; Fay and McKinley, 2014), and the clear separation between the Arabian Sea and the rest of the Indian Ocean (e.g. Reygondeau et al., 2011; Spalding et al., 2012). Particularly in regions of high environmental stability (PEU, TRP) or at high latitudes (HIL, HIT), our ocean biomes also strikingly align with partitionings based on biogeochemical or climatic criteria (e.g. Sarmiento et al., 2004; Oliver and Irwin, 2008; Fay and McKinley, 2014; Zhao et al., 2019). For instance, the TRP biome covers roughly the same geographical location as the subtropical seasonal biome in Fay and McKinley (2014), the subtropical ecosystem in Zhao et al. (2019), and oligotrophic provinces defined by Oliver and Irwin (2008). Moreover, our PEU biome roughly corresponds to the equatorial biome or province in Fay and McKinley (2014).

Biogeochemical partitionings of the ocean have usually resulted in a higher number of biomes or provinces compared to our phytoplankton-based partitioning. For instance, Oliver and Irwin (2008) reported 81 provinces based on monthly climatologies of water-leaving radiance and SST, with 17 provinces covering most of the ocean's surface area. Alternatively, Fay and McKinley (2014) defined 17 biomes based on multi-year climatologies of SST, chl-a concentration, maximum mixed layer depth, and sea ice cover. In both cases, the high latitudes of each hemisphere were separated into at least three provinces or biomes, and appeared to be nested within the geographic position of our HIL biome. Another example of biogeochemical provinces nested within one of our biomes are the five oligotrophic provinces defined in Oliver and Irwin (2008) and Irwin and Oliver (2009), which are located within the TRP biome. The observation that multiple provinces are nested within one of our biomes might be linked to the broad environmental niches of phytoplankton species (Righetti et al., 2019a; Brun et al., 2015), and is supported by previous evidence that biogeochemically-based provinces overlap in their biological properties (e.g. chl, and bacterial, and heterotrophic biomass; Vichi et al., 2011).

The dissimilarities between previous biogeochemical partitionings and our biomes might be explained by the fact that many (e.g. Sarmiento et al., 2004; Fay and McKinley, 2014) biogeochemical partitionings are based on annual or multi-year climatologies (Kavanaugh et al., 2016), and thus are not able to characterize the dynamic ocean ecosystem at monthly resolution (Hardman-Mountford et al., 2008), as in this study. The mismatch between other biogeochemical partitionings and the major Longhurst biomes might be linked to the temporal resolution of the data used for their definition (Kavanaugh et al., 2016), or the a priori fixed number of biomes (Devred et al., 2007). However, other biogeochemical partitionings such as those of Oliver and Irwin (2008) and Irwin and Oliver (2009), or the regional partitioning of Kavanaugh et al. (2014) included monthly-resolved data fields, which resemble our partitioning better. For instance, the basin scale partitioning of the North Pacific presented in Kavanaugh et al. (2014), defines seasonally alternating subtropical provinces similar to the SUS, and WIS biomes. In addition, in partitionings based on monthly input fields (SST, and ocean color data; Oliver and Irwin, 2008; Irwin et al., 2012), a province analogous to our PEU biome has been found. This stands in contrast to partitionings based on annual climatologies (e.g. Sarmiento et al., 2004; Fay and McKinley, 2014), where a partitioning analogous to the PEU biome emerges across all ocean basins.

The initial Longhurst partitioning defines four major, widely used open ocean biomes, i.e. a Polar, Westerlies, Trade wind, and Coastal biome using a mixture of physical variables (e.g. temperature, salinity, turbulence; Longhurst, 1995). The spatial distribution of these four major biomes corresponds partially to the spatial distribution of our biomes on an annual scale. The HIL biome corresponds to the Polar biome and parts of the Trade wind biome, the HIT biome corresponds to the equatorward part of the Westerlies biome in the North Atlantic Ocean and the Southern Hemisphere, and the remaining biomes to the Trade wind biome. Thus, unlike other biogeochemical partitionings mentioned above, our biomes are nested within the four major Longhurst biomes.

While our biomes at lower latitudes are nested within the major Longhurst biomes, the 57 biogeochemical provinces proposed by Longhurst (1995) are to a first approximation nested within our seven annual biomes. We established a many-to-one correspondence between Longhurst provinces and our annual biomes using the maximum pixel-wise overlap between a province and one of our biomes, i.e. a province can only correspond to one biome, but one biome can have multiple provinces. We found such a correspondence between Longhurst provinces and our biomes in 75.2% of the open ocean 1°-pixels assessed by our study (Fig. A.20a). Largest number of provinces were nested within the HIL (20) and TRP (17) biomes (i.e. within the biomes with the largest area coverage). The lowest correspondence between our biomes and Longhurst provinces occurred in regions where the WIS, HIT, and MTR biomes are located (Fig. A.20b). These regions are subject to seasonal changes in biome association and thus in phytoplankton community compositions (Fig. 3). The overall spatial correspondence between biomes and provinces suggests that the criteria chosen to define the biogeochemical provinces reflect general aspects of differences in the biogeography of phytoplankton species. This is likely linked to the bottom-up nature of both approaches. The spatial mismatch between our partitioning and the biogeochemical provinces can be explained by

the fact that the provinces do not take into account the highly dynamic nature of phytoplankton communities at the seasonal scale. The observation that the Longhurst provinces appear to be nested within our biomes indicates that we require a lower number of divisions to characterize phytoplankton biogeography compared to divisions based on the variation in ocean biogeochemistry, and certain bulk biological and zoogeographic features taken into account in Longhurst's approach, which might be linked to the large ecological niches of phytoplankton species (Brun et al., 2015; Righetti et al., 2019a).

Following the pioneering work by Tittensor et al. (2010), which took into account biological distribution data on a range of marine animals such as fish, sharks, and squids, a recent study by Costello et al. (2017) conducted a first biological clustering on presence-absence data patterns of pelagic and benthic species spanning eleven different phyla, with a large emphasis on heterotrophs. The partitioning of the ocean into 30 bioregions by Costello et al. (2017) are approximately nested within our seven annual biomes (Fig. A.20c). The correspondence between our biomes and those defined in Costello et al. (2017) amounts to 65.1% of the open ocean 1°-pixels on the annual scale. Eight out of the 30 bioregions of Costello et al. (2017) are nested within our TRP biome, and seven within our HIL biome. The poorest overlap between the two sets of biomes was found for regions where the WIS, HIT, and MTR biomes are located, as well as in the Atlantic Ocean, which, as previously mentioned, are regions characterized by seasonal changes in phytoplankton species communities (Fig. A.20d). The overall spatial correspondence between our biomes and the regions defined by different heterotrophic taxa suggests that factors controlling the presence of phytoplankton communities also control the higher trophic levels in the ocean, which suggests that similar spatial patterns exist across taxonomic groups (Spalding et al., 2012). However, the fact that many biomes based on the biogeography of heterotrophs are nested within our biomes, indicates that biological regionalization may increase with trophic level (Villarino et al., 2018). The observed nestedness of heterotrophs-based biomes within autotrophs-based ones is on one hand likely linked to the high degree of endemicity in the regions defined by Costello et al. (2017), and on the other hand to the large ecological niches of phytoplankton species (Brun et al., 2015; Righetti et al., 2019a).

Overall, our partitioning reduces the number of biogeographical units needed to characterize the open ocean. The nestedness of bioregions across trophic levels within our biologically-based biomes suggests that the open ocean flora is a suitable candidate for the partitioning of the open ocean into large ecologically meaningful units, as has been similarly discussed for the terrestrial realm (Bailey, 1998).

4.2. Can our biomes be characterized using a subset of species?

As a first characterization of our biomes, we explored their species richness, community overlap, top 100 most wide-spread/dominant species, indicator species and species co-occurrences. Not all biomes could be characterized unequivocally using their species richness (Section 3.2). Rather than showing a clear differentiation between biomes, the median and interquartile ranges in monthly species richness indicated differences at the scale of larger regions such as between high latitudes and low latitudes. The high latitudes and low latitudes displayed a relatively high versus low median species richness (Table 4). While the latitudinal difference in species richness is in coarse agreement with the equator-to-pole diversity decline described in Righetti et al. (2019a), the gradient is less steep at biome resolution. This is because our high latitude biome (HIL) has a wide latitudinal coverage, starting above 35° latitude, which integrates many species. Our tropical biomes PEU and MTR, instead, cover smaller latitudinal ranges.

In terms of species community composition (Fig. 4b), we found relatively little overlap in the species communities between the HIL, WIS, HIT, PEU, and SMN biomes. Globally, the species compositions of all 1°-pixels within a biome were more similar for biomes located at higher latitudes than at lower latitudes, i.e. the within biome difference in species compositions decreased with latitude. The relatively unique species communities of the HIL and HIT biomes may be a result of marked seasonality occurring from 35° to 60° latitude associated with elevated wind stress, turbulence, and seasonal light limitation that impose an ecological filter for species with certain traits and suppress the monthly active richness (Righetti et al., 2019a). Seasonal increases in the vertical mixing during winter at mid-latitudes, and the associated increase in nutrient supply to the surface layers (Cabré et al., 2016), result in environmental conditions favoring the phytoplankton communities of the WIS biome. Similarly, seasonal variations caused by monsoonal winds (Schott et al., 2009) and the associated changes in precipitation, and enhancement of coastal and open ocean upwelling (Sreeush et al., 2018) result in favorable conditions for the phytoplankton communities of the SMN biome. The unique, and high spatio-temporal variability of the tropical Pacific Ocean (Smith et al., 2019) likely results in a unique combination of environmental conditions favored by the phytoplankton community of the PEU biome. Thus, the unique communities of the HIL, HIT, WIS, SMN and PEU biomes appear to be linked to unique combinations of environmental factors, which are suitable for their respective communities.

Alternatively, we found substantial overlap in the communities of the TRP, SUS, and MTR biomes, where the absolute majority of neurons associated with the TRP and SUS biomes were within the envelope spanned by the phytoplankton communities of the MTR biome. Yet, the communities of the TRP and SUS biomes did not show substantial overlap. The overlap in species communities between the TRP and SUS biomes with the MTR biome suggests that the species communities of the MTR biome represents an intermediate composition with strong overlaps (Table A.15). Thus, with the exception of the MTR biome, we found that each biome hosts characteristic phytoplankton communities despite the large ecological niches of phytoplankton and the overlap in species communities between biomes.

For the dominant species (Section 3.4), we found similarities in their identity across biomes. The species found among the dominant species in most of our biomes are known to have cosmopolitan distributions (e.g. *Tripos extensus* or *Prochlorococcus*; Dodge and Marshall, 1994; Partensky et al., 1999). However, we were able to associate a few species among the top 100 most wide-spread species with in one or two biomes. For instance, the red-tide forming dinoflagellate *Lingulodinium polyedra* (Moorthi et al., 2006) dominates in biomes located near the Equator (MTR and PEU), or the haptophyte *Phaeocystis antarctica*, found in the HIL biome has its highest biomass in higher latitudes (HIL) (Vogt et al., 2012). The differences in the species dominance patterns between biomes suggest that we may be able to distinguish and characterize biomes based on a subset of its overall diversity using easily identifiable species, i.e. indicator species. Indicator species such as the three-spined stickleback *Gasterosteus aculeatus* (Lai et al., 2015), or the worm *Tubifex tubifex* (Vidal and Horne, 2003) are used as model organisms to monitor ecosystem health, and the impact of climate change or anthropogenic pollutants on local to regional communities (Angermeier and Karr, 2019). Indicator species that effectively characterize each of our biomes could be used to monitor ecosystem health at broad scales, or define future sampling efforts. Although we were able to associate certain species in the top 100 most wide-spread species with individual biomes, their area coverage remained relatively high in most other biomes. Thus, with our definition of indicator species, i.e. species that are wide-spread in one biome but rare in all other biomes, we were not able to identify indicator species for any of our biomes. This was likely due to the generally wide distributions and ecological niches of phytoplankton species (Brun et al., 2015; Righetti et al., 2019a). and strong dispersal ability (Cermeño and Falkowski, 2009; Whittaker and Rynearson, 2017). Changes in ocean properties (e.g. SST, nutrients, pH; Doney, 2010; Gruber, 2011; IPCC, 2013) over the past few decades may additionally have contributed to shifts in species' distributions and allowed some of the species to transcend biogeographic boundaries (Jönsson and

Watson, 2016; Thomas et al., 2012; Barton et al., 2016).

While none of our biomes could be identified based on single, common indicator species, we were able to find significant species pairs for all biomes except the SMN biome (Fig. 6), and characteristic species co-occurrences for the TRP, HIL, WIS and SUS biomes. In particular, for regions of low spatial and temporal variability (e.g. in the TRP, HIL, PEU biomes) we identified species pairs that may be used for monitoring purposes, and potentially relate to biome-wide ecosystem, ecological important species.

Biome definitions based on abiotic factors may represent a viable alternative, where biological information is limited or unavailable (Vichi et al., 2011; Kavanaugh et al., 2014). These abiotic factors may translate into phytoplankton biogeographic biomes via bottom-up effects on species' distributions (e.g. temperature, light availability, nutrients; Cermeño and Falkowski, 2009; Boyce et al., 2010; Brun et al., 2015; Righetti et al., 2019a).

Overall, SST and pCO_2 provided a suitable combination of environmental factors to differentiate each biome from all other biomes. SST and pCO_2, each were able to differentiate between some of our biomes. While temperature has been the potential key determinant of marine species distributions at broad scales (Tittensor et al., 2010; Righetti et al., 2019a), especially in plankton where physiological thermal tolerances may match species' latitudinal ranges (Beaugrand et al., 2008), pCO2 may serve as a proxy for the balance between nutrient inputs from deep ocean waters into surface surface waters through vertical mixing, and their plankton-driven removal (Sarmiento and Gruber, 2006). Using this premise, our biome-averaged pCO_2 values suggest that the PEU biome is associated with species communities do not reduce surface nutrient concentrations sufficiently fast relative to the vertical input of nutrients from deep ocean waters. Together, we propose temperature and pCO_2 as two factors useful for monitoring of marine biome boundaries in the presence of biological data limitation.

4.3. Caveats of our approach

We discuss potential caveats in our analysis in two steps. First, we address species' biogeographic projections obtained from the niche models that form the foundation for our clustering work. Second, we address the robustness of our biomes to choices made during the clustering method developed.

The presence-absence patterns of the phytoplankton species in our dataset have originally been subjected to methodological choices (see Righetti et al., 2019a). The SDMs developed in Righetti et al. (2019a) have been specifically designed to optimize the robustness of modelled distribution patterns in the presence of sparse, methodologically heterogeneous, and unevenly distributed field data of short-lived marine phytoplankton. Here, we discuss uncertainties, which significantly affect our results in terms of biome location and extent, including statistical algorithm choice. The latter comprising up to 73% of the uncertainty in the projection of future species habitat extent (Buisson et al., 2010; Benedetti et al., 2018). We compare results on biome distribution derived from GAM, with results derived from GLM- and RF-based species projections, and we examine uncertainty induced by predictor variable choice. Our biomes are mostly robust to the uncertainty in species presence patterns associated with algorithm choice (A.5; Fig. A.13) as well as to the uncertainties due to predictor variable uncertainty in species' original presence-absence projections (A.6).

In the original phytoplankton data, presence records of species have been aggregated to monthly climatological, 1° latitude × 1° longitude resolution. This aggregation, which treated species records (mostly originating from the decades 1950 to 2010 Righetti et al., 2019a) at monthly scales, has been an essential step to integrate the sparse data, and to balance inter-pixel differences in original sampling efforts. We used the same data (Righetti et al., 2019a) to diagnose the effect of this binning on emerging species co-occurrences, and find that 22 out of the 72 significant species pairs were indeed observed in the original dataset at the 1° × 1° scale, during same months. An additional 14, and 38 species pairs, co-occurred in the original dataset during the same month of different years, at the 1° × 1° scale, or at the biome scale, respectively. Six additional species pairs, co-occurred at the biome scale during different months and years. Overall, around one third of the significant species co-occurrences identified in our study were indeed observed at the 1°-pixel scale. While this analysis confirms that at least one-third of the significant species pairs indeed occurs in the field, we note that individual samples of phytoplankton in the original presence dataset are highly incomplete with regards to species detection (Righetti et al., 2019b; Righetti et al., 2019a). Thus, a lack of presence of species in the observational data at monthly 1° resolution, very likely reflects the selectivity of sampling efforts, rather than being evidence of true exclusion patterns of any given species or species associations. The SDMs have been specifically implemented to infer species' presence distributions in space and time from the limited sampling, in the first place. Sampling techniques that may detect most species present at any location and point in time, such as metagenomic sequencing (e.g. De Vargas et al., 2015), could further verify the total associations diagnosed in this work.

Indeed, recent analyses of metagenomic data from the Tara Oceans expeditions support our findings. Network analysis of co-located metagenomic observation revealed a clear differentiation in species associations between high latitudes and low latitudes, with higher species connectivity at higher latitudes, and a biome structure, similar to ours. These results based on independent data support that biomes at high latitudes may house characteristic species associations (Chaffron et al., 2020).

Other potential caveats relate to the spatial and temporal resolution of the SDM based presence-absence projections. The 1° resolution, at monthly scales cannot resolve fine scale variability in phytoplankton communities (Buitenhuis et al., 2013). Mesoscale and sub-mesoscale features modulate local phytoplankton community compositions in situ (Lévy et al., 2018), and short lived phytoplankton species (Padisák, 1994) exhibit rapid response times to perturbations (Kavanaugh et al., 2016). For instance, periods of pulsed phytoplankton growth (e.g. spring blooms) often occur on timescales from one to two weeks (Racault et al., 2012). Thus, our biomes do not consider small-scale changes in phytoplankton species communities (e.g. due to mesoscale forcing; Cotti-Rausch et al., 2016). However, such small scale phytoplankton variability might be primarily important in localized contexts and competitive interactions become negligible for the description of species associations at larger scales (Schimel, 1995). Thus, our biomes reflect large-scale differences in aggregated structures of species communities, which are composed of a multitude of small local subsets of larger species pools (Smith et al., 2005).

While our analysis includes more than 530 species, covering all major functional and taxonomic phytoplankton groups, many more phytoplankton species exist (ranging between 4′530 and 16′940; Falkowski et al., 2004; De Vargas et al., 2015). Thus, our analysis includes at most 12% of named marine phytoplankton species. Given that common sampling techniques in phytoplankton rely on small volumes of seawater, which incompletely detect species (Cermeño et al., 2014), the species included in our biomes are biased toward frequent, known, and abundant species of the ocean. However, small fractions of the total species diversity have been shown to account for a majority of phytoplankton abundance or biomass within different phytoplankton groups: 16 species of coccolithophore out of 195 or 43 diatom species out of 552, accounted for 75% of coccolithophore abundance (O'Brien et al., 2013) or 90% of total diatom biomass (Leblanc et al., 2012), respectively. Thus, although the original may not account for many rare species (Ser-Giacomi et al., 2018), the species considered likely represent abundant species, which support the most biogeographical patterns and ecologically relevant processes (Lyons et al., 2005; Cermeño et al., 2013; Ser-Giacomi et al., 2018). Similar to terrestrial biomes (e.g. Higgins et al., 2016; Silva de Miranda et al., 2018), biomes based on the most abundant

species likely capture differences in ecosystem function as well.

Potential circularities between the initial use of environmental predictor variables in SDMs and the analysis of environmental conditions at the level of biomes cannot be fully ruled out in our approach. However, the use of a broad initial variable pool and the randomized selection of small subsets of variables in each SDM, was designed to prevent predetermination or bias due to variable choice in the SDMs and the resulting biome boundaries (Righetti et al., 2019a). This variable selection approach also avoided over-representation of single predictors in SDMs by controlling for balanced predictor contribution in individual species' models. By integrating the spatial projections of hundreds of species, our biome boundaries integrated broad predictive power and emerged as higher-level aggregate properties of predicted species (Righetti et al., 2019a).

Caveats linked to our clustering methodology arise from choices in the different training parameters (i.e. number of neurons, number of epochs, distance metric) and the clustering of the neurons. Most likely, using different training parameters results in different clusters, and thus different biomes. However, we objectively selected the number of both the neurons and training epochs using best-practice recommendations (Vesanto and Alhoniemi, 2000), and we relied on previous findings that compared the behavior of different distance metrics with respect to high-dimensional data (Aggarwal et al., 2001). For the clustering of neurons, we decided to use the clustering with the highest robustness to information loss during training. Nevertheless, to reinforce the robustness of our choice, one could compare different clustering methods (k-means, fuzzy clustering; Jain et al., 1999) and choose the optimal number of clusters based on the overall agreement between the different methodologies.

5. Conclusion

We found that the open ocean can be partitioned into eight biomes based on monthly phytoplankton species presence-absence projections, with the largest differences in phytoplankton communities between biomes occurring across three major regions: the high latitudes, the Pacific equatorial region, and the low latitudes. A further partitioning of these three major regions is linked to subtler differences in phytoplankton species communities, including monthly shifts in the species communities throughout the year. We found that biomes can be discriminated based on a subset of their species diversity and associations between species. The inclusion of a larger set of species into future classifications, based on metagenomic data (De Vargas et al., 2015) may reveal further characteristic biological players and possibly unique biotic network associations (Lima-Mendez et al., 2015; Vincent and Bowler, 2020), which may enable in situ biome extent monitoring based on a subset of the full diversity and specific indicator species or their genes (e.g. Malviya et al., 2016). Our results further show a close association of plankton community structure with environmental drivers, which may guide the development of next-generation satellite products related to biodiversity and ocean health.

In our work, we took into account a subset of the full species existing, including preferentially the most common and well-detected taxa, which led to biome-wide patterns in terms of characteristic co-occurrences of species. The advent of metagenomic (Biller et al., 2018) and imaging data (Guidi et al., 2016), which resolve a greater variety of phytoplankton diversity, may promote future regionalizations of the open ocean that could inform monitoring plans and guide the development of bioindicators for ecosystem function and service provision. Additionally, with the higher temporal and spatial resolution of metagenomic and imaging data, this study could be extended to include dynamic coastal environments (Davis et al., 2007) or shorter timescales, which match the timescales of interactions between marine consumers, primary producers, and the environment (Kavanaugh et al., 2016). Such a high spatio-temporal resolution would allow for a dynamic management of ecosystem services, and thus a rapid response of marine resource users to ecosystem changes (Maxwell et al., 2015).

Ocean partitionings based on the spatial structure of marine ecosystems, rather than jurisdictional boundaries or physico-chemical properties (Crowder, 2006; Kavanaugh et al., 2016) divide the ocean into practical ecosystem management, conservation, and monitoring units (Spalding et al., 2007; Caldow et al., 2015; Breece et al., 2016; Lewison et al., 2015). Such partitionings characterize the spatio-temporal patterns of open ocean community structure and function (Kavanaugh et al., 2016), and can be used for marine ecosystem conservation purposes (Briones et al., 2009), or to investigate the responses of marine ecosystems to global environmental change and human activity (Gruber, 2011; Doney et al., 2011).

In this study we have provided a comprehensive global open ocean partitioning, which separates and summarizes important processes that govern phytoplankton community structure (Kavanaugh et al., 2014; Briones et al., 2009), and divides the open ocean into units whose consideration may help us to monitor, mitigate, or prevent potential present or future decline in marine ecosystem health (Halpern et al., 2014).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

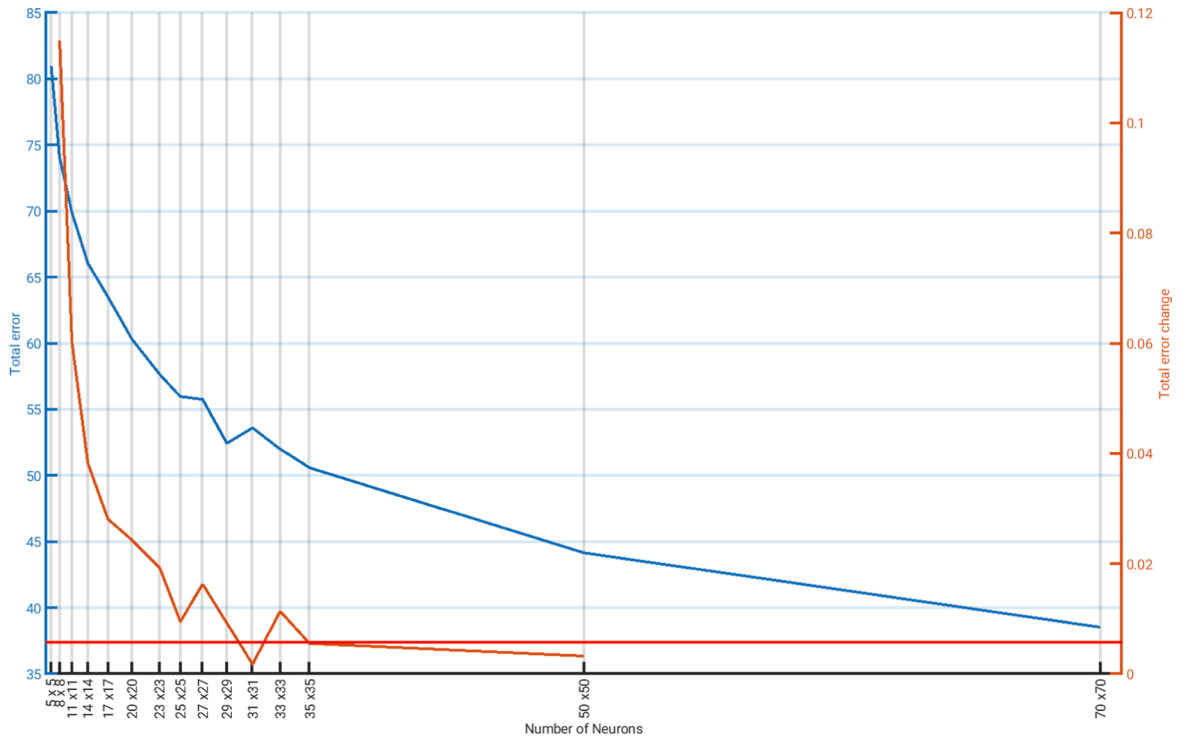
Acknowledgments

We thank N. Gruber and the Environmental Physics group for analytical and visualization advice. We thank the three anonymous reviewers for their fruitful and constructive comments that significantly improved the quality of our work. **Funding**: This research was supported by ETH Zürich under grant 175787 - 'X-EBUS: Extreme Ocean Weather Events and their Role for Ocean Biogeochemistry and Ecosystems in Eastern Boundary Upwelling Systems' (SNF). **Author contributions**: M.V. and U.H conceptualized the project. D.R. provided the phytoplankton data. D.R. and U.H. prepared the data. U.H. performed the statistical analysis with inputs by F.B. and M.V. U.H. prepared the data visualizations. U.H. wrote the first draft with substantial inputs by M.V. and D.R. All authors interpreted the results. All authors reviewed the manuscript. M.V. supervised the project. **Competing interests**: The authors declare that they have no competing interests. **Data and materials availability**: Phytoplankton data used for this project stems from Righetti et al. (2019a). Matlab scripts are freely available upon request.

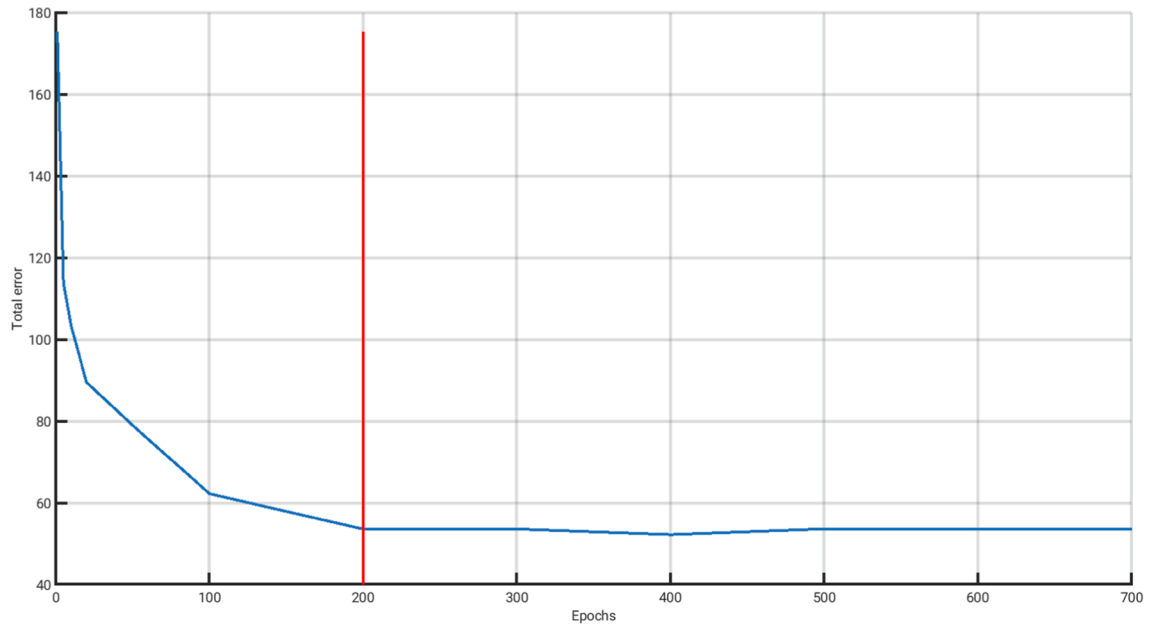
Appendix A

A.1. Self-organizing map

The quantization error (QE) (Kohonen, 2001), and the topological error (TE) (Kiviluoto, 1996) averaged over the whole mapping are defined as follows (Kohonen, 2001):



(a)



(b)

Fig. A.9. (a) Total error, and total error change as a function of the number of neurons. The neuron-lattice of 31 × 31 was chosen as an optimal amount of neurons. (b) Total error as a function of epochs, i.e. the number of times the input data was presented to the neural network, while using the optimal amount of neurons. The optimal number of epochs was chosen to be 200.

$$QE = \frac{1}{n} \cdot \sum_{i=1}^{n} d(v_i, w_{c,i}), \tag{A.1}$$

$$TE = \frac{1}{n} \cdot \sum_{i=1}^{n} u(v_i), \tag{A.2}$$

$$u(v_i) = \begin{cases} 1, & \text{best-and second-BMU are non-adjacent} \\ 0, & \text{otherwise} \end{cases}. \tag{A.3}$$

The distance function (e.g. Euclidean or Manhattan distance) is denoted by $d(v, w)$, n is the number of observations, v_i is the phytoplankton species presence projection, $W_{c,i}$ is the best matching neuron to the phytoplankton species presence projection i, and $i = 1, 2, 3, \ldots, n$.

A.2. Calculating the cluster centroid

Two methods for determining the value of a cluster mean, i.e. the centroid of a cluster were tested. Using HAC, we grouped 961 trained neurons into clusters. First, we calculated the centroid of a cluster as the mean of the trained neurons that were assigned to a cluster considering the number of times they appeared across all months (i.e. frequency weighted centroids). Second, we calculated the centroid of clusters as the mean of the trained neurons that were assigned to a cluster without considering their occurrence frequency, i.e. all member of a cluster had the same weight for the calculation of the centroid. We performed this test for two up to twenty clusters, and used the mean difference between the centroid to the assigned observation as quality measure (see Fig. A.10). We found that using the frequency weighted centroids resulted in a better representation of the phytoplankton species presence projections for all granularity levels of clustering.

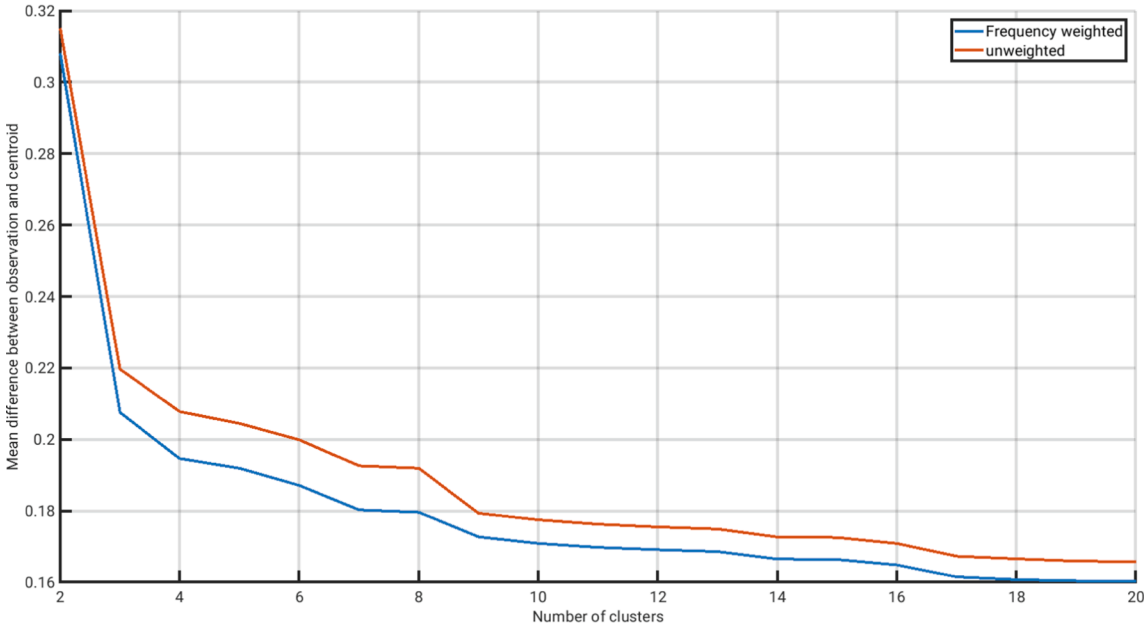


Fig. A.10. Comparison of two methods to calculate a cluster's centroid. The weighted version uses the occurrence frequency of the members of a clusters as a weighting factor, the unweighted version only uses the individual members. The comparison metric was calculated using the absolute difference between an observation and the associated centroid, calculating the mean over all species, and over all observations.

A.3. Environmental data

Table A.5.

Table A.5
Environmental parameters considered as potential determinants of the phytoplankton community compositions found in our biomes.

Abbr.	Parameter	Unit	Source
SST	Sea Surface Temperature	$^{\circ}C$	World Ocean Atlas 2013, Locarnini et al. (2013)
N	Nitrate concentration	$\frac{\mu mol}{L}$	World Ocean Atlas 2013, Garcia et al. (2013)
P	Phosphate concentration	$\frac{\mu mol}{L}$	World Ocean Atlas 2013, Garcia et al. (2013)
Si	Silicic acid concentration	$\frac{\mu mol}{L}$	World Ocean Atlas 2013, Garcia et al. (2013)
SSS	Sea Surface Salinity	PSU	World Ocean Atlas 2013, Zweng et al. (2013)
PAR	Photosynthetic Active Radiation	$\frac{Einstein}{m^2 d}$	Sea-Viewing Wide Field-of-View Sensor (NASA, 2018b)
chl	Chlorophyll-a concentration	$\frac{mg}{m^3}$	Sea-Viewing Wide Field-of-View Sensor (NASA, 2018a)
MLD	Mixed Layer Depth	m	Montégut (2004)
NPP	Net Primary Production	$\frac{mgC}{m^2 d}$	Behrenfeld and Falkowski (1997);Westberry et al. (2008); Ocean Productivity (2017)
wind	Sea surface wind stress	$\frac{m}{s}$	Atlas et al. (2011)
pCO_2	Sea surface CO_2 partial pressur	μatm	Landschützer et al. (2014)

A.4. Principal component analysis

Table A.6.

Table A.6
Loadings of the eight principal components (PC) found with Kaiser's rule (Wilks, 2011). The eight principal components were used to reduce the dimensionality prior to the hierarchical agglomerative clustering. Species are sorted in alphabetic order within classes or phyla.

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Achnanthes longipes	0.00	−0.02	0.01	−0.01	−0.03	0.01	0.02	0.00
Actinocyclus curvatulus	0.01	−0.07	0.07	−0.01	−0.03	−0.03	−0.06	−0.08
Actinocyclus octonarius	0.04	0.00	0.03	0.01	0.10	0.08	−0.07	0.00
Actinoptychus octonarius	−0.02	−0.02	0.02	0.02	0.00	0.00	0.09	−0.02
Actinoptychus senarius	0.06	−0.01	0.05	−0.11	0.02	−0.02	0.08	−0.04
Actinoptychus splendens	0.01	−0.05	0.02	−0.03	−0.04	0.01	0.01	−0.02
Amphiprora gigantea	0.03	0.08	0.00	−0.03	−0.02	0.02	0.05	0.05
Asterionellopsis glacialis	0.01	−0.03	0.05	−0.08	0.02	−0.01	0.00	−0.02
Asterolampra marylandica	0.06	0.09	0.02	0.00	−0.01	0.02	0.05	−0.03
Asteromphalus brookei	0.00	−0.03	0.01	−0.01	−0.04	0.01	0.02	0.00
Asteromphalus cleveanus	0.06	0.10	0.00	0.00	−0.05	0.01	−0.05	−0.02
Asteromphalus flabellatus	0.06	−0.05	0.08	−0.04	0.03	−0.02	−0.01	0.04
Asteromphalus heptactis	0.05	−0.02	0.09	−0.03	0.01	−0.01	0.05	0.04
Asteromphalus hyalinus	−0.01	0.00	0.02	0.04	0.03	−0.02	0.08	−0.04
Asteromphalus parvulus	−0.02	0.00	0.02	0.05	0.04	−0.03	0.10	−0.03
Asteromphalus robustus	0.01	0.02	0.07	0.03	−0.03	0.03	−0.04	−0.06
Asteroplanus karianus	−0.01	0.00	0.02	0.01	0.00	−0.01	−0.01	−0.03
Attheya septentrionalis	−0.01	0.00	0.02	0.01	0.02	−0.02	0.03	−0.02
Aulacoseira granulata	0.05	0.04	−0.06	−0.01	−0.01	−0.05	0.00	−0.04
Azpeitia nodulifera	0.07	0.03	0.00	−0.03	0.07	0.03	−0.01	−0.10
Bacteriastrum comosum	0.06	0.09	0.03	0.01	−0.02	0.06	0.01	0.01
Bacteriastrum delicatulum	0.06	0.04	0.04	0.01	−0.07	0.08	0.03	−0.07
Bacteriastrum elegans	0.05	0.10	0.02	0.02	−0.06	0.05	0.04	0.05
Bacteriastrum elongatum	0.06	0.08	0.07	0.01	−0.04	0.02	0.00	0.04
Bacteriastrum furcatum	0.07	0.07	0.05	0.00	0.00	0.00	−0.01	0.02
Bacteriastrum hyalinum	0.06	0.05	0.06	0.05	−0.07	−0.03	−0.03	0.00
Bacteriastrum mediterraneum	0.04	0.07	0.08	0.01	−0.05	0.03	0.04	0.06
Bacterosira bathyomphala	−0.02	0.00	0.03	0.04	0.03	−0.02	0.03	−0.06
Bellerochea malleus	0.00	−0.02	0.02	0.01	0.01	0.05	−0.09	−0.01
Biddulphia alternans	0.00	−0.01	0.00	0.00	0.00	0.01	−0.01	0.00
Cerataulina pelagica	0.07	0.01	0.01	−0.09	−0.01	0.01	0.01	−0.04
Chaetoceros aequatorialis	0.06	0.09	0.04	0.01	−0.05	−0.03	−0.04	0.01
Chaetoceros affinis	0.07	0.01	0.07	−0.04	−0.07	−0.02	0.00	0.05
Chaetoceros anastomosans	0.06	0.05	0.02	−0.05	−0.03	−0.02	−0.07	−0.03
Chaetoceros atlanticus	0.02	−0.07	0.12	0.06	0.02	−0.03	0.03	−0.09
Chaetoceros bacteriastroides	0.04	0.07	0.07	0.02	−0.03	−0.01	0.03	0.12
Chaetoceros borealis	−0.01	−0.01	0.01	0.01	0.01	0.02	−0.03	−0.02
Chaetoceros brevis	0.06	0.02	0.07	−0.06	−0.03	−0.06	−0.02	0.04
Chaetoceros bulbosus	−0.01	0.00	0.02	0.04	0.03	−0.02	0.09	−0.04
Chaetoceros castracanei	0.02	0.03	0.09	0.03	0.01	−0.02	0.02	0.02
Chaetoceros cinctus	0.00	0.00	0.00	0.00	0.00	0.01	−0.01	0.01
Chaetoceros coarctatus	0.07	−0.01	0.01	−0.08	0.00	0.01	0.04	−0.03
Chaetoceros compressus	0.07	0.01	0.00	−0.06	−0.03	0.04	0.04	−0.06
Chaetoceros concavicornis	−0.01	−0.03	0.07	0.04	−0.04	−0.01	−0.01	−0.11
Chaetoceros constrictus	0.00	−0.04	0.02	−0.01	−0.02	0.03	−0.02	−0.03
Chaetoceros convolutus	−0.01	−0.04	0.08	0.04	−0.05	−0.03	0.01	−0.13
Chaetoceros costatus	0.06	0.04	0.07	−0.01	−0.09	−0.01	0.02	0.03
Chaetoceros criophilus	−0.01	0.00	0.02	0.03	0.03	−0.02	0.06	−0.06
Chaetoceros curvisetus	0.07	0.02	0.00	−0.07	−0.06	0.02	0.02	−0.05
Chaetoceros dadayi	0.07	0.07	0.03	−0.01	0.00	−0.02	−0.01	−0.01
Chaetoceros danicus	0.06	0.00	0.04	−0.08	−0.02	0.11	0.07	−0.05
Chaetoceros debilis	0.02	−0.07	0.08	0.01	−0.07	−0.02	−0.03	−0.09
Chaetoceros decipiens	0.06	0.03	0.06	−0.03	−0.05	0.05	0.01	0.01
Chaetoceros densus	0.05	0.06	0.04	0.04	0.00	0.11	0.03	0.07
Chaetoceros diadema	0.03	0.08	0.08	0.02	−0.04	−0.04	0.01	0.09
Chaetoceros dichaeta	0.04	0.02	0.09	0.03	−0.05	−0.10	0.05	−0.02
Chaetoceros didymus	0.05	0.01	0.10	0.03	−0.07	−0.01	−0.02	0.04
Chaetoceros diversus	0.07	0.07	0.02	0.00	−0.01	0.04	0.01	−0.08
Chaetoceros eibenii	0.04	0.05	0.08	0.02	−0.07	0.07	0.04	0.00
Chaetoceros furcellatus	−0.02	0.01	0.06	0.03	0.00	−0.03	−0.01	−0.10
Chaetoceros gracilis	0.03	0.01	0.10	0.02	0.02	−0.03	−0.10	−0.01
Chaetoceros hyalochaetae	0.05	0.02	−0.05	−0.07	−0.02	0.00	−0.02	−0.01
Chaetoceros indicus	0.02	0.02	0.04	0.03	0.02	0.04	0.02	0.06
Chaetoceros laciniosus	0.06	0.05	0.00	−0.01	−0.02	−0.04	0.03	−0.02
Chaetoceros laevis	0.05	0.06	0.05	0.05	−0.05	−0.02	0.00	0.03
Chaetoceros lauderi	0.05	0.00	0.10	−0.04	−0.01	0.00	0.05	0.12
Chaetoceros lorenzianus	0.07	−0.01	0.05	−0.05	−0.03	−0.04	0.00	0.04
Chaetoceros messanensis	0.06	0.08	0.03	0.02	−0.05	0.03	−0.03	0.01

(continued on next page)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Chaetoceros neglectus	−0.02	0.01	0.02	0.04	0.04	−0.03	0.12	−0.02
Chaetoceros pacificus	0.03	0.05	0.04	0.02	−0.03	0.04	0.07	0.09
Chaetoceros paradoxus	0.03	0.08	0.08	−0.04	−0.04	−0.05	−0.06	0.03
Chaetoceros pelagicus	0.05	−0.03	−0.02	0.09	−0.02	0.01	0.14	0.03
Chaetoceros pendulus	0.06	0.06	0.01	−0.01	−0.06	−0.03	−0.05	−0.03
Chaetoceros peruvianus	0.07	0.00	0.01	−0.07	−0.02	0.00	0.04	−0.05
Chaetoceros phaeoceros	0.04	0.07	−0.07	−0.07	0.03	0.03	0.06	−0.07
Chaetoceros pseudocrinitus	0.04	0.03	0.03	0.05	0.01	0.08	0.05	0.07
Chaetoceros pseudocurvisetus	0.06	0.06	0.07	0.00	−0.03	0.03	−0.02	0.07
Chaetoceros pseudodichaetus	0.05	0.08	0.05	0.02	−0.04	0.05	0.06	0.05
Chaetoceros radicans	0.05	0.01	0.11	0.02	−0.05	−0.02	−0.01	−0.01
Chaetoceros rostratus	0.06	0.06	0.07	0.01	−0.05	0.01	0.02	0.05
Chaetoceros saltans	0.05	0.07	0.08	0.01	−0.01	−0.03	0.01	0.08
Chaetoceros seiracanthus	0.01	0.05	0.05	−0.05	0.00	0.03	0.00	0.00
Chaetoceros seriacanthus	0.01	0.03	0.06	−0.02	0.03	0.00	−0.03	0.02
Chaetoceros seychellarum	0.06	0.07	0.05	0.00	−0.03	0.05	0.05	0.02
Chaetoceros similis	−0.01	−0.02	0.01	0.02	0.00	0.00	0.03	−0.02
Chaetoceros simplex	0.01	0.08	0.08	−0.07	0.01	−0.01	−0.06	−0.01
Chaetoceros socialis	0.01	−0.03	0.04	0.00	−0.08	0.05	0.00	−0.05
Chaetoceros subsecundus	0.04	0.04	0.11	0.04	0.03	−0.04	−0.07	0.06
Chaetoceros teres	0.00	0.02	0.01	0.00	−0.03	0.06	−0.02	−0.03
Chaetoceros tetrastichon	0.06	0.04	0.03	−0.04	0.01	−0.05	−0.01	0.04
Chaetoceros tortissimus	0.04	0.04	0.09	0.02	0.03	0.05	0.03	0.04
Chaetoceros vanheurckii	0.06	0.08	0.04	0.02	−0.04	0.00	0.01	0.03
Chaetoceros wighamii	0.06	0.06	0.00	−0.04	−0.01	−0.03	0.04	0.03
Chaetoceros willei	0.06	0.07	0.06	0.02	−0.02	0.06	0.05	0.04
Climacodium biconcavum	0.06	0.09	0.03	0.01	−0.04	−0.02	−0.01	−0.02
Climacodium frauenfeldianum	0.07	0.05	0.00	−0.05	−0.02	−0.02	−0.03	−0.02
Climacosphenia moniligera	0.07	−0.01	0.02	−0.03	−0.04	−0.07	−0.02	0.05
Corethron hystrix	−0.01	−0.01	0.02	0.02	−0.01	0.00	−0.02	−0.05
Corethron pennatum	−0.01	0.02	0.13	0.08	0.05	−0.07	0.01	0.03
Coscinodiscus antarcticus	0.00	0.02	0.02	0.05	0.04	−0.01	0.11	0.01
Coscinodiscus asteromphalus	0.03	0.00	0.07	−0.06	0.07	−0.01	−0.06	−0.02
Coscinodiscus centralis	0.00	−0.05	0.02	−0.04	−0.04	0.01	0.02	−0.01
Coscinodiscus concinnus	−0.01	−0.01	0.01	0.02	0.01	0.01	−0.01	−0.01
Coscinodiscus curvatulus	−0.02	−0.02	0.03	0.04	0.01	−0.01	0.08	−0.05
Coscinodiscus gigas	0.04	−0.01	0.04	0.01	0.10	0.10	−0.05	0.00
Coscinodiscus granii	0.03	−0.04	0.06	−0.03	−0.03	0.01	0.08	0.11
Coscinodiscus marginatus	−0.01	−0.01	0.04	0.04	0.00	−0.02	0.01	−0.09
Coscinodiscus oculus-iridis	0.04	−0.01	0.05	0.05	0.13	0.03	−0.02	0.04
Coscinodiscus radiatus	0.05	−0.07	0.05	−0.05	0.07	0.03	0.02	−0.01
Coscinodiscus subbulliens	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Coscinodiscus wailesii	0.00	−0.01	0.01	0.01	0.00	0.01	−0.01	−0.02
Cylindrotheca fusiformis	−0.01	0.00	0.03	0.02	0.01	−0.01	−0.01	−0.07
Dactyliosolen antarcticus	−0.01	0.01	0.01	−0.01	0.01	0.00	0.01	0.04
Dactyliosolen fragilissimus	0.06	0.06	0.01	0.02	0.00	0.15	0.04	−0.04
Detonula confervacea	−0.01	−0.01	0.01	0.01	−0.01	0.01	0.00	−0.02
Detonula pumila	0.07	0.01	0.03	−0.06	0.01	0.05	0.03	−0.06
Diatoma rhombica	−0.02	0.00	0.02	0.04	0.03	−0.02	0.09	−0.04
Ditylum brightwellii	0.04	0.02	0.03	−0.15	0.08	0.00	0.02	−0.07
Ditylum sol	0.04	0.08	0.06	0.00	−0.06	−0.04	0.02	0.08
Entomoneis alata	0.01	−0.04	0.04	0.00	−0.03	0.01	−0.03	−0.05
Entomoneis paludosa	−0.02	0.00	0.02	−0.03	0.00	0.01	0.01	−0.06
Ephemera planamembranacea	−0.01	0.00	0.00	0.01	0.01	0.01	−0.02	0.01
Ethmodiscus gazellae	0.01	−0.07	0.00	−0.01	−0.06	0.00	−0.02	−0.03
Eucampia antarctica	−0.02	0.00	0.03	0.05	0.05	−0.03	0.10	−0.02
Eucampia cornuta	0.05	0.04	0.03	0.00	−0.05	−0.03	0.09	0.01
Eucampia zodiacus	0.06	0.01	0.08	−0.04	−0.04	0.03	0.00	−0.04
Eucampia zoodiacus	0.04	0.08	0.09	0.02	−0.06	0.00	−0.04	0.05
Eupyxidicula turris	0.00	−0.02	0.01	−0.02	−0.01	0.02	−0.03	−0.03
Fragilaria striatula	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.01
Fragilariopsis curta	−0.01	0.00	0.02	0.04	0.03	−0.02	0.08	−0.04
Fragilariopsis cylindrus	−0.02	0.00	0.02	0.05	0.04	−0.02	0.10	−0.04
Fragilariopsis doliolus	0.01	−0.03	0.04	0.00	−0.06	0.01	0.00	−0.03
Fragilariopsis kerguelensis	−0.02	0.00	0.02	0.05	0.04	−0.03	0.10	−0.03
Fragilariopsis obliquecostata	−0.01	0.00	0.02	0.04	0.03	−0.02	0.08	−0.04
Fragilariopsis oceanica	−0.01	−0.01	0.06	0.05	0.04	−0.06	−0.03	−0.08
Gossleriella tropica	0.06	0.01	0.02	0.01	−0.04	0.05	−0.05	−0.03
Grammatophora angulosa	0.00	−0.01	0.01	0.00	−0.01	0.01	0.00	−0.01
Grammatophora marina	0.00	−0.03	0.02	0.01	−0.02	0.01	0.03	−0.03
Grammatophora oceanica	0.01	−0.05	0.04	0.00	−0.05	−0.03	−0.05	−0.06

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Guinardia cylindrus	0.06	0.08	0.00	0.02	−0.01	−0.02	0.02	−0.09
Guinardia delicatula	0.06	−0.02	0.06	−0.07	0.10	0.09	0.01	0.02
Guinardia flaccida	0.06	0.04	−0.01	−0.01	−0.05	0.04	0.04	−0.05
Guinardia striata	0.05	0.03	0.06	−0.04	−0.07	0.03	0.00	0.12
Haslea wawrikae	0.06	0.08	0.02	0.01	−0.04	−0.01	−0.02	−0.04
Helicotheca tamesis	0.00	−0.03	0.01	−0.02	−0.04	0.01	0.04	0.00
Hemiaulus chinensis	0.07	0.06	0.03	0.00	−0.04	0.07	0.02	−0.05
Hemiaulus hauckii	0.07	0.05	−0.01	−0.03	−0.01	0.00	−0.02	−0.07
Hemiaulus membranaceus	0.07	0.06	0.07	0.02	−0.02	0.02	−0.03	0.01
Hemidiscus cuneiformis	0.01	−0.03	−0.01	0.02	−0.06	0.00	0.02	−0.02
Lauderia annulata	0.06	−0.01	0.05	−0.08	0.11	0.07	0.03	0.01
Leptocylindrus danicus	0.07	0.02	−0.02	−0.05	−0.04	0.01	−0.02	−0.07
Leptocylindrus minimus	0.06	0.06	0.07	0.00	−0.01	0.00	−0.07	0.00
Licmophora abbreviata	0.00	−0.05	0.02	−0.02	−0.05	0.02	0.01	−0.02
Licmophora lyngbyei	0.03	0.02	0.10	0.01	0.00	−0.07	−0.06	−0.07
Lioloma delicatulum	0.01	−0.09	0.00	−0.05	−0.04	−0.03	−0.01	−0.03
Lioloma pacificum	0.02	−0.09	0.02	−0.04	−0.03	−0.05	−0.02	−0.05
Lithodesmium undulatum	0.03	−0.04	0.07	−0.03	−0.03	0.07	−0.01	−0.04
Mastogloia rostrata	0.06	0.09	−0.02	0.00	−0.03	0.02	0.01	−0.06
Melosira borreri	0.02	0.04	0.06	0.01	−0.08	−0.04	0.00	0.01
Melosira moniliformis	0.00	−0.01	0.01	0.01	0.00	0.00	0.02	−0.02
Membraneis challengeri	0.02	0.09	0.03	0.07	−0.03	0.00	0.17	−0.09
Meuniera membranacea	0.06	0.08	0.05	0.02	−0.03	0.01	−0.02	0.01
Navicula pelagica	−0.01	0.00	0.02	0.03	0.01	−0.01	0.02	−0.04
Neocalyptrella robusta	0.05	−0.05	0.08	−0.06	0.05	0.00	0.02	0.03
Nitzschia bicapitata	0.05	0.00	−0.03	−0.07	−0.01	−0.04	0.00	−0.01
Nitzschia bilobata	0.00	−0.03	0.02	−0.02	−0.04	0.00	0.01	−0.01
Nitzschia frigida	0.00	0.00	0.01	0.01	0.01	0.00	0.01	−0.02
Nitzschia longissima	0.05	0.02	0.01	−0.01	−0.03	−0.04	0.05	−0.05
Nitzschia pacifica	0.02	−0.08	0.00	−0.01	−0.04	−0.08	0.03	−0.01
Nitzschia sicula	0.06	0.03	−0.04	−0.04	0.10	0.06	0.06	0.01
Nitzschia sigma	0.07	0.05	−0.01	0.01	0.02	0.01	−0.05	−0.05
Nitzschia tenuirostris	0.06	0.02	−0.03	−0.05	0.06	−0.02	0.01	−0.07
Nitzschia vitrea	0.04	0.04	0.06	0.05	−0.01	−0.02	0.02	0.07
Odontella aurita	−0.01	−0.05	0.11	0.04	−0.02	−0.06	0.00	−0.16
Odontella longicruris	0.00	−0.02	0.02	0.00	−0.03	0.00	0.00	−0.04
Plagiotropis lepidoptera	0.04	0.02	0.05	0.08	0.02	0.01	−0.02	0.09
Planktoniella sol	0.07	−0.04	0.01	−0.05	−0.01	−0.07	0.03	0.03
Pleurosigma angulatum	0.05	0.09	−0.03	0.00	−0.04	0.05	−0.05	−0.04
Pleurosigma directum	0.03	0.00	0.06	−0.03	0.08	−0.14	0.12	−0.02
Pleurosigma diversestriatum	0.04	0.00	−0.04	0.05	0.03	0.10	0.02	−0.05
Pleurosigma elongatum	0.05	0.07	−0.06	−0.05	−0.02	0.01	−0.03	−0.05
Pleurosigma nicobaricum	0.00	−0.01	0.00	0.00	−0.01	0.00	0.01	0.00
Pleurosigma normanii	0.06	0.02	−0.02	−0.06	0.01	0.00	−0.04	−0.03
Pleurosigma simonsenii	−0.01	−0.01	0.00	0.00	0.01	0.01	0.00	0.01
Podosira stelligera	−0.01	0.09	−0.07	−0.08	−0.10	−0.02	0.00	−0.01
Porosira glacialis	−0.01	0.00	0.02	0.02	0.00	−0.01	0.00	−0.05
Proboscia indica	−0.01	0.02	−0.02	−0.01	−0.05	0.00	−0.04	0.05
Proboscia inermis	−0.03	−0.04	0.03	0.03	0.03	−0.02	0.11	0.01
Proboscia truncata	−0.01	0.00	−0.01	0.00	0.00	0.01	−0.03	0.03
Pseudo-nitzschia lineola	−0.01	−0.03	0.02	0.01	−0.03	0.00	0.07	−0.02
Pseudo-nitzschia pseudodelicatissima	0.04	0.00	−0.01	0.06	0.04	0.12	−0.06	−0.06
Pseudo-nitzschia pungens	0.07	0.03	0.02	0.02	−0.01	0.03	−0.05	−0.05
Pseudo-nitzschia subfraudulenta	0.05	0.01	−0.06	0.01	0.11	0.11	−0.02	−0.02
Pseudosolenia calcar-avis	0.06	0.02	−0.02	−0.06	−0.04	0.01	−0.05	−0.06
Rhabdonema arcuatum	0.00	−0.01	0.08	0.05	0.01	−0.05	−0.03	−0.12
Rhizosolenia acuminata	0.04	0.00	0.05	−0.07	−0.05	−0.09	−0.02	0.17
Rhizosolenia bergonii	0.05	−0.02	0.01	−0.04	−0.05	−0.04	−0.02	0.10
Rhizosolenia castracanei	0.07	0.00	0.05	−0.03	−0.06	−0.03	0.02	0.06
Rhizosolenia chunii	−0.02	−0.03	0.03	0.03	0.00	−0.01	0.10	−0.03
Rhizosolenia clevei	0.06	0.04	0.06	0.01	0.09	0.00	−0.01	0.04
Rhizosolenia formosa	0.06	0.00	0.08	−0.06	0.02	−0.04	0.02	0.07
Rhizosolenia hyalina	0.02	−0.07	0.05	−0.04	−0.02	−0.01	−0.04	0.00
Rhizosolenia setigera	0.06	0.01	−0.01	−0.06	−0.06	−0.03	−0.03	0.00
Rhizosolenia temperei	0.01	−0.05	0.02	−0.02	−0.05	0.01	0.01	−0.01
Roperia tesselata	0.00	−0.04	0.01	−0.03	−0.03	0.02	0.01	−0.01
Shionodiscus gracilis	−0.02	0.00	0.01	0.03	0.03	−0.01	0.06	0.00
Shionodiscus oestrupii	0.05	0.04	−0.03	0.01	0.03	−0.06	0.00	−0.04
Stellarima stellaris	−0.01	0.00	0.05	0.04	0.01	−0.03	0.00	−0.11
Stephanopyxis nipponica	−0.01	−0.01	0.07	0.05	0.05	−0.05	−0.07	−0.10
Stephanopyxis palmeriana	0.06	0.02	0.06	−0.03	−0.08	−0.06	0.02	0.08
Striatella unipunctata	0.07	0.03	−0.02	0.07	0.02	−0.01	−0.04	−0.05

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Thalassionema bacillare	0.02	−0.07	0.01	−0.02	−0.06	0.01	0.02	−0.02
Thalassionema frauenfeldii	0.06	0.08	0.02	0.03	−0.05	0.00	−0.05	−0.02
Thalassiosira aestivalis	0.05	−0.01	−0.06	−0.03	−0.08	−0.03	0.09	−0.03
Thalassiosira angulata	0.02	0.06	0.12	0.05	−0.02	−0.01	0.08	−0.09
Thalassiosira angustelineata	0.00	−0.01	0.01	0.00	−0.01	0.01	0.00	−0.02
Thalassiosira antarctica	0.01	0.04	0.03	0.05	0.08	−0.06	0.11	0.02
Thalassiosira decipiens	0.03	0.04	0.14	0.02	0.05	−0.03	0.00	−0.06
Thalassiosira eccentrica	0.04	−0.04	0.09	−0.02	0.09	−0.05	0.11	−0.08
Thalassiosira gracilis	−0.01	0.00	0.02	0.04	0.04	−0.02	0.09	−0.04
Thalassiosira gravida	0.00	−0.03	0.07	0.02	−0.06	−0.02	−0.03	−0.10
Thalassiosira hyalina	−0.01	0.00	0.05	0.05	0.00	−0.05	−0.04	−0.10
Thalassiosira leptopus	0.01	0.02	0.11	0.10	0.03	0.01	0.10	−0.06
Thalassiosira mendiolana	0.00	−0.02	0.01	−0.01	−0.02	0.00	0.01	0.00
Thalassiosira minima	0.00	−0.01	0.00	0.00	−0.01	0.01	0.00	0.00
Thalassiosira nordenskioeldii	−0.02	−0.01	0.02	0.03	0.02	0.00	0.01	−0.06
Thalassiosira partheneia	0.01	−0.05	0.02	−0.02	−0.06	0.01	0.02	−0.01
Thalassiosira subtilis	0.02	−0.06	0.05	−0.01	−0.08	0.05	0.04	−0.02
Thalassiothrix heteromorpha	0.06	0.08	0.05	0.02	−0.03	0.03	0.04	0.01
Trieres chinensis	−0.02	0.02	0.03	−0.16	−0.02	0.05	0.01	0.03
Trieres mobiliensis	0.03	−0.07	0.05	−0.08	0.07	0.00	−0.05	−0.02
Trieres regia	−0.01	0.02	−0.01	−0.04	−0.01	0.08	−0.01	−0.01
Halosphaera viridis	−0.03	0.00	0.04	0.06	0.02	−0.02	0.06	−0.07
Pterosperma moebii	−0.01	0.00	0.01	0.03	0.02	0.00	0.03	−0.03
Pterosperma vanhoeffenii	0.00	0.00	0.01	0.01	0.00	0.01	−0.02	0.00
Ulva fasciata	0.03	0.00	0.00	0.03	−0.02	−0.03	−0.04	0.02
Dictyocha fibula	−0.01	−0.01	0.10	0.08	−0.02	−0.03	−0.02	−0.14
Dinobryon balticum	0.00	0.00	0.01	0.01	0.00	0.01	−0.01	−0.01
Octactis speculum	−0.01	−0.02	0.06	0.06	0.00	−0.01	0.01	−0.10
Hillea fusiformis	0.01	−0.02	0.03	0.02	0.01	−0.03	−0.08	−0.04
Leucocryptos marina	0.00	−0.03	0.01	−0.01	−0.04	0.01	0.03	0.00
Prochlorococcus	0.07	0.03	−0.05	−0.06	0.00	−0.02	−0.02	−0.03
Synechococcus	0.03	−0.02	0.04	0.00	0.09	−0.02	0.04	−0.06
Trichodesmium erythraeum	0.03	0.00	0.05	0.02	0.07	−0.01	−0.04	0.02
Trichodesmium thiebautii	0.03	0.00	0.05	0.06	0.00	0.02	−0.02	−0.01
Akashiwo sanguinea	0.00	−0.02	0.01	−0.01	−0.03	0.01	0.01	−0.01
Alexandrium monilatum	0.02	−0.04	0.05	0.00	0.02	0.01	−0.06	0.02
Amphidinium acutissimum	0.06	−0.07	−0.02	0.02	0.06	0.07	0.04	0.04
Amphidinium carterae	0.00	−0.01	−0.02	0.00	0.03	0.05	−0.04	0.00
Amphidinium sphenoides	0.00	0.00	0.00	−0.01	0.02	0.00	−0.02	0.00
Amphisolenia bidentata	0.07	−0.04	0.03	−0.05	0.07	−0.01	0.05	0.00
Amphisolenia bispinosa	0.00	−0.02	0.01	−0.01	−0.03	0.01	0.02	0.00
Archaeperidinium minutum	0.03	−0.02	0.02	0.04	−0.02	0.10	0.06	0.02
Azadinium caudatum	0.00	0.00	0.00	0.00	0.00	0.01	−0.01	0.00
Bitectatodinium spongium	0.04	−0.03	−0.04	−0.04	0.10	0.04	0.04	−0.06
Blepharocysta splendor-maris	0.07	0.02	−0.03	0.10	0.04	0.02	−0.01	0.00
Ceratium arcticum	−0.01	0.00	0.01	0.02	0.01	0.01	−0.04	−0.01
Ceratium breve	0.05	−0.08	0.02	0.03	0.01	0.02	0.00	0.00
Ceratium contrarium	0.08	0.02	−0.02	−0.05	0.00	−0.02	−0.01	−0.06
Ceratium falcatiforme	0.07	−0.05	0.02	0.01	−0.03	0.00	0.02	0.00
Ceratium falcatum	0.07	−0.05	0.00	0.02	0.01	0.03	0.03	0.03
Ceratium gibberum	0.02	−0.08	−0.02	0.02	−0.08	0.01	0.04	0.03
Ceratium gravidum	0.06	−0.05	−0.01	0.09	0.01	0.09	−0.02	0.03
Ceratium longirostrum	0.03	−0.06	−0.02	0.07	−0.09	0.05	0.05	0.01
Ceratium massiliense	0.06	0.00	−0.07	0.00	−0.07	0.06	−0.01	0.00
Ceratium pavillardii	0.03	0.05	−0.02	0.03	−0.01	0.16	0.08	0.02
Ceratium setaceum	0.07	−0.02	−0.05	−0.05	−0.01	−0.02	0.03	−0.01
Ceratium symmetricum	0.06	−0.06	0.03	−0.05	0.06	−0.01	0.03	0.05
Ceratium trichoceros	0.05	−0.01	−0.08	−0.03	−0.11	0.11	0.03	−0.03
Ceratocorys armata	0.04	−0.07	0.04	0.05	−0.01	0.06	−0.05	0.01
Ceratocorys bipes	0.00	−0.02	0.01	−0.01	−0.03	0.01	0.01	−0.01
Ceratocorys horrida	0.07	−0.08	−0.01	0.01	0.04	0.00	0.00	0.01
Ceratocorys reticulata	0.01	−0.06	0.02	−0.03	−0.05	0.01	0.01	0.00
Ceratoperidinium falcatum	0.07	0.00	−0.02	0.00	0.06	−0.03	−0.03	0.00
Cladopyxis brachiolata	0.07	0.02	−0.03	−0.04	0.06	0.00	0.05	−0.03
Cochlodinium pupa	0.00	0.00	0.00	−0.09	0.06	−0.01	−0.03	0.02
Cochlodinium vinctum	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00
Corythodinium tesselatum	0.07	−0.04	−0.03	0.05	0.01	0.01	−0.04	−0.02
Dinophysis acuminata	−0.01	−0.04	0.03	0.01	−0.02	0.03	0.00	−0.05
Dinophysis acuta	−0.01	0.00	0.01	0.02	0.01	0.01	−0.01	0.00
Dinophysis apicata	0.00	−0.02	0.01	−0.01	−0.02	0.00	0.03	0.01
Dinophysis argus	0.05	−0.07	−0.02	0.02	−0.01	0.13	0.05	0.00
Dinophysis caudata	0.02	−0.09	0.00	−0.02	−0.06	0.02	0.00	−0.01
Dinophysis fortii	0.07	0.03	−0.01	0.09	−0.01	−0.05	−0.05	−0.05

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Dinophysis hastata	0.05	−0.07	0.01	0.06	−0.01	0.08	−0.03	0.01
Dinophysis norvegica	−0.01	0.00	0.01	0.02	0.01	0.01	−0.02	−0.01
Dinophysis ovum	0.05	0.02	−0.07	0.08	−0.05	−0.05	0.03	0.02
Dinophysis parvula	0.07	0.03	−0.04	0.00	−0.01	−0.01	0.02	−0.03
Dinophysis porodictyum	0.05	−0.05	−0.02	0.02	0.08	0.10	−0.01	0.05
Dinophysis sacculus	0.04	0.04	−0.09	0.04	−0.07	0.11	0.04	−0.09
Dinophysis schroederi	0.05	0.07	−0.08	0.00	−0.04	−0.01	0.01	−0.04
Dinophysis schuettii	0.07	0.00	−0.04	−0.05	−0.02	−0.02	0.00	−0.03
Dinophysis tripos	−0.01	0.01	−0.04	0.03	−0.05	−0.04	−0.07	0.04
Diplopelta asymmetrica	0.00	−0.05	0.01	−0.03	−0.05	0.00	0.03	0.01
Diplopelta steinii	0.00	−0.03	0.01	−0.02	−0.04	0.01	0.02	0.00
Diplopsalis lenticula	0.03	0.01	−0.07	0.10	−0.06	−0.01	0.05	0.00
Diplopsalopsis bomba	0.02	−0.05	0.04	−0.03	0.02	0.00	−0.03	0.02
Echinidinium delicatum	0.05	−0.03	−0.05	−0.02	0.03	−0.04	0.03	0.00
Goniodoma sphaericum	0.06	−0.03	−0.03	0.02	0.06	0.10	−0.01	0.01
Gonyaulax birostris	0.07	0.02	−0.02	0.03	0.08	0.02	−0.02	−0.04
Gonyaulax diegensis	0.06	−0.04	−0.02	0.11	−0.03	−0.01	−0.04	0.01
Gonyaulax digitalis	0.07	−0.04	−0.02	0.02	0.03	−0.04	−0.02	0.05
Gonyaulax elongata	−0.01	0.00	0.02	0.03	0.00	0.00	−0.02	−0.04
Gonyaulax membranacea	0.07	−0.03	−0.03	−0.03	0.02	−0.07	0.02	0.06
Gonyaulax monacantha	0.07	−0.02	−0.07	0.01	−0.03	0.00	0.02	−0.02
Gonyaulax pacifica	0.05	−0.08	0.04	−0.03	0.04	−0.01	−0.01	−0.02
Gonyaulax polygramma	0.06	−0.08	0.01	0.03	0.04	0.07	0.01	0.03
Gonyaulax scrippsae	0.07	0.04	−0.04	−0.04	−0.03	−0.01	0.00	−0.04
Gonyaulax spinifera	0.06	−0.01	−0.02	−0.03	0.00	−0.06	0.04	−0.03
Gymnodinium agiliforme	0.00	−0.01	0.01	0.01	0.02	0.03	−0.06	0.02
Gymnodinium arcticum	−0.02	0.00	0.02	0.04	0.03	−0.02	0.07	−0.05
Gymnodinium aureolum	0.00	−0.02	0.00	−0.01	0.01	0.01	−0.04	−0.02
Gymnodinium catenatum	0.06	0.00	−0.02	0.00	0.08	−0.03	−0.08	−0.05
Gymnodinium gracile	0.01	−0.08	0.02	−0.03	−0.05	−0.02	−0.01	−0.03
Gymnodinium marinum	0.06	0.00	−0.06	0.08	0.01	−0.12	−0.03	0.04
Gymnodinium simplex	0.03	0.04	−0.06	0.06	−0.06	−0.08	0.04	0.02
Gymnodinium uberrimum	0.01	0.00	−0.05	0.11	−0.02	−0.01	0.12	−0.01
Gymnodinium wulffii	−0.03	0.04	0.03	0.00	−0.03	−0.06	0.08	0.03
Gyrodinium flagellare	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	0.00
Gyrodinium fusiforme	0.05	−0.04	0.02	−0.04	0.14	0.03	0.06	0.06
Gyrodinium pingue	0.00	−0.02	0.04	0.02	0.03	0.01	−0.05	−0.03
Gyrodinium prunus	0.00	0.00	0.00	0.00	0.01	0.00	−0.01	−0.01
Gyrodinium spirale	0.06	0.02	−0.05	0.04	0.03	0.07	−0.09	−0.06
Gyrodinium wulffii	0.00	−0.01	0.00	−0.01	0.01	0.00	−0.02	0.01
Heterocapsa niei	0.02	0.00	−0.02	0.05	−0.02	0.09	−0.05	0.01
Heterocapsa rotundata	0.00	0.01	−0.01	−0.05	0.04	0.01	−0.02	0.00
Heterocapsa triquetra	−0.01	0.00	0.01	0.02	0.02	−0.01	0.01	−0.01
Heterodinium blackmanii	0.01	−0.06	0.03	−0.04	−0.05	−0.02	0.01	−0.03
Impagidinium aculeatum	0.05	0.00	−0.03	−0.04	0.03	−0.09	0.09	0.02
Impagidinium patulum	0.05	0.01	−0.04	−0.04	0.01	−0.09	0.07	0.01
Impagidinium sphaericum	0.04	0.02	−0.03	−0.02	0.00	−0.10	0.12	0.03
Kapelodinium vestifici	−0.01	−0.01	0.02	0.03	0.03	0.00	0.03	−0.02
Karenia brevis	0.06	0.01	−0.03	0.09	0.01	0.02	−0.06	−0.06
Karenia mikimotoi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Karlodinium veneficum	0.00	0.00	0.01	0.00	0.01	−0.01	−0.02	−0.01
Kofoidinium velleloides	0.00	−0.05	0.00	−0.04	−0.04	0.01	0.03	0.00
Lebessphaera urania	0.00	0.01	0.00	−0.01	−0.02	0.02	0.00	0.00
Lebouridinium glaucum	0.06	0.00	−0.06	−0.06	0.10	0.00	0.05	0.00
Leonella granifera	0.07	0.01	−0.03	−0.02	0.09	−0.02	0.00	−0.04
Lingulodinium polyedra	0.07	−0.02	−0.04	−0.05	0.00	−0.03	0.00	−0.04
Mesoporos perforatus	0.01	−0.01	−0.01	0.06	−0.08	−0.06	−0.05	0.05
Neoceratium breve	0.03	−0.01	−0.02	0.08	−0.02	0.00	0.02	0.02
Neoceratium hexacanthum	0.02	−0.01	0.02	0.01	0.03	−0.01	−0.16	−0.04
Ornithocercus heteroporus	0.07	0.05	−0.02	0.04	0.02	0.00	−0.02	−0.06
Ornithocercus magnificus	0.07	−0.05	−0.01	0.00	0.03	−0.01	0.01	−0.01
Ornithocercus quadratus	0.05	−0.09	−0.01	0.00	0.05	0.01	0.03	0.03
Ornithocercus splendidus	0.05	−0.02	0.00	0.02	0.03	0.15	0.05	0.01
Ornithocercus steinii	0.03	−0.06	0.03	0.03	−0.01	0.06	0.00	0.01
Ornithocercus thumii	0.04	−0.08	0.03	−0.01	0.01	0.08	0.01	−0.01
Ornithocercus thurnii	0.06	−0.05	−0.03	0.10	0.02	0.01	−0.04	0.04
Oxytoxum caudatum	0.06	−0.06	−0.07	0.05	−0.01	−0.02	0.01	0.05
Oxytoxum constrictum	0.03	0.01	−0.12	−0.06	0.02	0.06	0.02	−0.01
Oxytoxum curvatum	0.06	−0.03	−0.05	0.10	−0.03	−0.07	0.00	0.08
Oxytoxum elegans	0.07	−0.03	−0.04	0.04	−0.03	0.04	0.04	0.03
Oxytoxum gracile	0.05	0.01	−0.10	0.07	0.02	0.03	0.02	−0.06
Oxytoxum laticeps	0.04	−0.05	0.03	0.03	0.11	0.03	−0.04	0.07
Oxytoxum longiceps	0.08	−0.03	0.00	−0.05	0.06	−0.03	0.03	−0.02

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Oxytoxum milneri	0.08	0.02	−0.05	0.00	0.05	−0.03	0.00	−0.02
Oxytoxum parvum	0.07	−0.01	−0.03	0.02	0.02	−0.09	−0.01	0.00
Oxytoxum reticulatum	0.04	−0.05	−0.06	0.07	−0.06	−0.02	0.04	0.05
Oxytoxum sceptrum	0.06	0.03	−0.05	0.11	0.01	−0.01	0.03	0.03
Oxytoxum scolopax	0.07	0.00	−0.04	−0.05	−0.02	−0.05	−0.01	−0.01
Oxytoxum sphaeroideum	0.05	−0.01	−0.05	−0.03	−0.01	−0.15	−0.07	0.00
Oxytoxum tesselatum	0.05	0.06	−0.06	0.03	−0.06	−0.04	0.00	0.00
Oxytoxum turbo	0.07	0.04	−0.06	−0.03	−0.04	−0.04	0.05	0.00
Oxytoxum variabile	0.06	−0.02	−0.03	−0.02	0.01	−0.10	−0.02	0.02
Pentapharsodinium dalei	0.04	−0.05	−0.01	−0.03	0.04	−0.13	0.02	0.00
Peridinium breve	0.00	−0.01	0.05	0.04	−0.02	−0.02	−0.03	−0.10
Phalacroma doryphorum	0.06	−0.03	0.00	0.09	0.03	0.05	−0.02	0.02
Phalacroma favus	0.06	−0.05	−0.03	0.10	−0.02	−0.04	−0.05	0.05
Phalacroma lens	0.04	−0.04	0.04	−0.04	−0.09	−0.06	0.04	0.01
Phalacroma mitra	0.07	0.00	0.00	0.10	0.04	0.00	−0.07	0.02
Phalacroma oxytoxoides	0.01	−0.04	0.03	0.01	−0.06	0.02	−0.01	−0.03
Phalacroma rapa	0.03	−0.07	0.03	0.00	−0.02	0.02	0.00	0.03
Phalacroma rotundatum	0.01	0.01	−0.02	0.17	−0.06	−0.05	−0.01	0.02
Podolampas bipes	0.05	−0.08	0.03	−0.02	0.01	0.01	0.00	−0.03
Podolampas elegans	0.07	0.00	−0.04	−0.06	−0.01	0.01	0.03	−0.05
Podolampas palmipes	0.07	−0.02	−0.04	−0.03	−0.05	−0.04	0.05	0.00
Podolampas spinifera	0.07	0.00	−0.03	−0.06	−0.01	−0.03	0.01	−0.04
Polykrikos schwartzii	−0.01	−0.01	0.00	0.01	0.01	0.02	−0.04	0.03
Preperidinium meunieri	0.03	−0.06	0.03	0.06	−0.01	0.02	0.06	0.02
Pronoctiluca pelagica	0.07	−0.04	−0.02	0.02	−0.02	−0.05	−0.01	0.02
Pronoctiluca spinifera	0.07	0.00	−0.02	0.11	0.01	0.01	−0.02	0.00
Prorocentrum arcuatum	0.01	−0.03	0.03	0.01	0.00	0.03	−0.04	−0.02
Prorocentrum balticum	0.06	0.00	0.04	0.05	0.11	−0.03	0.07	−0.08
Prorocentrum cordatum	0.03	−0.02	0.03	0.07	0.07	−0.04	0.05	0.07
Prorocentrum dentatum	0.07	0.01	−0.03	−0.01	−0.02	−0.05	−0.04	0.00
Prorocentrum gracile	0.04	−0.08	0.04	0.01	0.01	0.04	−0.03	−0.01
Prorocentrum lima	0.06	0.00	−0.03	−0.03	−0.01	−0.06	0.08	−0.01
Prorocentrum mexicanum	0.08	−0.01	0.02	−0.06	0.09	0.02	0.05	−0.03
Prorocentrum micans	0.07	−0.06	−0.01	0.00	0.04	−0.01	0.00	0.01
Prorocentrum rostratum	0.08	−0.02	−0.04	−0.01	0.02	−0.07	−0.01	−0.01
Prorocentrum scutellum	0.06	0.05	−0.05	−0.04	−0.03	−0.02	−0.06	−0.04
Prorocentrum triestinum	0.06	0.02	0.05	0.07	0.01	0.07	−0.01	−0.03
Protoceratium reticulatum	0.05	−0.01	−0.02	0.00	0.05	−0.11	0.05	−0.01
Protoperidinium abei	0.07	−0.04	0.00	0.05	−0.02	−0.10	−0.04	0.08
Protoperidinium americanum	0.02	−0.04	0.00	−0.08	0.02	−0.03	0.00	−0.03
Protoperidinium bipes	−0.02	−0.01	0.08	0.07	0.10	−0.03	0.05	0.01
Protoperidinium brevipes	−0.01	0.01	0.02	0.09	0.02	−0.06	0.12	0.06
Protoperidinium brochii	0.06	−0.08	0.01	0.02	−0.01	−0.05	−0.03	0.01
Protoperidinium cerasus	0.04	0.03	0.06	0.08	0.03	−0.07	−0.03	−0.08
Protoperidinium claudicans	0.01	−0.05	0.04	0.00	−0.05	0.00	−0.04	−0.05
Protoperidinium conicoides	0.00	−0.02	0.02	0.01	−0.02	0.01	0.00	−0.03
Protoperidinium crassipes	0.02	−0.07	0.05	−0.01	−0.07	−0.01	−0.03	−0.06
Protoperidinium curtipes	0.01	−0.05	0.00	0.02	−0.06	0.04	−0.02	−0.01
Protoperidinium defectum	−0.01	0.00	0.02	0.04	0.03	−0.02	0.08	−0.05
Protoperidinium depressum	0.00	−0.08	0.01	−0.02	−0.03	0.02	0.01	−0.05
Protoperidinium diabolus	0.01	−0.06	−0.03	0.02	−0.10	0.03	0.04	0.04
Protoperidinium divergens	0.05	−0.09	0.02	0.03	−0.01	0.01	0.00	0.06
Protoperidinium elegans	0.05	−0.09	−0.01	0.05	−0.01	−0.02	−0.01	0.02
Protoperidinium excentricum	0.00	−0.03	0.01	−0.02	−0.04	0.01	0.02	0.00
Protoperidinium fatulipes	0.00	−0.04	0.01	−0.02	−0.05	0.01	0.02	0.00
Protoperidinium grahamii	0.00	−0.02	0.01	0.00	−0.02	0.01	0.01	−0.01
Protoperidinium grande	0.03	−0.10	0.01	−0.04	−0.01	−0.02	−0.02	0.01
Protoperidinium granii	0.03	−0.05	0.04	0.05	−0.05	0.04	−0.02	−0.02
Protoperidinium latidorsale	0.07	−0.01	−0.01	−0.08	0.05	0.00	0.04	−0.02
Protoperidinium latispinum	0.03	−0.01	0.05	0.01	0.05	−0.07	−0.08	0.05
Protoperidinium leonis	0.03	−0.06	0.03	0.01	0.00	0.08	−0.02	0.01
Protoperidinium longipes	0.02	−0.08	0.02	−0.03	−0.03	0.02	0.00	0.01
Protoperidinium longispinum	0.00	−0.01	0.01	−0.01	−0.02	0.01	0.01	0.00
Protoperidinium mediterraneum	0.07	0.03	−0.03	0.02	0.02	0.00	−0.02	−0.06
Protoperidinium mendiolae	0.00	−0.01	0.00	0.00	−0.02	0.01	0.01	0.00
Protoperidinium mite	0.04	0.01	0.04	0.06	0.04	0.12	−0.05	0.02
Protoperidinium oblongum	0.00	−0.03	0.01	−0.01	−0.04	0.02	0.01	−0.01
Protoperidinium obtusum	0.01	−0.03	0.02	0.00	−0.04	0.01	0.00	−0.02
Protoperidinium oceanicum	0.01	−0.06	0.02	−0.02	−0.06	0.00	0.01	−0.01
Protoperidinium ovatum	0.00	0.00	0.00	0.00	0.00	0.01	−0.01	0.00
Protoperidinium ovum	0.07	−0.02	−0.02	0.06	−0.01	0.00	−0.07	−0.03
Protoperidinium pallidum	0.00	−0.03	0.02	−0.01	−0.03	0.02	0.00	−0.02

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Protoperidinium pedunculatum	0.05	−0.08	0.00	0.02	−0.01	−0.04	−0.05	0.02
Protoperidinium pellucidum	0.00	−0.04	0.03	0.00	−0.05	0.03	−0.02	−0.04
Protoperidinium pentagonum	0.00	−0.06	0.02	−0.03	−0.04	0.02	0.00	−0.02
Protoperidinium peruvianum	0.00	0.00	0.00	0.00	−0.01	0.01	0.00	0.00
Protoperidinium punctulatum	0.08	0.01	−0.01	0.00	0.02	−0.05	−0.02	−0.05
Protoperidinium pyriforme	0.06	−0.05	0.00	0.09	−0.01	0.02	−0.05	0.01
Protoperidinium pyrum	0.00	−0.04	0.02	−0.02	−0.05	0.01	0.01	−0.01
Protoperidinium quarnerense	0.06	−0.09	0.00	0.04	0.01	−0.03	0.02	0.00
Protoperidinium steinii	0.03	−0.08	0.02	0.02	−0.04	0.03	−0.01	0.02
Protoperidinium stellatum	0.02	−0.02	−0.04	−0.01	−0.04	−0.02	0.03	−0.06
Protoperidinium subinerme	0.00	−0.04	0.02	−0.02	−0.06	0.00	0.01	−0.03
Protoperidinium tenuissimum	0.00	−0.05	0.01	−0.03	−0.05	0.01	0.03	0.01
Protoperidinium tristylum	0.00	−0.04	0.01	−0.03	−0.05	0.01	0.03	0.01
Protoperidinium tuba	0.07	−0.02	−0.04	0.04	0.01	0.05	0.00	−0.04
Protoperidinium venustum	0.01	−0.02	0.02	0.02	−0.01	0.01	−0.05	−0.01
Ptychodiscus noctiluca	0.00	−0.03	0.00	−0.03	−0.01	0.02	−0.02	0.02
Pyrocystis elegans	0.00	−0.03	0.01	−0.02	−0.05	0.01	0.02	−0.01
Pyrocystis fusiformis	0.02	−0.09	0.01	−0.07	−0.02	−0.03	0.00	−0.01
Pyrocystis hamulus	0.06	0.00	0.00	0.03	0.04	0.07	0.00	−0.01
Pyrocystis lunula	0.05	−0.04	0.03	0.06	0.04	0.04	−0.10	0.01
Pyrocystis robusta	0.06	−0.03	0.07	−0.02	0.07	−0.04	−0.02	0.09
Pyrophacus horologium	0.06	−0.08	0.02	0.05	0.00	0.07	−0.03	0.02
Pyrophacus steinii	0.06	−0.09	0.02	0.00	0.06	0.03	−0.01	0.03
Pyrophacus vancampoae	0.06	−0.07	−0.05	0.09	−0.01	0.00	−0.05	0.03
Schuettiella mitra	0.02	−0.08	0.03	−0.04	−0.02	0.01	0.00	−0.02
Scrippsiella acuminata	0.00	−0.05	0.02	0.00	−0.05	0.03	−0.01	−0.03
Scrippsiella regalis	0.06	0.06	−0.02	−0.09	0.04	−0.01	0.04	−0.02
Spiniferites pachydermus	0.05	−0.05	−0.08	0.07	−0.01	0.01	0.00	−0.05
Spiraulax kofoidii	0.06	−0.09	0.03	−0.02	0.05	−0.03	−0.06	−0.02
Thoracosphaera heimii	0.07	0.01	−0.03	−0.05	0.06	−0.01	0.01	−0.07
Torodinium robustum	0.01	0.00	0.03	−0.04	0.10	−0.02	−0.05	0.04
Triadinium polyedricum	0.05	−0.08	−0.04	0.04	−0.02	0.05	0.05	0.05
Trinovantidinium applanatum	0.06	0.00	−0.05	−0.06	0.03	−0.05	0.01	0.02
Tripos azoricus	−0.01	−0.04	0.00	−0.02	−0.06	0.00	−0.01	0.03
Tripos balechii	0.03	0.00	0.04	0.04	0.08	0.01	−0.06	0.04
Tripos belone	0.07	−0.02	−0.02	−0.03	0.00	0.01	0.01	−0.02
Tripos bigelowii	0.00	−0.01	0.00	0.00	−0.03	0.00	0.00	−0.01
Tripos bucephalus	−0.01	0.00	−0.01	0.01	0.00	0.01	−0.04	0.04
Tripos buceros	0.02	−0.05	−0.02	0.03	−0.10	0.00	0.08	0.03
Tripos candelabrum	0.05	−0.06	−0.09	0.05	−0.09	0.01	−0.02	−0.01
Tripos carnegiei	0.00	−0.01	0.01	−0.01	−0.02	0.01	0.01	0.00
Tripos carriensis	0.07	−0.01	−0.04	−0.03	−0.01	0.02	−0.01	0.00
Tripos compressus	0.00	0.00	0.00	0.00	0.00	0.01	−0.01	0.02
Tripos concilians	0.05	−0.05	−0.06	0.00	0.00	0.08	0.09	0.03
Tripos contortus	0.07	−0.07	0.01	0.03	0.01	0.02	0.00	0.02
Tripos declinatus	0.07	−0.01	−0.02	0.03	−0.04	−0.01	−0.06	−0.01
Tripos deflexus	0.06	−0.05	0.08	−0.03	0.05	−0.02	0.00	0.07
Tripos dens	0.00	−0.05	0.01	−0.03	−0.05	0.01	0.03	0.01
Tripos digitatus	0.01	−0.02	0.02	0.00	0.00	0.03	−0.02	0.00
Tripos euarcuatus	0.06	−0.08	0.02	0.05	0.02	0.06	−0.03	0.01
Tripos eugrammus	0.00	−0.05	0.01	−0.03	−0.06	0.00	0.02	0.00
Tripos extensus	0.05	0.01	−0.03	−0.08	−0.06	0.00	−0.06	0.01
Tripos geniculatus	0.00	−0.03	0.01	−0.02	−0.03	0.00	0.01	0.00
Tripos gravidus	0.01	−0.06	0.02	−0.04	−0.06	0.00	0.04	0.01
Tripos hexacanthus	0.00	−0.05	−0.06	−0.03	−0.07	0.01	0.01	0.06
Tripos incisus	0.01	−0.06	0.01	−0.03	−0.06	−0.01	0.02	−0.01
Tripos inflatus	0.04	−0.08	0.04	0.00	0.03	0.00	−0.06	0.02
Tripos karstenii	0.07	0.05	−0.04	0.01	−0.01	0.03	0.01	−0.03
Tripos kofoidii	0.07	−0.06	0.00	0.02	0.01	−0.01	0.00	0.01
Tripos lamellicornis	−0.01	0.00	−0.01	0.00	0.00	0.02	−0.03	0.03
Tripos limulus	0.03	−0.07	0.04	−0.01	−0.03	0.06	0.00	−0.01
Tripos lunula	0.01	−0.08	0.00	−0.05	−0.06	−0.01	0.02	−0.02
Tripos massiliensis	0.04	0.03	−0.07	−0.02	0.07	0.04	−0.01	0.04
Tripos minutus	−0.01	−0.01	0.00	0.00	0.01	0.03	−0.05	0.03
Tripos paradoxides	0.03	−0.07	0.03	−0.05	0.04	0.02	0.01	0.02
Tripos pentagonus	0.06	−0.02	0.00	−0.02	−0.02	−0.01	0.01	−0.10
Tripos platycornis	−0.01	−0.01	0.00	−0.01	0.00	0.03	−0.02	0.02
Tripos pulchellus	0.08	−0.04	0.00	0.01	−0.01	0.00	0.04	0.03
Tripos ranipes	0.02	−0.07	−0.02	−0.06	−0.05	0.04	0.04	−0.06
Tripos strictus	0.00	−0.06	0.01	−0.04	−0.06	0.00	0.04	0.01
Tripos teres	0.06	0.05	−0.05	−0.04	−0.02	−0.01	−0.05	−0.05
Tripos trichoceros	0.05	0.06	−0.06	−0.03	0.05	0.03	0.02	0.04
Tripos vultur	0.05	−0.09	0.01	0.02	−0.01	0.06	0.01	0.04

(*continued on next page*)

Table A.6 (*continued*)

Species	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Tryblionella compressa	0.06	−0.01	0.02	−0.02	0.12	−0.01	0.01	0.00
Eutreptiella gymnastica	0.00	−0.02	0.01	−0.01	−0.03	0.01	0.02	0.00
Acanthoica acanthifera	0.02	−0.01	−0.01	0.03	−0.02	−0.03	0.02	0.02
Algirosphaera robusta	0.04	−0.03	−0.04	0.01	0.03	−0.11	−0.08	−0.05
Calciopappus rigidus	0.03	−0.04	−0.03	−0.02	0.04	−0.16	−0.12	−0.06
Calciosolenia brasiliensis	0.01	−0.08	−0.03	−0.04	−0.04	−0.02	0.01	−0.05
Calciosolenia granii	0.00	−0.02	0.00	−0.01	−0.03	0.00	0.01	0.00
Calciosolenia murrayi	0.03	−0.06	−0.04	0.05	−0.07	−0.02	0.03	−0.05
Coronosphaera mediterranea	0.01	0.00	−0.07	0.06	−0.02	−0.02	−0.01	−0.02
Discosphaera tubifera	0.07	0.03	−0.04	0.00	0.00	−0.06	0.01	−0.01
Florisphaera profunda	0.05	−0.05	−0.07	−0.01	0.04	−0.09	0.04	0.02
Gephyrocapsa caribbeanica	0.01	0.03	−0.07	0.06	−0.06	0.00	0.02	−0.01
Gephyrocapsa ericsonii	0.03	0.01	−0.09	0.08	−0.03	−0.08	0.04	0.00
Gephyrocapsa muellerae	0.01	−0.04	−0.07	0.04	−0.06	−0.09	−0.02	0.00
Gladiolithus flabellatus	0.07	−0.03	−0.05	0.06	0.02	−0.06	0.01	0.07
Hayaster perplexus	0.07	−0.04	−0.06	−0.02	0.08	0.00	0.05	0.06
Helladosphaera cornifera	0.06	0.00	−0.06	0.10	0.02	−0.01	−0.02	0.09
Michaelsarsia adriatica	0.04	−0.07	−0.03	0.06	−0.06	−0.02	0.03	0.01
Michaelsarsia elegans	0.04	−0.06	0.01	−0.05	0.09	−0.08	0.00	0.02
Oolithotus fragilis	0.03	−0.02	−0.12	0.04	−0.02	0.03	0.02	−0.10
Ophiaster hydroideus	0.00	−0.04	0.01	−0.01	−0.04	0.02	0.01	0.01
Phaeocystis antarctica	−0.02	0.00	0.02	0.04	0.04	−0.02	0.09	−0.05
Phaeocystis pouchetii	−0.02	0.00	0.01	0.02	0.00	0.02	−0.06	0.03
Reticulofenestra sessilis	0.06	−0.01	−0.06	0.10	0.00	−0.03	0.05	0.04
Rhabdolithes claviger	0.03	0.05	−0.12	−0.04	0.05	0.11	−0.03	−0.04
Rhabdosphaera hispida	0.01	0.01	−0.02	0.01	0.03	0.01	0.02	0.00
Rhabdosphaera xiphos	0.03	−0.01	−0.04	0.10	−0.02	−0.06	−0.04	0.00
Syracosphaera molischii	0.03	0.00	−0.02	0.08	0.00	−0.05	−0.16	0.05
Syracosphaera nodosa	0.06	−0.01	−0.10	0.04	0.01	−0.01	0.02	−0.08
Syracosphaera prolongata	0.04	0.06	−0.03	−0.02	0.01	−0.01	−0.02	−0.06
Umbellosphaera irregularis	0.07	0.06	−0.07	−0.01	0.03	−0.01	0.02	−0.04
Umbellosphaera tenuis	0.06	−0.01	−0.08	0.01	0.00	−0.05	0.02	0.02
Umbilicosphaera hulburtiana	0.06	0.01	−0.07	0.08	0.01	−0.07	−0.03	0.00
Umbilicosphaera sibogae	0.06	0.04	−0.04	−0.04	−0.02	−0.02	−0.06	−0.05

A.5. Uncertainty in species' presence projections due to algorithm choice and effects of algorithm choice on biome boundaries

We tested to what degree uncertainty in species' geographic distributions predicted by SDMs translates into uncertainties in our biome boundaries. Particularly, we explored the effect of algorithm choice at the level of SDMs on resulting biome boundaries. This test is relevant, as algorithm choice has been identified as a major source of uncertainty (around 73%) in SDM projections (Buisson et al., 2010; Benedetti et al., 2018, and references therein). Species presence patterns were originally projected in Righetti et al. (2019a) using three algorithms of increasing statistical response shape complexity (Merow et al., 2014), i.e. Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), and Random Forests (RFs), which delineate species' distributions with less or more flexibility. Simpler models (e.g. GLMs, GAMs) with reduced statistical response complexity and few predictors are generally able to perform better under data sparseness than complex ones (e.g. RFs; Merow et al., 2014). We compared differences between algorithms as follows: (1) we characterized how species' richness projected by GLMs, and RFs, differed from the richness projected by GAMs. We calculate absolute richness differences, normalized to the richness predicted by the GAM-approach, at monthly 1° resolution, and we show the annual median pattern of the twelve months (e.g. 0.5 indicates that half of the local species richness is now missing, or half of the previously absent richness is now projected as present). (2) Similar to other application of self-organizing maps (Weber et al., 2010), we use our trained self-organizing map to predict the cluster association of the presence-absence patterns from GLM and RF projections. We expected results in comparisons between GLMs and GAMs to be more similar than those between GAMs and RFs, since the former two algorithms are statistically more similar to each other than the other two.

We chose the GAM as the basis for our biome partitioning given its intermediate statistical response shape complexity relative to GLM and RF (Merow et al., 2014). We find that species' presences projected by GLMs and RFs (Fig. A.11) both differ the least from GAM-projections in the tropics, and the most in mid latitude regions of both hemispheres. Regions of small/large differences in presence projections of GLMs and RFs relative to GAMs are closely associated with regions of low/high species turnover and regions of high/low species richness. This shows that models agree more closely where species patterns are stable throughout the year and predicted species richness is higher (compare to Fig. 1 A, and B in Righetti et al., 2019a). Within the tropics, including the equatorial Pacific (Fig. A.11B) and the equatorial Atlantic region (Fig. A.11B), we also observe regions of relatively larger differences between presence projections of GLMs and RFs relative to those of GAM-projections.

Analyzed by latitude (Fig. A.12), richness differences of GML and RF relative to GAM projections are both largest between 40°S and 50°S. For the Northern Hemisphere, we find a similar peak around 40°N in differences between the projections, but with smaller magnitude than in the Southern Hemisphere. Overall, the latitudinal patterns of differences in species richness projections coincide with latitudinal pattern of species' temporal (month-to-month) turnover reported previously (Righetti et al., 2019a). Thus, we would expect that the geographic location of biomes is most sensitive to algorithmn choice at around 40° in both hemispheres, and around the equatorial Pacific region.

We tested whether uncertainties in species' projections translate into uncertainties in biome boundaries, at the suspected locations as mentioned above. The annual biomes based on presence patterns projected by GLMs, GAMs, and RFs are shown in Fig. A.13A, B, and C, respectively. We

quantified the spatio-temporal agreement between sets of biomes at monthly pairwise-comparative resolution using the Kappa index (Cohen, 1960; Landis and Koch, 1977). This index quantifies the pixel-wise, and area weighted agreement in the locations of the biomes at monthly scales, with a Kappa index above 0.7 suggesting a very good agreement between patterns, and above 0.85 suggesting excellent agreement. The mean (± standard deviation) Kappa index across all 12 months for the comparison between GLM and GAM derived biomes was 0.81 ± 0.14 with a minimum of 0.58 (February) and a maximum of 0.97 (September). For RF compared to GAM, the average Kappa index was 0.76 ± 0.20 with a minimum of 0.39 (January) and a maximum of 0.95 (August). Overall, the agreement between both GLM- and GAM-based biomes, and RF- and GAM-based biomes was very good. As expected, GLM- and GAM-derived biome patterns were more consistent than RF- vs GAM-derived biome patterns. GLM- and GAM-derived biomes overlapped across 70 ± 8% of the global surface ocean when considering all months together (Fig. A.13D), and RF- and GAM-derived biomes across 64 ± 12% of the global surface ocean area (Fig. A.13E). Differences between the biome patterns based on algorithm emerge mainly in the Southern Hemisphere at around 40°, and around the equatorial Pacific region (Fig. A.13D, E). We conclude that our biomes are largely robust to uncertainty in species presence patterns induced by algorithm choice, yet the exact geographical location of the biome boundaries was less certain at around 40° in both hemispheres, and in the equatorial Pacific.

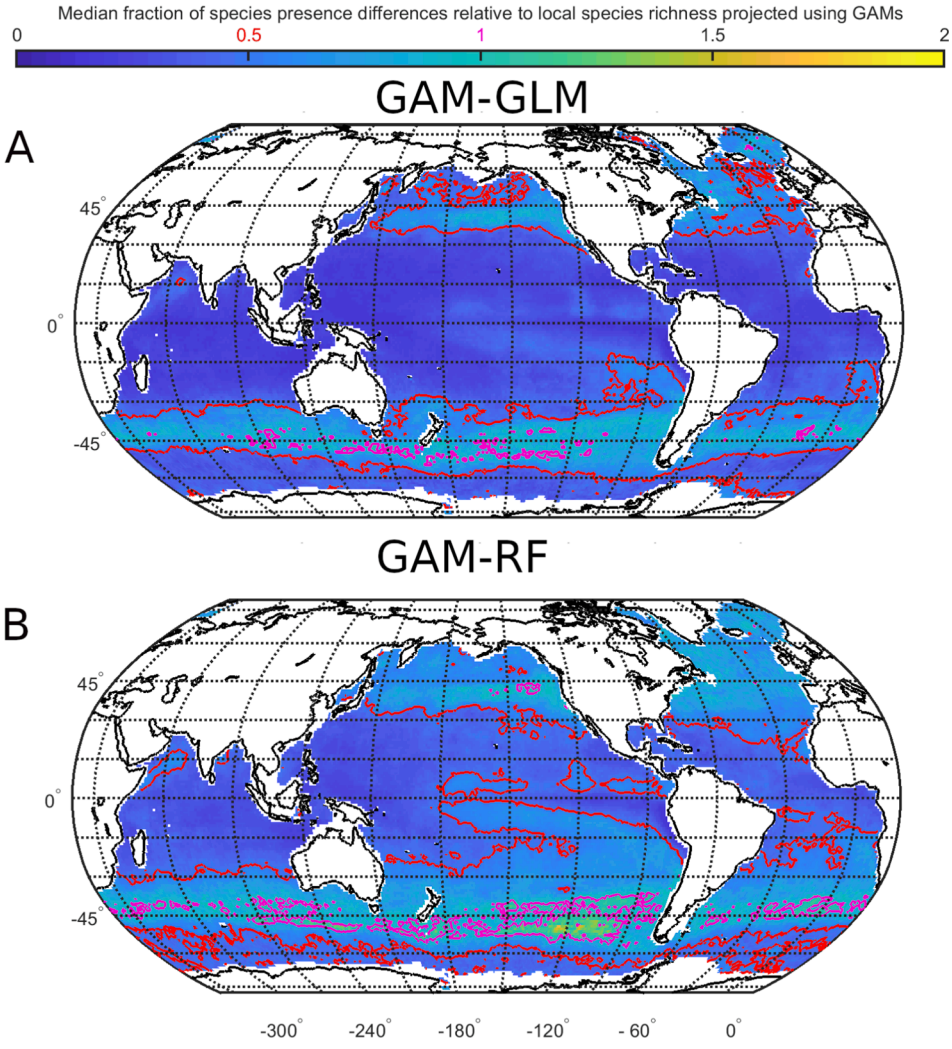


Fig. A.11. A shows the median fraction of species presence differences between the species presence projected using General Linear Models (GLMs) relative to those projected using GAMs. B compares Random Forests (RFs) to GAMs. Values represent monthly median fraction of species presence differences normalized by the local (pixel-wise) monthly species richness projected using General Additive Models (GAMs). For instance, 0.5 indicates that half of the local species richness is either missing or previously absent species are now projected as present. We show in red the contour line for 0.5 median fraction of species, and in magenta the contour line for 1 median fraction of species. The median fraction of species presence differences was calculated for each 1°-pixel and across all twelve month.

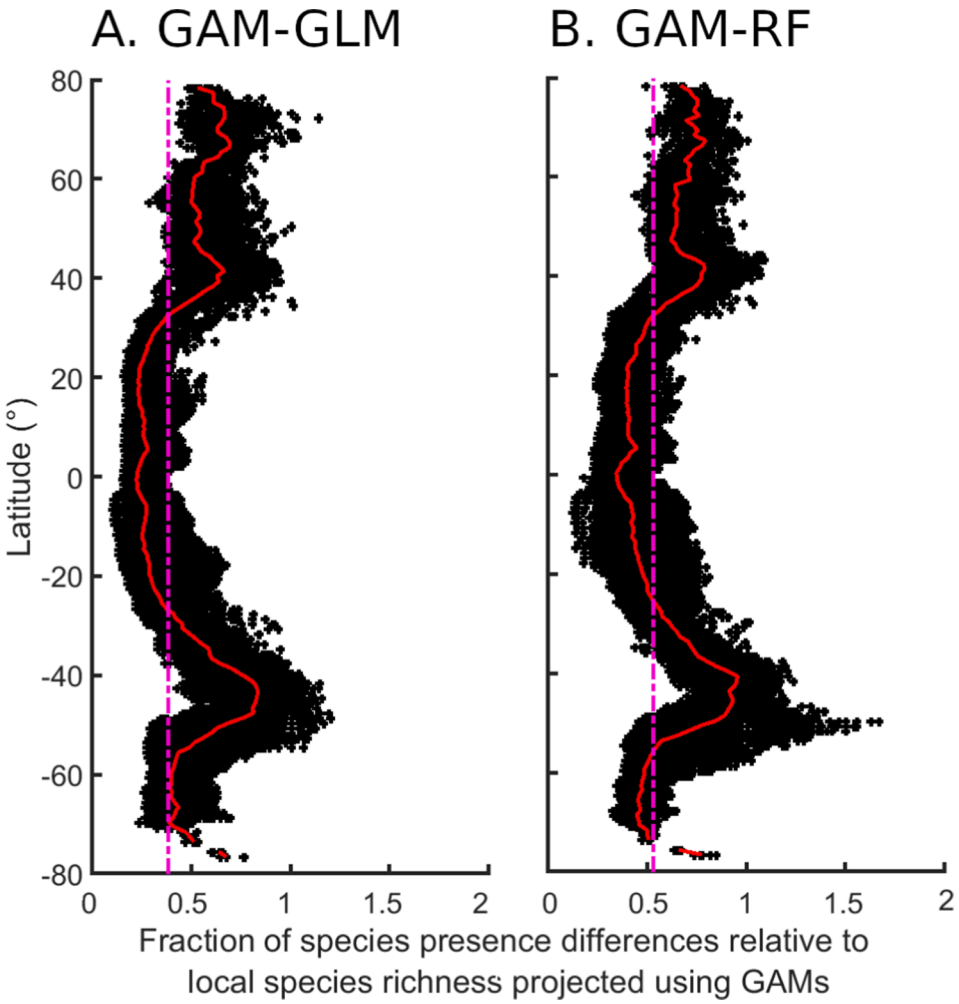
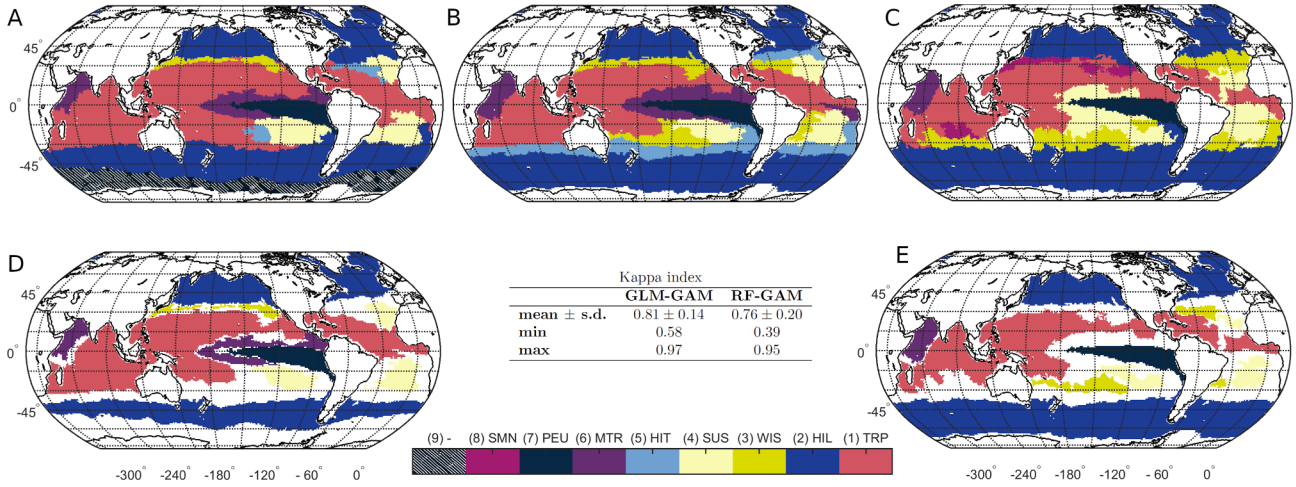


Fig. A.12. A shows the latitudinal gradients of the median fraction of species presence differences between the species presence projected using General Linear Models (GLMs) relative to those projected using GAMs. B shows the same as A but for Random Forests (RFs) instead of GLMs. The black dots show the amplitude of the median fraction of species presence differences relative to the local species richness projected using GAMs across all months. The red line shows the median of all black dots at each latitude. The magenta line shows the global median. Latitudinal gradients of monthly median fraction of species presence differences normalized by the local (pixel-wise), monthly species richness projected using General Additive Models (GAMs). For instance, 0.5 indicates that half of the local species richness is either missing or previously absent species are now projected as present.



Kappa index

	GLM-GAM	RF-GAM
mean ± s.d.	0.81 ± 0.14	0.76 ± 0.20
min	0.58	0.39
max	0.97	0.95

Fig. A.13. Similarity of biomes obtained from species presence distribution patterns projected by General Linear Models (GLMs), General Additive Models (GAMs), and Random Forests (RFs). Similarity is shown as the mean and standard deviation (s.d.) of the twelve monthly pair-wise comparisons of the year. A, B, and C show the annual biomes obtained by our trained self-organizing map with species presence patterns predicted by GLM, GAMs, and RFs, respectively. D shows in color where annual biomes using GLM-predictions overlap with the biomes using GAM-predictions. E compares the RF-predictions to GLM-predictions.

A.6. Uncertainty in species' presence projections due to variable choice

To visualize the uncertainty in presence projections induced by predictor variable choice, we mapped the projection error calculated as the standard deviation of the presence projections of all ensemble members (occurring within individual species) averaged across all 536 species, at monthly 1° × 1° resolution (Fig. A.14). For species with five ensemble members, the maximum value that this uncertainty measure reaches is 0.49,

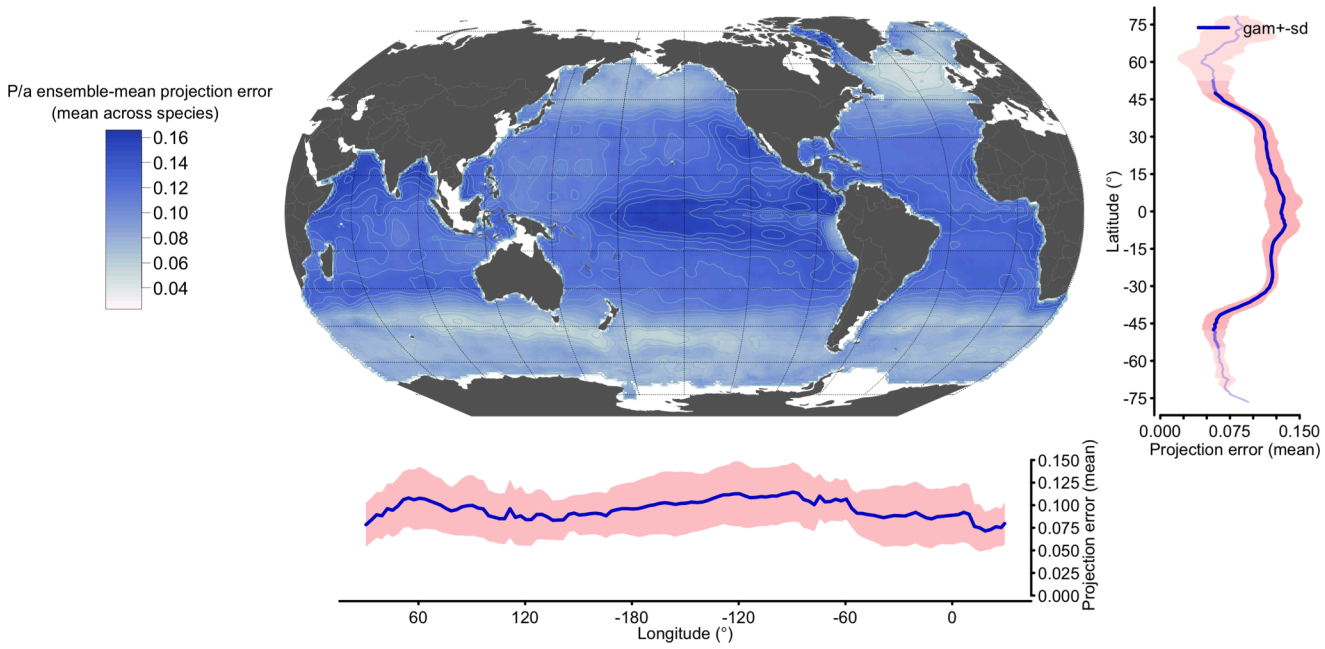


Fig. A.14. Projection error calculated as the standard deviation of the presence projections of species' ensemble members (GAMs) averaged across 536 species. Panels to the right and bottom show the latitudinal and longitudinal patterns of this uncertainty measure on the right, and below the global map, respectively. The analysis was performed at monthly 1° resolution, showing the annual mean of the twelve months.

which indicates that two out of five (or three out of five) ensemble members project a presence and the remaining members an absence, on average across species.

Analyzed by latitude, predictor based uncertainty is highest around the equator, with a maximum value of 0.13 at 5°S. At the resolution of 1° longitude bands, the uncertainty reaches local minima of 0.05, and 0.06 at 60° in the Northern Hemisphere, and at 42° in the Southern Hemisphere, respectively. Longitudinally, the predictor based uncertainty is lowest in the Western Pacific (0.07) and highest in the Eastern Pacific (0.12). At the biome scale, the MTR (0.14 ± 0.01) and PEU (0.14 ± 0.02) biomes are associated with the largest predictor induced uncertainty in species' presence projections (Table A.7), followed by the seasonally alternating subtropical biomes (0.13 ± 0.02 and 0.13 ± 0.01 for the WIS, and SUS biome). The smallest uncertainty is found in the high latitudes, i.e. in the HIL biome (0.07 ± 0.03). Thus, contrary to the uncertainty analyzed as a function of algorithm choice (A.5), we find that variable induced uncertainty (at the level of individual species projection) is weakest between 45° and 60° North and South, and generally enhanced between 35° North and 35° South. Yet, a similar feature of uncertainty, as found for algorithm based uncertainty, is found in the equatorial Pacific.

Table A.7
Monthly means and standard deviations of variable-induced uncertainty in species' presence projections across the 1° pixels per biome. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	Uncertainty
TRP	0.11 ± 0.03
HIL	0.07 ± 0.03
WIS	0.13 ± 0.02
SUS	0.13 ± 0.01
HIT	0.11 ± 0.03
MTR	0.14 ± 0.01
PEU	0.14 ± 0.02
SMN	0.12 ± 0.02

A.7. Monthly biome area

Table A.8.

Table A.8
Relative area coverage of biomes at the monthly scale of analysis. Biomes are sorted according to their mean area coverage across months where the biome is present. Note that the Southern Hemisphere was shifted by six months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon. Also included is the ninth biome (9), which is only present for four months throughout the year and is absent in both the seasonal and annual scale.

Month	(1) TRP	(2) HIL	(3) WIS	(4) SUS	(5) HIT	(6) MTR	(7) PEU	(8) SMN	(9) -
1	6.5	24.4	35.2	0.0	14.4	10.0	4.7	2.5	2.3
2	11.0	29.9	27.2	0.0	15.6	9.6	4.4	1.5	0.7
3	23.5	30.9	14.4	0.0	16.4	9.2	4.4	1.3	0.0
4	30.5	29.0	3.6	9.8	13.5	8.5	4.1	1.0	0.0
5	32.2	27.0	0.6	20.4	7.3	6.9	4.4	1.1	0.0
6	33.3	24.9	0.0	23.4	5.9	7.3	4.7	0.5	0.0
7	37.8	25.3	0.0	19.7	4.6	6.4	4.1	1.5	0.6
8	41.8	26.1	0.0	14.2	4.5	8.1	4.3	1.0	0.0
9	43.0	27.2	3.3	10.8	4.8	7.1	3.9	0.0	0.0
10	32.2	21.3	23.0	3.6	7.1	9.1	3.8	0.0	0.0
11	20.4	20.1	33.3	0.0	10.7	10.5	5.0	0.0	0.0
12	9.2	18.2	38.6	0.0	14.3	12.7	5.0	0.7	1.3

A.8. Number of species found per biome

Tables A.9–A.12.

Table A.9
Total number of species found in our biomes, integrating the species' projected present across the months per biome. Note that the Southern Hemisphere was shifted by six months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon. Also included is the ninth biome (9), which is only present for four months throughout the year and is absent in both the seasonal and annual scale.

Month	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN	(9) -
1	478	351	456	NaN	490	511	478	396	450
2	494	402	498	NaN	509	515	482	412	426
3	511	398	459	NaN	505	495	495	395	NaN
4	513	451	431	487	481	502	496	342	NaN
5	515	472	384	499	483	473	482	371	NaN
6	513	493	NaN	513	465	502	489	374	NaN
7	516	497	NaN	513	446	508	476	466	391
8	510	514	NaN	511	492	512	482	404	NaN
9	515	507	424	507	477	512	474	NaN	NaN
10	505	439	475	349	504	507	464	NaN	NaN
11	498	375	462	NaN	484	502	470	NaN	NaN
12	508	354	490	NaN	484	505	469	334	461

Table A.10
Total number of *Dinoflagellata* species found in our biomes, integrating the species' projected present across the months per biome. Note that the Southern Hemisphere was shifted by six months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon. Also included is the ninth biome (9), which is only present for four months throughout the year and is absent in both the seasonal and annual scale.

Month	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN	(9) -
1	232	130	218	NaN	239	248	235	182	218
2	240	165	241	NaN	250	252	239	189	208
3	249	164	217	NaN	246	244	245	178	NaN
4	253	196	205	243	232	249	249	149	NaN
5	252	214	178	249	239	239	239	173	NaN
6	251	232	NaN	252	226	247	243	172	NaN
7	253	230	NaN	252	222	249	238	226	185
8	247	243	NaN	252	239	253	241	189	NaN
9	249	241	195	249	234	250	239	NaN	NaN
10	248	189	232	179	243	249	237	NaN	NaN
11	243	151	224	NaN	236	246	236	NaN	NaN
12	249	137	236	NaN	234	242	232	157	227

Table A.11
Total number of *Bacillariophyceae* species found in our biomes, integrating the species' projected present across the months per biome. Note that the Southern Hemisphere was shifted by six months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon. Also included is the ninth biome (9), which is only present for four months throughout the year and is absent in both the seasonal and annual scale.

Month	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN	(9) -
1	206	197	200	NaN	214	220	204	185	194
2	212	211	214	NaN	221	219	203	189	182

(*continued on next page*)

Table A.11 (*continued*)

Month	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN	(9) -
3	218	203	201	NaN	217	209	207	185	NaN
4	216	218	188	201	208	212	205	160	NaN
5	218	218	174	209	203	196	201	164	NaN
6	219	220	NaN	216	196	213	203	175	NaN
7	218	224	NaN	216	184	215	198	203	180
8	218	227	NaN	215	210	215	198	183	NaN
9	222	226	190	215	203	218	194	NaN	NaN
10	216	218	203	138	221	214	188	NaN	NaN
11	213	197	201	NaN	214	215	193	NaN	NaN
12	215	195	212	NaN	215	219	201	158	199

Table A.12
Total number of *Haptophyta* species found in our biomes, integrating the species' projected present across the months per biome. Note that the Southern Hemisphere was shifted by six months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon. Also included is the ninth biome (9), which is only present for four months throughout the year and is absent in both the seasonal and annual scale.

Month	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN	(9) -
1	28	14	27	NaN	25	29	28	19	28
2	30	15	30	NaN	26	30	28	23	28
3	31	20	30	NaN	31	30	29	22	NaN
4	31	26	27	31	31	30	29	24	NaN
5	31	29	23	30	31	30	29	25	NaN
6	31	30	NaN	31	31	30	30	20	NaN
7	31	31	NaN	31	30	30	28	26	19
8	31	32	NaN	31	31	30	30	23	NaN
9	31	28	27	30	30	30	29	NaN	NaN
10	30	21	28	27	28	30	28	NaN	NaN
11	30	16	26	NaN	23	30	29	NaN	NaN
12	30	13	29	NaN	22	30	25	12	27

A.9. Seasonal changes in biomes

To characterize the seasonal distribution of our biomes relative to their annual distribution, we compared each $1^\circ \times 1^\circ$ pixel in the seasonal scale of distribution to the annual distribution. For pixels where the associated biome was not equal, we assigned a missing value, and we calculated the total missing area per season (Fig. A.15). This analysis showed that the largest seasonal changes relative to the annual distributions are found in the

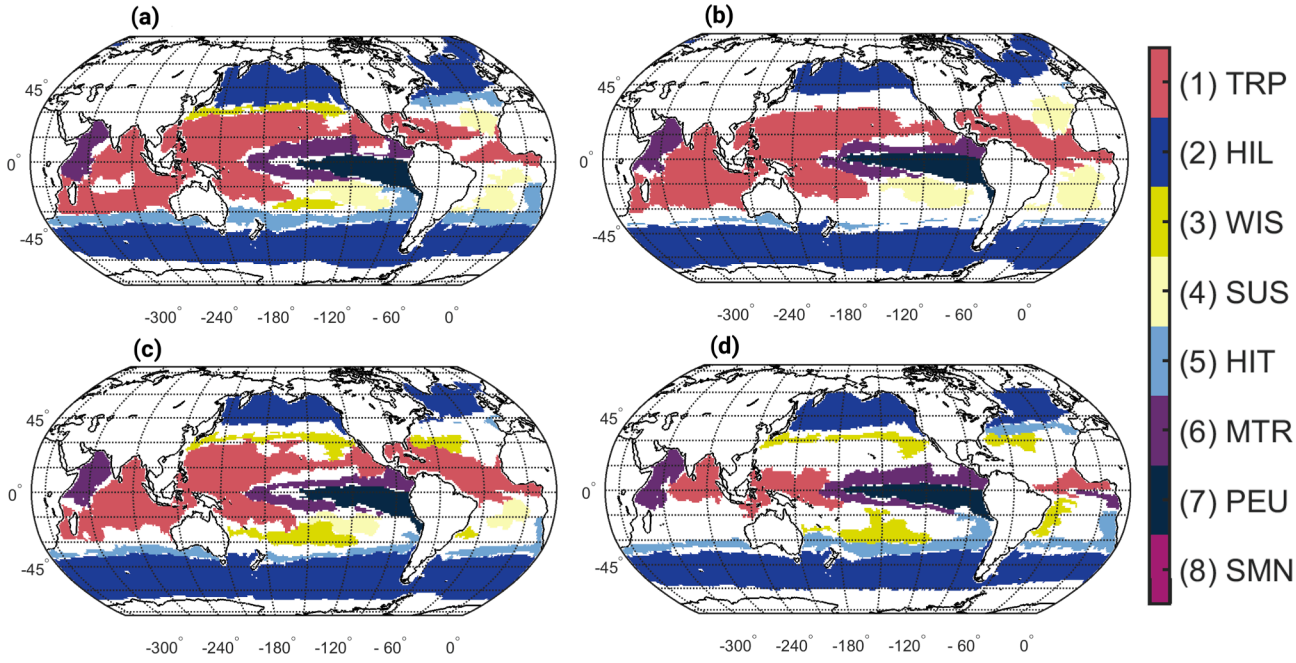


Fig. A.15. Difference between the annual and seasonal distributions of our biomes during (a) spring, (b) summer, (c) fall, and (d) winter. White spaces denote $1^\circ \times 1^\circ$ pixels where the association to a biome is different between the annual and seasonal distributions of our biomes. The Southern Hemisphere was shifted by six months for the calculation of the seasonally integrated biomes based on the monthly scale distributions of our biomes. For each 1°-pixel, the distribution of each biome was calculated based on the monthly scale distribution of the biomes, where for each 1°-pixel the most frequent, unique biome was assigned as its annual biome, and the final location of the annual biomes was determined by reducing the patchiness of the biomes (see Section 2.4). Biomes are sorted according to their mean area coverage across months. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

subtropics, especially during winter and summer.

A.10. NMDS

Tables A.13–A.15.

Table A.13
Degree of overlap between biomes after projecting the trained neurons onto two dimensions using the NMDS. The percentages denote the fraction of neurons associated with one biome that were also located within the envelope of another biome. Marked in blue are the comparisons with an overlap above 50.0%, and marked in yellow are the comparisons with an overlap below 50.0%. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
TRP	NaN	NaN	NaN	4.08%	NaN	57.15%	NaN	NaN
HIL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WIS	NaN	NaN	NaN	NaN	NaN	6.76%	NaN	10.96%
SUS	20.60%	NaN	NaN	NaN	NaN	96.70%	NaN	NaN
HIT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MTR	7.11%	NaN	5.08%	29.69%	NaN	NaN	4.79%	NaN
PEU	NaN	NaN	NaN	NaN	NaN	7.05%	NaN	NaN
SMN	NaN	NaN	4.15%	NaN	NaN	NaN	NaN	NaN

Table A.14
Spread of the envelopes of our biomes along the two dimensions resulting from the NMDS analysis. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	Dimension 1	Dimension 2
TRP	81.4	55.5
HIL	37.4	21.2
WIS	95.4	57.0
SUS	106.5	63.8
HIT	71.0	55.0
MTR	130.1	141.7
PEU	112.8	109.2
SMN	88.7	30.5

Table A.15
Mean overlap in species' presence patterns between biomes. For each month we calculated the overlap in phytoplankton species presence between each pixel in one biome and all pixels in another biome. From all comparisons we calculated the mean overlap on a monthly scale, and from the monthly averages we calculated an overall average. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
TRP		0.38	0.84	0.83	0.74	0.79	0.63	0.81
HIL	0.38		0.31	0.32	0.25	0.33	0.33	0.42
WIS	0.84	0.31		0.72	0.67	0.69	0.50	0.77
SUS	0.83	0.32	0.72		0.71	0.74	0.61	0.65
HIT	0.74	0.25	0.67	0.71		0.63	0.49	0.63
MTR	0.79	0.33	0.69	0.74	0.63		0.69	0.71
PEU	0.63	0.33	0.50	0.61	0.49	0.69		0.54
SMN	0.81	0.42	0.77	0.65	0.63	0.71	0.54	

A.11. Core species and indicator species

Tables A.16–A.19.
Fig. A.16

Table A.16
Total number of core, subordinate, and satellite species per biome found across all months. Core species are defined as species found in > 90% of the biome area on average, satellite species as those with an area coverage ⩽10%, and subordinate species as those between these two thresholds (Table A.18). The core/subordinate/satellite species were identified for each month and each biome. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Biome	*Core*	Subordinate	Satellite
TRP	190	336	194
HIL	1	209	511
WIS	135	353	291
SUS	148	338	272
HIT	20	339	473
MTR	128	410	201
PEU	153	373	211
SMN	246	380	277

Table A.17
Spearman correlation coefficients between the global and biome specific bar-code-like pattern showing the biogeography of the 536 phytoplankton species on the global or biome scale. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

	global	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
global	1.00	0.95	0.04	0.90	0.91	0.78	0.94	0.57	0.85
TRP	0.95	1.00	−0.01	0.92	0.87	0.71	0.89	0.41	0.88
HIL	0.04	−0.01	1.00	0.10	−0.06	0.27	−0.08	−0.21	0.14
WIS	0.90	0.92	0.10	1.00	0.82	0.81	0.77	0.24	0.87
SUS	0.91	0.87	−0.06	0.82	1.00	0.79	0.85	0.49	0.65
HIT	0.78	0.71	0.27	0.81	0.79	1.00	0.65	0.26	0.63
MTR	0.94	0.89	−0.08	0.77	0.85	0.65	1.00	0.68	0.73
PEU	0.57	0.41	−0.21	0.24	0.49	0.26	0.68	1.00	0.28
SMN	0.85	0.88	0.14	0.87	0.65	0.63	0.73	0.28	1.00

Table A.18
List of 536 phytoplankton species modelled. Species are sorted based on their global normalized area coverage in descending order. The numbers indicate the average normalized area coverage of each species and biome, separately based on the monthly area coverage of the respective species. The following abbreviations were used: Phy for phylum; cl for class; din for *Dinoflagellata*; bac for *Bacillariophyceae*; hap for *Haptophyta*; chl for *Chlorophyta*; cya for *Cyanobacteria*; cry for *Cryptophyta*; dic for Dictyochophyceae; chr for *Chrysophyceae*; eug for *Euglenozoa*; TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
din	*Tripos extensus*	1.00	0.11	1.00	1.00	0.96	1.00	1.00	0.98
bac	*Nitzschia bicapitata*	0.98	0.26	0.94	0.97	0.75	1.00	0.98	0.90
din	*Gonyaulax spinifera*	1.00	0.24	0.87	1.00	0.46	1.00	1.00	0.93
din	*Prorocentrum lima*	0.99	0.20	0.91	0.99	0.45	0.95	0.98	0.98
din	*Impagidinium patulum*	0.97	0.30	0.89	0.98	0.63	0.93	0.84	0.83
din	*Oxytoxum scolopax*	0.99	0.05	0.89	1.00	0.52	0.99	0.95	0.97
din	*Impagidinium sphaericum*	0.94	0.42	0.91	0.95	0.55	0.85	0.86	0.88
bac	*Rhizosolenia setigera*	0.99	0.05	0.91	0.91	0.54	0.95	0.99	0.98
bac	*Nitzschia longissima*	0.98	0.26	0.91	0.91	0.43	0.89	0.93	0.99
bac	*Chaetoceros hyalochaetae*	0.97	0.07	0.94	0.97	0.79	0.92	0.79	0.81
bac	*Pseudosolenia calcar-avis*	1.00	0.04	0.98	0.98	0.56	0.94	0.77	0.98
din	*Impagidinium aculeatum*	0.96	0.31	0.85	0.96	0.55	0.92	0.87	0.73
din	*Lingulodinium polyedra*	0.99	0.02	0.80	0.99	0.38	1.00	1.00	0.91
din	*Podolampas spinifera*	0.99	0.02	0.86	0.98	0.35	0.97	0.97	0.92
bac	*Chaetoceros peruvianus*	0.99	0.06	0.96	0.89	0.18	0.99	0.99	0.98
din	*Prorocentrum dentatum*	0.99	0.07	0.86	0.99	0.52	0.88	0.72	0.99
din	*Tripos pentagonus*	0.96	0.16	0.80	0.94	0.30	0.90	0.97	1.00
bac	*Chaetoceros curvisetus*	0.98	0.03	0.98	0.85	0.35	0.88	0.96	0.99
din	*Dinophysis schuettii*	0.99	0.02	0.87	0.99	0.41	0.96	0.96	0.81
bac	*Cerataulina pelagica*	0.99	0.02	0.95	0.83	0.25	0.99	0.98	0.98
din	*Oxytoxum variabile*	0.97	0.11	0.65	0.94	0.57	0.96	0.93	0.83
bac	*Leptocylindrus danicus*	0.99	0.02	0.95	0.96	0.40	0.88	0.78	0.97

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
din	*Prorocentrum rostratum*	0.99	0.02	0.69	0.99	0.38	0.98	0.91	0.97
din	*Protoceratium reticulatum*	0.96	0.33	0.74	0.93	0.41	0.93	0.79	0.84
cya	*Prochlorococcus*	0.99	0.02	0.93	0.98	0.53	0.97	0.65	0.82
din	*Ceratium setaceum*	0.98	0.01	0.79	0.98	0.40	0.98	1.00	0.75
hap	*Umbilicosphaera sibogae*	0.99	0.03	0.99	1.00	0.62	0.94	0.30	0.99
din	*Tripos belone*	0.98	0.02	0.75	0.97	0.24	0.97	1.00	0.89
bac	*Chaetoceros coarctatus*	0.98	0.01	0.86	0.85	0.17	0.99	1.00	0.93
bac	*Chaetoceros laciniosus*	0.98	0.22	0.97	0.92	0.34	0.82	0.55	0.99
cya	*Synechococcus*	0.87	0.41	0.67	0.80	0.31	0.97	0.78	0.96
din	*Prorocentrum scutellum*	0.96	0.04	0.95	0.98	0.67	0.82	0.37	0.98
din	*Podolampas elegans*	0.98	0.01	0.88	0.94	0.31	0.89	0.90	0.84
din	*Ceratium contrarium*	0.99	0.01	0.89	0.95	0.22	0.96	0.76	0.97
din	*Tripos teres*	0.99	0.01	0.98	0.99	0.62	0.90	0.29	0.95
din	*Podolampas palmipes*	0.90	0.01	0.75	0.97	0.37	0.86	0.99	0.85
bac	*Chaetoceros compressus*	0.97	0.03	0.94	0.86	0.15	0.85	0.95	0.92
bac	*Climacodium frauenfeldianum*	0.99	0.02	0.97	0.86	0.39	0.90	0.53	0.99
din	*Oxytoxum parvum*	0.98	0.06	0.62	0.98	0.38	0.94	0.72	0.97
din	*Gonyaulax scrippsae*	0.97	0.01	0.93	0.97	0.45	0.88	0.54	0.86
din	*Trinovantidinium applanatum*	0.92	0.05	0.73	0.89	0.61	0.89	0.77	0.69
din	*Pentapharsodinium dalei*	0.77	0.29	0.52	0.84	0.47	0.96	0.99	0.68
bac	*Shionodiscus oestrupii*	0.96	0.27	0.85	0.95	0.48	0.83	0.19	0.98
hap	*Discosphaera tubifera*	0.96	0.07	0.84	0.98	0.38	0.79	0.58	0.88
din	*Tripos declinatus*	0.97	0.03	0.70	0.95	0.45	0.77	0.72	0.88
bac	*Pleurosigma normanii*	0.95	0.05	0.82	0.87	0.51	0.90	0.60	0.75
din	*Gonyaulax membranacea*	0.95	0.07	0.60	0.82	0.40	0.95	0.95	0.71
bac	*Rhizosolenia acuminata*	0.89	0.21	0.62	0.41	0.50	0.89	0.97	0.97
bac	*Rhizosolenia bergonii*	0.89	0.10	0.54	0.57	0.63	0.84	0.99	0.89
din	*Tripos carriensis*	0.97	0.02	0.75	0.98	0.42	0.93	0.90	0.47
bac	*Eucampia cornuta*	0.96	0.18	0.90	0.74	0.32	0.73	0.62	0.99
bac	*Hemiaulus hauckii*	0.99	0.01	0.96	0.94	0.25	0.87	0.38	0.98
din	*Oxytoxum sphaeroideum*	0.79	0.06	0.59	0.89	0.54	0.90	0.76	0.85
bac	*Pleurosigma elongatum*	0.90	0.06	0.97	0.93	0.71	0.74	0.18	0.89
din	*Protoperidinium latidorsale*	0.94	0.01	0.74	0.79	0.27	0.98	0.90	0.73
bac	*Guinardia flaccida*	0.96	0.07	0.93	0.86	0.31	0.67	0.64	0.91
bac	*Chaetoceros anastomosans*	0.97	0.03	0.93	0.69	0.43	0.86	0.43	1.00
din	*Prorocentrum cordatum*	0.86	0.58	0.46	0.75	0.34	0.77	0.67	0.87
din	*Gonyaulax digitalis*	0.95	0.07	0.51	0.88	0.31	0.92	0.89	0.76
bac	*Chaetoceros lorenzianus*	0.96	0.02	0.63	0.61	0.20	0.93	0.97	0.97
din	*Oxytoxum longiceps*	0.96	0.00	0.60	0.85	0.10	0.99	0.98	0.78
din	*Dinophysis parvula*	0.96	0.01	0.85	0.97	0.26	0.78	0.54	0.88
din	*Oxytoxum turbo*	0.91	0.01	0.88	0.96	0.44	0.72	0.57	0.75
bac	*Thalassiosira eccentrica*	0.90	0.42	0.52	0.39	0.12	0.97	0.98	0.93
bac	*Climacosphenia moniligera*	0.95	0.01	0.61	0.61	0.28	0.90	0.90	0.95
bac	*Planktoniella sol*	0.93	0.02	0.55	0.74	0.24	0.93	1.00	0.79
bac	*Detonula pumila*	0.96	0.01	0.82	0.72	0.14	0.88	0.78	0.88
din	*Thoracosphaera heimii*	0.98	0.00	0.78	0.94	0.18	0.98	0.52	0.79
bac	*Guinardia striata*	0.95	0.10	0.84	0.37	0.38	0.73	0.82	0.96
din	*Pronoctiluca pelagica*	0.92	0.02	0.48	0.93	0.22	0.88	0.96	0.71
bac	*Rhizosolenia castracanei*	0.95	0.02	0.62	0.59	0.17	0.84	0.96	0.97
bac	*Chaetoceros affinis*	0.97	0.02	0.74	0.39	0.18	0.88	0.93	0.99
din	*Ornithocercus magnificus*	0.93	0.01	0.44	0.88	0.09	0.94	0.99	0.82
din	*Ceratoperidinium falcatum*	0.96	0.07	0.63	0.79	0.37	0.90	0.54	0.84
bac	*Pleurosigma directum*	0.85	0.49	0.57	0.52	0.21	0.87	0.82	0.76
din	*Protoperidinium punctulatum*	0.98	0.02	0.68	0.89	0.10	0.89	0.53	0.96
bac	*Chaetoceros danicus*	0.84	0.02	0.83	0.59	0.15	0.76	0.94	0.90
din	*Prorocentrum mexicanum*	0.94	0.00	0.69	0.70	0.03	0.99	0.84	0.85
bac	*Chaetoceros pendulus*	0.96	0.03	0.89	0.81	0.37	0.67	0.31	0.99
bac	*Pleurosigma angulatum*	0.96	0.07	0.98	0.89	0.65	0.54	0.00	0.92
din	*Tripos kofoidii*	0.91	0.02	0.41	0.86	0.16	0.88	0.99	0.76
din	*Ceratium massiliense*	0.82	0.02	0.77	0.99	0.61	0.60	0.78	0.36
bac	*Pseudo-nitzschia pungens*	0.97	0.07	0.81	0.76	0.17	0.76	0.41	0.99
din	*Tripos pulchellus*	0.95	0.01	0.48	0.81	0.11	0.82	0.99	0.77
bac	*Guinardia cylindrus*	0.97	0.12	0.95	0.88	0.31	0.61	0.11	0.97
din	*Protoperidinium ovum*	0.91	0.03	0.53	0.93	0.29	0.76	0.62	0.85
bac	*Chaetoceros tetrastichon*	0.95	0.04	0.76	0.64	0.29	0.84	0.41	0.98
din	*Cladopyxis brachiolata*	0.96	0.00	0.77	0.88	0.10	0.88	0.60	0.71
din	*Scrippsiella regalis*	0.89	0.04	0.90	0.77	0.40	0.85	0.33	0.73
bac	*Chaetoceros brevis*	0.94	0.02	0.71	0.37	0.22	0.89	0.75	0.99
din	*Prorocentrum micans*	0.92	0.01	0.42	0.82	0.09	0.95	0.95	0.72
din	*Prorocentrum balticum*	0.96	0.28	0.58	0.80	0.02	0.89	0.45	0.90
bac	*Actinoptychus senarius*	0.81	0.04	0.72	0.46	0.13	0.90	0.98	0.83
bac	*Chaetoceros dichaeta*	0.92	0.29	0.57	0.42	0.28	0.70	0.67	1.00
din	*Tryblionella compressa*	0.94	0.16	0.57	0.66	0.10	0.96	0.63	0.81
din	*Ceratium trichoceros*	0.72	0.02	0.80	0.97	0.63	0.53	0.95	0.21
bac	*Stephanopyxis palmeriana*	0.91	0.02	0.66	0.42	0.22	0.74	0.89	0.96
din	*Oxytoxum milneri*	0.97	0.01	0.72	0.97	0.24	0.89	0.40	0.62

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
bac	*Chaetoceros wighamii*	0.94	0.01	0.90	0.71	0.23	0.76	0.36	0.92
din	*Ceratium falcatum*	0.90	0.01	0.44	0.78	0.13	0.83	0.98	0.73
bac	*Eucampia zodiacus*	0.90	0.05	0.71	0.41	0.15	0.73	0.87	0.97
bac	*Nitzschia tenuirostris*	0.90	0.02	0.70	0.80	0.26	0.82	0.49	0.75
bac	*Gossleriella tropica*	0.91	0.02	0.66	0.72	0.30	0.65	0.56	0.89
hap	*Umbellosphaera tenuis*	0.80	0.06	0.58	0.98	0.38	0.80	0.74	0.36
bac	*Nitzschia sigma*	0.96	0.04	0.82	0.83	0.21	0.74	0.20	0.89
din	*Tripos karstenii*	0.97	0.01	0.84	0.94	0.29	0.61	0.31	0.71
din	*Amphisolenia bidentata*	0.91	0.00	0.43	0.64	0.03	0.93	0.97	0.73
bac	*Chaetoceros diversus*	0.98	0.04	0.93	0.77	0.19	0.65	0.17	0.92
din	*Protoperidinium mediterraneum*	0.96	0.01	0.77	0.93	0.20	0.70	0.28	0.74
bac	*Asteromphalus heptactis*	0.88	0.14	0.55	0.36	0.08	0.82	0.96	0.80
bac	*Asteromphalus cleveanus*	0.95	0.04	0.88	0.74	0.46	0.52	0.01	0.96
din	*Gonyaulax monacantha*	0.83	0.01	0.56	0.97	0.35	0.73	0.79	0.32
bac	*Mastogloia rostrata*	0.94	0.01	0.96	0.85	0.27	0.54	0.02	0.93
din	*Corythodinium tesselatum*	0.90	0.02	0.43	0.93	0.13	0.76	0.76	0.59
bac	*Chaetoceros decipiens*	0.95	0.02	0.76	0.40	0.13	0.75	0.58	0.92
din	*Protoperidinium abei*	0.89	0.03	0.20	0.68	0.24	0.83	0.80	0.82
hap	*Gladiolithus flabellatus*	0.88	0.06	0.38	0.92	0.32	0.77	0.68	0.49
hap	*Umbilicosphaera hulburtiana*	0.80	0.09	0.55	0.99	0.50	0.65	0.25	0.63
bac	*Azpeitia nodulifera*	0.92	0.00	0.77	0.80	0.07	0.79	0.26	0.84
bac	*Hemiaulus chinensis*	0.95	0.02	0.87	0.69	0.13	0.54	0.34	0.90
bac	*Chaetoceros messanensis*	0.96	0.04	0.86	0.62	0.33	0.53	0.12	0.98
bac	*Ditylum brightwellii*	0.73	0.04	0.74	0.42	0.25	0.86	0.53	0.83
bac	*Rhizosolenia formosa*	0.92	0.00	0.49	0.26	0.04	0.88	0.86	0.95
din	*Echinidinium delicatum*	0.72	0.08	0.48	0.86	0.28	0.80	0.75	0.43
din	*Ceratocorys horrida*	0.85	0.00	0.25	0.84	0.09	0.89	1.00	0.47
din	*Lebouridinium glaucum*	0.82	0.04	0.64	0.79	0.28	0.85	0.52	0.44
din	*Oxytoxum tesselatum*	0.80	0.02	0.78	0.94	0.50	0.48	0.16	0.70
bac	*Bacteriastrum delicatulum*	0.91	0.02	0.81	0.61	0.11	0.46	0.51	0.94
bac	*Thalassionema frauenfeldii*	0.96	0.03	0.84	0.72	0.27	0.52	0.05	0.97
bac	*Asterolampra marylandica*	0.94	0.06	0.94	0.73	0.19	0.52	0.05	0.92
din	*Protoperidinium tuba*	0.89	0.00	0.58	0.92	0.12	0.67	0.65	0.45
din	*Ceratium falcatiforme*	0.81	0.01	0.34	0.63	0.08	0.78	0.92	0.72
bac	*Chaetoceros dadayi*	0.97	0.01	0.83	0.63	0.14	0.68	0.06	0.96
bac	*Lauderia annulata*	0.88	0.02	0.62	0.32	0.08	0.94	0.71	0.70
bac	*Haslea wawrikae*	0.94	0.01	0.85	0.72	0.21	0.50	0.04	0.98
din	*Leonella granifera*	0.92	0.00	0.58	0.86	0.11	0.89	0.21	0.68
din	*Oxytoxum elegans*	0.85	0.01	0.49	0.77	0.29	0.60	0.85	0.37
hap	*Hayaster perplexus*	0.82	0.02	0.49	0.91	0.15	0.92	0.87	0.05
din	*Oxytoxum curvatum*	0.80	0.09	0.28	0.89	0.49	0.55	0.65	0.48
bac	*Aulacoseira granulata*	0.76	0.01	0.72	0.81	0.47	0.56	0.20	0.70
din	*Gyrodinium fusiforme*	0.83	0.14	0.39	0.48	0.07	0.88	0.86	0.56
bac	*Chaetoceros costatus*	0.90	0.01	0.64	0.45	0.11	0.58	0.58	0.95
bac	*Striatella unipunctata*	0.95	0.02	0.61	0.91	0.19	0.54	0.18	0.81
din	*Dinophysis fortii*	0.94	0.04	0.55	0.90	0.18	0.55	0.07	0.98
bac	*Chaetoceros aequatorialis*	0.92	0.02	0.76	0.60	0.29	0.53	0.04	0.99
bac	*Bacteriastrum furcatum*	0.97	0.03	0.79	0.52	0.11	0.64	0.11	0.98
bac	*Asteromphalus flabellatus*	0.80	0.02	0.33	0.26	0.08	0.85	0.96	0.81
din	*Tripos candelabrum*	0.55	0.03	0.39	0.92	0.49	0.57	0.84	0.33
bac	*Chaetoceros didymus*	0.88	0.11	0.51	0.32	0.12	0.58	0.62	0.97
hap	*Umbellosphaera irregularis*	0.91	0.01	0.82	0.96	0.28	0.72	0.05	0.36
din	*Tripos contortus*	0.82	0.01	0.27	0.67	0.05	0.81	0.93	0.54
din	*Gymnodinium catenatum*	0.80	0.02	0.51	0.81	0.22	0.81	0.23	0.65
bac	*Leptocylindrus minimus*	0.93	0.03	0.65	0.42	0.25	0.64	0.12	0.99
bac	*Climacodium biconcavum*	0.91	0.02	0.78	0.65	0.18	0.48	0.03	0.99
bac	*Meuniera membranacea*	0.95	0.03	0.83	0.50	0.17	0.50	0.04	1.00
din	*Ornithocercus heteroporus*	0.95	0.01	0.75	0.87	0.14	0.57	0.03	0.70
hap	*Florisphaera profunda*	0.61	0.12	0.36	0.88	0.26	0.82	0.67	0.29
bac	*Chaetoceros atlanticus*	0.64	0.45	0.19	0.16	0.10	0.67	0.96	0.83
bac	*Chaetoceros pelagicus*	0.77	0.22	0.39	0.84	0.13	0.45	0.86	0.34
din	*Tripos deflexus*	0.85	0.03	0.23	0.30	0.06	0.82	0.89	0.80
bac	*Bacteriastrum comosum*	0.94	0.01	0.87	0.60	0.14	0.45	0.07	0.90
din	*Protoperidinium brochii*	0.73	0.02	0.12	0.59	0.15	0.81	0.98	0.58
bac	*Chaetoceros phaeoceros*	0.75	0.03	0.86	0.82	0.37	0.61	0.06	0.47
din	*Oxytoxum caudatum*	0.69	0.01	0.31	0.89	0.28	0.72	0.85	0.22
bac	*Chaetoceros pseudocurvisetus*	0.94	0.04	0.75	0.30	0.16	0.60	0.21	0.95
bac	*Thalassiosira aestivalis*	0.56	0.01	0.55	0.82	0.17	0.59	0.89	0.35
din	*Gyrodinium spirale*	0.84	0.03	0.59	0.89	0.43	0.61	0.07	0.48
din	*Gonyaulax birostris*	0.92	0.00	0.61	0.88	0.08	0.73	0.11	0.58
bac	*Guinardia delicatula*	0.82	0.00	0.52	0.23	0.08	0.88	0.64	0.68
din	*Spiraulax kofoidii*	0.68	0.01	0.12	0.55	0.08	0.86	0.93	0.62
din	*Goniodoma sphaericum*	0.80	0.03	0.40	0.74	0.28	0.71	0.65	0.26
bac	*Bacteriastrum hyalinum*	0.91	0.06	0.62	0.48	0.17	0.48	0.21	0.91
din	*Dinophysis schroederi*	0.75	0.01	0.78	0.93	0.43	0.46	0.04	0.44
bac	*Coscinodiscus radiatus*	0.66	0.06	0.27	0.31	0.03	0.88	0.98	0.64

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
hap	*Rhabdolithes claviger*	0.65	0.09	0.78	0.92	0.62	0.56	0.08	0.11
din	*Ceratium symmetricum*	0.78	0.00	0.29	0.40	0.07	0.87	0.99	0.42
din	*Gymnodinium marinum*	0.76	0.02	0.31	0.89	0.42	0.62	0.22	0.56
bac	*Hemiaulus membranaceus*	0.95	0.01	0.69	0.36	0.08	0.58	0.14	0.95
din	*Pyrophacus vancampoae*	0.71	0.02	0.13	0.93	0.24	0.65	0.80	0.28
bac	*Membraneis challengeri*	0.74	0.42	0.71	0.62	0.17	0.14	0.10	0.86
bac	*Chaetoceros vanheurckii*	0.92	0.01	0.74	0.51	0.09	0.45	0.05	0.99
din	*Oxytoxum constrictum*	0.48	0.04	0.63	0.86	0.59	0.58	0.50	0.06
din	*Pyrocystis robusta*	0.86	0.00	0.22	0.28	0.03	0.87	0.77	0.71
bac	*Podosira stelligera*	0.34	0.28	0.87	0.58	0.78	0.24	0.04	0.56
bac	*Neocalyptrella robusta*	0.74	0.00	0.25	0.16	0.02	0.85	0.96	0.71
bac	*Chaetoceros lauderi*	0.80	0.01	0.42	0.07	0.04	0.64	0.85	0.87
din	*Amphidinium acutissimum*	0.74	0.00	0.21	0.73	0.04	0.80	0.97	0.19
din	*Protoperidinium quarnerense*	0.65	0.03	0.09	0.69	0.05	0.77	0.95	0.39
bac	*Bacteriastrum elongatum*	0.93	0.02	0.79	0.29	0.13	0.46	0.07	0.93
din	*Protoperidinium pedunculatum*	0.67	0.01	0.11	0.62	0.13	0.79	0.98	0.27
din	*Tripos trichoceros*	0.78	0.02	0.69	0.72	0.33	0.61	0.11	0.33
bac	*Nitzschia sicula*	0.83	0.00	0.68	0.72	0.09	0.72	0.35	0.17
din	*Podolampas bipes*	0.56	0.00	0.12	0.54	0.07	0.74	0.98	0.54
din	*Tripos concilians*	0.62	0.01	0.34	0.78	0.19	0.56	0.87	0.13
din	*Pyrophacus steinii*	0.67	0.00	0.09	0.57	0.03	0.79	0.96	0.39
bac	*Chaetoceros seychellarum*	0.87	0.01	0.80	0.37	0.04	0.35	0.21	0.85
hap	*Helladosphaera cornifera*	0.81	0.05	0.33	0.86	0.31	0.53	0.36	0.24
din	*Gonyaulax pacifica*	0.63	0.00	0.14	0.42	0.03	0.81	0.96	0.50
bac	*Chaetoceros rostratus*	0.88	0.00	0.62	0.32	0.06	0.46	0.16	0.91
din	*Phalacroma mitra*	0.89	0.02	0.26	0.74	0.13	0.59	0.15	0.61
bac	*Bacteriastrum elegans*	0.87	0.01	0.81	0.48	0.20	0.27	0.01	0.74
din	*Ornithocercus quadratus*	0.62	0.02	0.10	0.65	0.03	0.75	0.97	0.24
bac	*Chaetoceros radicans*	0.79	0.08	0.40	0.13	0.08	0.40	0.50	0.97
din	*Pyrophacus horologium*	0.64	0.00	0.10	0.56	0.04	0.65	0.91	0.44
din	*Phalacroma lens*	0.50	0.01	0.26	0.18	0.09	0.59	0.95	0.72
bac	*Chaetoceros paradoxus*	0.74	0.03	0.66	0.18	0.31	0.35	0.04	0.99
din	*Phalacroma favus*	0.71	0.04	0.14	0.76	0.18	0.58	0.62	0.25
bac	*Thalassiosira angulata*	0.73	0.37	0.61	0.10	0.12	0.30	0.06	0.98
hap	*Syracosphaera prolongata*	0.65	0.03	0.68	0.73	0.35	0.38	0.01	0.44
bac	*Dactyliosolen fragilissimus*	0.84	0.02	0.79	0.59	0.10	0.30	0.12	0.49
din	*Gonyaulax polygramma*	0.69	0.01	0.11	0.55	0.02	0.70	0.92	0.25
din	*Karenia brevis*	0.78	0.00	0.41	0.81	0.14	0.47	0.06	0.56
din	*Blepharocysta splendor-maris*	0.89	0.00	0.42	0.87	0.10	0.47	0.04	0.44
din	*Ornithocercus thurnii*	0.76	0.02	0.13	0.80	0.24	0.55	0.49	0.22
din	*Triadinium polyedricum*	0.54	0.02	0.10	0.74	0.18	0.54	0.96	0.14
bac	*Thalassiosira decipiens*	0.75	0.30	0.48	0.05	0.08	0.47	0.09	0.98
bac	*Thalassiothrix heteromorpha*	0.82	0.00	0.65	0.42	0.06	0.31	0.01	0.93
bac	*Trieres chinensis*	0.21	0.41	0.57	0.04	0.48	0.43	0.56	0.45
din	*Oxytoxum sceptrum*	0.81	0.04	0.43	0.88	0.17	0.37	0.05	0.42
bac	*Eucampia zoodiacus*	0.81	0.04	0.62	0.20	0.18	0.27	0.02	0.99
din	*Pronoctiluca spinifera*	0.83	0.01	0.29	0.86	0.03	0.45	0.26	0.40
din	*Protoperidinium pyriforme*	0.76	0.00	0.11	0.64	0.12	0.60	0.58	0.30
din	*Dinophysis argus*	0.48	0.00	0.26	0.59	0.07	0.57	0.95	0.17
din	*Protoperidinium cerasus*	0.77	0.18	0.38	0.28	0.04	0.49	0.02	0.93
din	*Gonyaulax diegensis*	0.67	0.03	0.12	0.75	0.18	0.47	0.53	0.33
din	*Tripos vultur*	0.51	0.00	0.06	0.55	0.07	0.64	1.00	0.23
hap	*Syracosphaera nodosa*	0.55	0.01	0.49	0.95	0.10	0.62	0.19	0.15
din	*Tripos euarcuatus*	0.69	0.01	0.09	0.50	0.04	0.63	0.81	0.29
bac	*Chaetoceros willei*	0.83	0.00	0.66	0.29	0.03	0.30	0.05	0.88
bac	*Rhizosolenia clevei*	0.88	0.00	0.46	0.28	0.01	0.60	0.01	0.78
bac	*Chaetoceros saltans*	0.87	0.00	0.56	0.14	0.07	0.41	0.02	0.94
din	*Dinophysis ovum*	0.59	0.02	0.42	0.82	0.32	0.28	0.18	0.39
din	*Dinophysis porodictyum*	0.67	0.01	0.21	0.64	0.10	0.66	0.68	0.02
bac	*Chaetoceros simplex*	0.58	0.12	0.63	0.10	0.37	0.30	0.01	0.87
din	*Protoperidinium elegans*	0.58	0.04	0.02	0.62	0.07	0.63	0.90	0.11
din	*Ceratium breve*	0.54	0.00	0.09	0.47	0.02	0.63	0.87	0.35
din	*Protoperidinium americanum*	0.34	0.04	0.33	0.38	0.20	0.62	0.67	0.38
din	*Protoperidinium divergens*	0.59	0.02	0.03	0.37	0.07	0.65	0.96	0.21
bac	*Corethron pennatum*	0.55	0.61	0.24	0.09	0.14	0.26	0.14	0.86
hap	*Michaelsarsia elegans*	0.48	0.00	0.07	0.39	0.01	0.77	0.74	0.40
din	*Pyrocystis hamulus*	0.76	0.01	0.39	0.59	0.08	0.52	0.33	0.18
hap	*Syracosphaera molischii*	0.59	0.08	0.18	0.55	0.46	0.58	0.06	0.32
din	*Ceratium gravidum*	0.71	0.01	0.18	0.62	0.07	0.51	0.56	0.18
bac	*Chaetoceros laevis*	0.78	0.01	0.44	0.43	0.06	0.27	0.02	0.83
hap	*Reticulofenestra sessilis*	0.66	0.00	0.19	0.89	0.07	0.39	0.37	0.26
din	*Tripos massiliensis*	0.58	0.02	0.53	0.65	0.29	0.54	0.06	0.14
bac	*Ditylum sol*	0.78	0.00	0.56	0.15	0.14	0.28	0.01	0.87

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
bac	*Coscinodiscus granii*	0.56	0.07	0.23	0.15	0.06	0.47	0.96	0.29
hap	*Algirosphaera robusta*	0.51	0.01	0.15	0.71	0.21	0.73	0.26	0.19
din	*Pyrocystis lunula*	0.69	0.01	0.17	0.35	0.06	0.63	0.43	0.42
bac	*Chaetoceros eibenii*	0.68	0.02	0.58	0.15	0.08	0.15	0.16	0.94
hap	*Calciopappus rigidus*	0.33	0.03	0.17	0.59	0.24	0.72	0.29	0.38
din	*Dinophysis sacculus*	0.43	0.01	0.62	0.82	0.22	0.20	0.07	0.36
bac	*Odontella aurita*	0.20	0.46	0.12	0.04	0.08	0.35	0.62	0.85
din	*Prorocentrum triestinum*	0.73	0.01	0.30	0.43	0.01	0.32	0.07	0.83
bac	*Trieres mobiliensis*	0.36	0.00	0.18	0.06	0.03	0.78	0.78	0.49
bac	*Chaetoceros diadema*	0.76	0.05	0.52	0.08	0.16	0.21	0.04	0.84
hap	*Gephyrocapsa ericsonii*	0.41	0.17	0.29	0.85	0.37	0.24	0.05	0.28
bac	*Amphiprora gigantea*	0.63	0.01	0.65	0.37	0.27	0.23	0.02	0.45
din	*Oxytoxum gracile*	0.58	0.01	0.46	0.91	0.18	0.36	0.03	0.09
bac	*Chaetoceros pseudodichaetus*	0.74	0.00	0.59	0.26	0.05	0.15	0.00	0.82
bac	*Chaetoceros subsecundus*	0.77	0.05	0.29	0.06	0.03	0.42	0.02	0.97
din	*Neoceratium hexacanthum*	0.43	0.03	0.27	0.21	0.33	0.54	0.13	0.62
bac	*Chaetoceros densus*	0.82	0.02	0.59	0.28	0.08	0.26	0.01	0.48
din	*Oxytoxum reticulatum*	0.37	0.02	0.08	0.59	0.20	0.44	0.64	0.16
bac	*Chaetoceros tortissimus*	0.76	0.05	0.47	0.06	0.04	0.31	0.08	0.74
bac	*Pseudo-nitzschia subfraudulenta*	0.62	0.01	0.46	0.70	0.12	0.48	0.05	0.06
din	*Ornithocercus thumii*	0.37	0.00	0.10	0.27	0.01	0.60	0.92	0.22
din	*Dinophysis hastata*	0.53	0.01	0.06	0.50	0.04	0.48	0.70	0.14
hap	*Calciosolenia murrayi*	0.27	0.01	0.12	0.61	0.20	0.29	0.73	0.23
din	*Oxytoxum laticeps*	0.60	0.06	0.06	0.24	0.05	0.61	0.55	0.27
din	*Spiniferites pachydermus*	0.42	0.01	0.16	0.83	0.10	0.47	0.39	0.06
bac	*Chaetoceros debilis*	0.24	0.07	0.10	0.08	0.04	0.39	0.85	0.67
bac	*Licmophora lyngbyei*	0.54	0.09	0.26	0.07	0.07	0.43	0.07	0.91
din	*Phalacroma rotundatum*	0.42	0.31	0.20	0.48	0.42	0.10	0.01	0.48
din	*Prorocentrum gracile*	0.42	0.00	0.07	0.23	0.01	0.57	0.78	0.36
bac	*Pseudo-nitzschia pseudodelicatissima*	0.61	0.04	0.34	0.65	0.14	0.38	0.04	0.24
din	*Ornithocercus splendidus*	0.57	0.00	0.35	0.44	0.05	0.38	0.45	0.16
bac	*Bacteriastrum mediterraneum*	0.69	0.01	0.53	0.08	0.06	0.11	0.01	0.90
din	*Tripos ranipes*	0.13	0.01	0.20	0.42	0.13	0.38	0.90	0.21
din	*Tripos inflatus*	0.47	0.00	0.01	0.25	0.02	0.63	0.82	0.17
bac	*Thalassiosira leptopus*	0.54	0.43	0.27	0.03	0.02	0.12	0.09	0.88
din	*Bitectatodinium spongium*	0.42	0.00	0.34	0.46	0.02	0.65	0.34	0.13
din	*Preperidinium meunieri*	0.49	0.15	0.06	0.38	0.02	0.37	0.80	0.10
bac	*Lioloma pacificum*	0.15	0.05	0.02	0.29	0.05	0.57	0.89	0.33
din	*Protoperidinium grande*	0.23	0.00	0.01	0.26	0.06	0.65	0.99	0.11
bac	*Chaetoceros bacteriastroides*	0.74	0.00	0.39	0.13	0.08	0.23	0.02	0.70
hap	*Oolithotus fragilis*	0.24	0.04	0.31	0.85	0.26	0.30	0.19	0.12
din	*Ceratocorys armata*	0.48	0.00	0.01	0.34	0.01	0.49	0.66	0.25
bac	*Coscinodiscus asteromphalus*	0.49	0.07	0.28	0.07	0.03	0.54	0.31	0.47
din	*Ceratium gibberum*	0.15	0.07	0.02	0.34	0.23	0.34	0.95	0.12
din	*Pyrocystis fusiformis*	0.08	0.00	0.02	0.20	0.03	0.62	0.99	0.26
hap	*Gephyrocapsa muellerae*	0.15	0.12	0.06	0.69	0.35	0.28	0.35	0.18
bac	*Chaetoceros gracilis*	0.52	0.04	0.20	0.03	0.11	0.32	0.11	0.83
bac	*Actinocyclus curvatulus*	0.19	0.08	0.06	0.05	0.04	0.42	0.74	0.58
din	*Cochlodinium pupa*	0.23	0.08	0.28	0.15	0.39	0.50	0.17	0.33
bac	*Lithodesmium undulatum*	0.34	0.02	0.21	0.08	0.05	0.33	0.62	0.49
din	*Gymnodinium simplex*	0.43	0.03	0.31	0.61	0.28	0.10	0.01	0.33
hap	*Gephyrocapsa caribbeanica*	0.23	0.18	0.36	0.47	0.35	0.04	0.03	0.42
din	*Ceratium longirostrum*	0.29	0.03	0.08	0.56	0.09	0.18	0.76	0.09
hap	*Michaelsarsia adriatica*	0.32	0.01	0.01	0.47	0.07	0.38	0.76	0.05
din	*Protoperidinium stellatum*	0.19	0.01	0.26	0.39	0.22	0.28	0.46	0.22
bac	*Nitzschia pacifica*	0.16	0.06	0.01	0.33	0.07	0.46	0.87	0.04
din	*Protoperidinium leonis*	0.38	0.00	0.11	0.28	0.05	0.46	0.59	0.15
din	*Tripos hexacanthus*	0.04	0.05	0.11	0.30	0.47	0.20	0.84	0.00
din	*Dinophysis caudata*	0.18	0.00	0.02	0.31	0.05	0.46	0.92	0.07
bac	*Asterionellopsis glacialis*	0.22	0.01	0.16	0.01	0.03	0.45	0.52	0.61
hap	*Calciosolenia brasiliensis*	0.07	0.00	0.02	0.41	0.10	0.43	0.91	0.05
din	*Phalacroma doryphorum*	0.62	0.00	0.04	0.54	0.02	0.39	0.22	0.15
din	*Torodinium robustum*	0.43	0.13	0.19	0.11	0.19	0.40	0.09	0.44
bac	*Thalassiosira subtilis*	0.23	0.02	0.10	0.13	0.03	0.26	0.86	0.33
din	*Tripos paradoxides*	0.32	0.00	0.03	0.10	0.02	0.59	0.82	0.08
din	*Mesoporos perforatus*	0.30	0.16	0.14	0.16	0.31	0.22	0.30	0.35
din	*Tripos limulus*	0.32	0.00	0.08	0.13	0.05	0.42	0.74	0.20
bac	*Actinocyclus octonarius*	0.60	0.00	0.23	0.33	0.02	0.44	0.04	0.26
chl	*Ulva fasciata*	0.46	0.01	0.15	0.31	0.19	0.37	0.16	0.25
bac	*Rhizosolenia hyalina*	0.19	0.02	0.04	0.05	0.02	0.44	0.83	0.30
din	*Protoperidinium steinii*	0.26	0.04	0.01	0.21	0.05	0.40	0.76	0.15
din	*Ornithocercus steinii*	0.33	0.02	0.04	0.27	0.02	0.37	0.59	0.24
din	*Ceratium pavillardii*	0.53	0.01	0.50	0.39	0.17	0.06	0.00	0.21

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
din	*Schuettiella mitra*	0.18	0.00	0.01	0.07	0.01	0.52	0.86	0.20
bac	*Nitzschia vitrea*	0.67	0.00	0.24	0.15	0.04	0.13	0.01	0.61
bac	*Coscinodiscus gigas*	0.56	0.01	0.22	0.26	0.03	0.42	0.16	0.18
bac	*Chaetoceros convolutus*	0.09	0.29	0.09	0.03	0.05	0.10	0.58	0.59
dic	*Dictyocha fibula*	0.21	0.34	0.13	0.07	0.05	0.12	0.19	0.73
din	*Gymnodinium wulffii*	0.08	0.43	0.29	0.07	0.36	0.10	0.07	0.41
din	*Diplopsalis lenticula*	0.38	0.04	0.21	0.67	0.10	0.11	0.14	0.15
bac	*Thalassionema bacillare*	0.11	0.02	0.04	0.25	0.06	0.28	0.83	0.19
din	*Phalacroma rapa*	0.29	0.01	0.01	0.11	0.02	0.48	0.73	0.12
bac	*Chaetoceros castracanei*	0.48	0.09	0.21	0.01	0.03	0.10	0.00	0.84
bac	*Lioloma delicatulum*	0.05	0.01	0.01	0.23	0.03	0.43	0.90	0.07
bac	*Coscinodiscus oculus-iridis*	0.67	0.10	0.05	0.16	0.00	0.46	0.03	0.25
din	*Protoperidinium depressum*	0.06	0.10	0.02	0.23	0.03	0.36	0.89	0.05
bac	*Entomoneis paludosa*	0.04	0.31	0.22	0.05	0.32	0.20	0.25	0.30
din	*Protoperidinium crassipes*	0.15	0.01	0.02	0.11	0.02	0.31	0.78	0.32
din	*Protoperidinium mite*	0.56	0.02	0.22	0.19	0.09	0.22	0.05	0.34
din	*Tripos buceros*	0.20	0.01	0.03	0.34	0.06	0.11	0.82	0.11
din	*Diplopsalopsis bomba*	0.27	0.00	0.04	0.02	0.01	0.44	0.58	0.27
bac	*Thalassiosira gravida*	0.10	0.14	0.09	0.03	0.04	0.09	0.47	0.64
din	*Protoperidinium longipes*	0.11	0.00	0.00	0.08	0.01	0.42	0.87	0.09
bac	*Rhabdonema arcuatum*	0.20	0.24	0.07	0.03	0.02	0.19	0.11	0.72
din	*Protoperidinium diabolus*	0.11	0.01	0.05	0.41	0.14	0.10	0.73	0.02
bac	*Melosira borreri*	0.33	0.01	0.28	0.06	0.08	0.09	0.02	0.71
bac	*Thalassiosira antarctica*	0.48	0.29	0.24	0.12	0.03	0.14	0.00	0.26
bac	*Chaetoceros seiracanthus*	0.33	0.01	0.40	0.03	0.13	0.07	0.01	0.58
din	*Gymnodinium uberrimum*	0.29	0.31	0.07	0.62	0.06	0.04	0.09	0.06
din	*Protoperidinium bipes*	0.33	0.59	0.03	0.01	0.02	0.17	0.11	0.28
din	*Protoperidinium granii*	0.32	0.02	0.02	0.20	0.00	0.20	0.54	0.23
bac	*Chaetoceros pseudocrinitus*	0.58	0.01	0.32	0.12	0.04	0.11	0.02	0.31
bac	*Asteromphalus robustus*	0.27	0.11	0.23	0.06	0.05	0.06	0.01	0.72
bac	*Pleurosigma diversestriatum*	0.37	0.00	0.23	0.53	0.01	0.21	0.10	0.04
bac	*Chaetoceros pacificus*	0.53	0.00	0.33	0.12	0.02	0.04	0.00	0.44
cya	*Trichodesmium erythraeum*	0.45	0.00	0.08	0.05	0.00	0.29	0.01	0.59
din	*Tripos lunula*	0.03	0.00	0.01	0.17	0.03	0.25	0.91	0.06
bac	*Ethmodiscus gazellae*	0.06	0.01	0.02	0.24	0.03	0.25	0.74	0.11
bac	*Chaetoceros socialis*	0.18	0.02	0.14	0.09	0.02	0.13	0.52	0.35
din	*Protoperidinium curtipes*	0.10	0.02	0.03	0.20	0.04	0.19	0.60	0.27
din	*Heterodinium blackmanii*	0.08	0.00	0.02	0.05	0.02	0.24	0.77	0.28
hap	*Rhabdosphaera xiphos*	0.32	0.00	0.02	0.58	0.12	0.21	0.05	0.15
din	*Gymnodinium gracile*	0.04	0.03	0.00	0.13	0.04	0.28	0.83	0.08
bac	*Plagiotropis lepidoptera*	0.66	0.00	0.09	0.11	0.01	0.17	0.02	0.36
cya	*Trichodesmium thiebautii*	0.37	0.00	0.07	0.10	0.00	0.15	0.05	0.64
bac	*Chaetoceros concavicornis*	0.08	0.18	0.07	0.02	0.03	0.07	0.36	0.57
din	*Protoperidinium latispinum*	0.44	0.00	0.06	0.06	0.01	0.41	0.11	0.29
bac	*Stephanopyxis nipponica*	0.18	0.29	0.04	0.05	0.04	0.21	0.06	0.47
bac	*Chaetoceros seriacanthus*	0.34	0.01	0.17	0.00	0.04	0.12	0.00	0.56
bac	*Proboscia inermis*	0.00	0.48	0.00	0.02	0.08	0.08	0.53	0.04
bac	*Grammatophora oceanica*	0.11	0.01	0.01	0.06	0.04	0.23	0.51	0.26
bac	*Fragilariopsis oceanica*	0.14	0.24	0.04	0.04	0.02	0.17	0.03	0.53
bac	*Thalassiosira hyalina*	0.08	0.23	0.11	0.04	0.07	0.05	0.02	0.63
din	*Archaeperidinium minutum*	0.35	0.01	0.04	0.22	0.00	0.06	0.41	0.13
din	*Tripos azoricus*	0.01	0.05	0.04	0.06	0.25	0.05	0.56	0.17
din	*Alexandrium monilatum*	0.34	0.00	0.01	0.02	0.00	0.34	0.39	0.06
hap	*Coronosphaera mediterranea*	0.14	0.04	0.05	0.56	0.21	0.13	0.01	0.02
din	*Neoceratium breve*	0.35	0.00	0.06	0.41	0.01	0.14	0.15	0.05
din	*Dinophysis tripos*	0.04	0.09	0.12	0.18	0.48	0.00	0.00	0.24
din	*Tripos gravidus*	0.03	0.00	0.00	0.03	0.01	0.17	0.84	0.06
bac	*Chaetoceros furcellatus*	0.06	0.24	0.13	0.00	0.08	0.03	0.01	0.59
din	*Tripos incisus*	0.02	0.00	0.00	0.09	0.01	0.17	0.79	0.05
din	*Protoperidinium oceanicum*	0.04	0.00	0.00	0.07	0.02	0.18	0.80	0.00
din	*Tripos balechii*	0.51	0.00	0.01	0.13	0.00	0.26	0.02	0.14
din	*Tripos strictus*	0.01	0.00	0.00	0.03	0.00	0.15	0.89	0.00
bac	*Fragilariopsis doliolus*	0.07	0.01	0.05	0.01	0.01	0.05	0.52	0.36
dic	*Octactis speculum*	0.08	0.26	0.03	0.01	0.02	0.07	0.17	0.40
din	*Protoperidinium pellucidum*	0.03	0.03	0.01	0.01	0.01	0.13	0.53	0.28
din	*Protoperidinium claudicans*	0.08	0.01	0.00	0.04	0.02	0.18	0.53	0.15
din	*Phalacroma oxytoxoides*	0.09	0.01	0.01	0.03	0.01	0.11	0.51	0.25
din	*Kofoidinium velleloides*	0.01	0.00	0.02	0.09	0.05	0.16	0.65	0.01
din	*Protoperidinium brevipes*	0.22	0.41	0.02	0.05	0.09	0.02	0.08	0.09
din	*Peridinium breve*	0.11	0.11	0.08	0.04	0.02	0.09	0.02	0.51
din	*Ceratocorys reticulata*	0.02	0.00	0.00	0.06	0.00	0.18	0.70	0.00
din	*Protoperidinium pentagonum*	0.02	0.01	0.00	0.01	0.01	0.18	0.72	0.01

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
bac	*Entomoneis alata*	0.13	0.01	0.01	0.01	0.01	0.16	0.44	0.17
din	*Heterocapsa niei*	0.20	0.02	0.09	0.25	0.13	0.11	0.01	0.13
din	*Heterocapsa rotundata*	0.09	0.04	0.19	0.11	0.20	0.22	0.01	0.07
bac	*Coscinodiscus centralis*	0.01	0.01	0.02	0.01	0.01	0.17	0.65	0.05
hap	*Acanthoica acanthifera*	0.18	0.02	0.04	0.16	0.07	0.16	0.19	0.09
bac	*Chaetoceros teres*	0.13	0.02	0.31	0.06	0.10	0.04	0.01	0.24
bac	*Proboscia indica*	0.05	0.12	0.17	0.06	0.43	0.00	0.00	0.06
din	*Diplopelta asymmetrica*	0.01	0.00	0.00	0.04	0.01	0.12	0.71	0.00
bac	*Chaetoceros indicus*	0.40	0.00	0.11	0.02	0.01	0.01	0.01	0.33
bac	*Actinoptychus splendens*	0.02	0.00	0.00	0.03	0.01	0.18	0.63	0.01
bac	*Dactyliosolen antarcticus*	0.08	0.21	0.09	0.01	0.09	0.06	0.11	0.24
bac	*Hemidiscus cuneiformis*	0.07	0.00	0.00	0.22	0.01	0.07	0.37	0.11
bac	*Rhizosolenia temperei*	0.01	0.00	0.00	0.03	0.01	0.14	0.62	0.04
bac	*Stellarima stellaris*	0.06	0.20	0.07	0.02	0.01	0.02	0.00	0.46
din	*Tripos dens*	0.00	0.00	0.00	0.01	0.01	0.07	0.74	0.00
din	*Protoperidinium venustum*	0.10	0.01	0.01	0.03	0.01	0.10	0.17	0.40
din	*Amphidinium carterae*	0.08	0.08	0.11	0.20	0.13	0.12	0.08	0.02
bac	*Trieres regia*	0.02	0.22	0.26	0.04	0.25	0.02	0.02	0.01
din	*Scrippsiella acuminata*	0.02	0.03	0.00	0.05	0.01	0.13	0.52	0.07
bac	*Licmophora abbreviata*	0.02	0.02	0.00	0.01	0.01	0.09	0.66	0.02
din	*Protoperidinium tenuissimum*	0.00	0.00	0.00	0.01	0.01	0.10	0.68	0.00
bac	*Thalassiosira partheneia*	0.04	0.00	0.00	0.06	0.01	0.08	0.57	0.04
din	*Tripos eugrammus*	0.01	0.00	0.00	0.07	0.02	0.08	0.63	0.00
bac	*Coscinodiscus antarcticus*	0.28	0.23	0.11	0.05	0.03	0.04	0.00	0.06
din	*Protoperidinium subinerme*	0.02	0.00	0.01	0.03	0.02	0.05	0.52	0.13
din	*Protoperidinium tristylum*	0.00	0.00	0.00	0.01	0.00	0.07	0.68	0.00
din	*Ptychodiscus noctiluca*	0.02	0.05	0.00	0.01	0.15	0.20	0.29	0.03
chl	*Halosphaera viridis*	0.01	0.38	0.05	0.01	0.02	0.01	0.01	0.25
bac	*Pseudo-nitzschia lineola*	0.00	0.12	0.01	0.03	0.01	0.01	0.44	0.11
bac	*Rhizosolenia chunii*	0.00	0.30	0.00	0.00	0.01	0.01	0.41	0.00
bac	*Bellerochea malleus*	0.11	0.07	0.03	0.04	0.13	0.14	0.08	0.14
cry	*Hillea fusiformis*	0.14	0.05	0.02	0.07	0.02	0.17	0.10	0.16
din	*Protoperidinium pyrum*	0.02	0.00	0.00	0.02	0.01	0.09	0.54	0.03
din	*Dinophysis acuminata*	0.02	0.12	0.00	0.01	0.01	0.11	0.44	0.00
bac	*Coscinodiscus marginatus*	0.01	0.22	0.02	0.02	0.02	0.02	0.09	0.29
bac	*Roperia tesselata*	0.01	0.00	0.00	0.02	0.02	0.12	0.52	0.01
bac	*Nitzschia bilobata*	0.02	0.00	0.00	0.01	0.00	0.09	0.51	0.06
bac	*Grammatophora marina*	0.02	0.08	0.00	0.01	0.01	0.10	0.33	0.16
din	*Protoperidinium obtusum*	0.03	0.00	0.00	0.01	0.00	0.09	0.40	0.15
hap	*Ophiaster hydroideus*	0.00	0.06	0.00	0.01	0.03	0.05	0.53	0.00
din	*Gyrodinium pingue*	0.11	0.06	0.03	0.01	0.01	0.16	0.13	0.17
din	*Protoperidinium fatulipes*	0.00	0.00	0.00	0.02	0.00	0.05	0.58	0.00
hap	*Rhabdosphaera hispida*	0.18	0.00	0.13	0.17	0.05	0.11	0.00	0.01
din	*Gymnodinium agiliforme*	0.13	0.05	0.03	0.04	0.09	0.15	0.12	0.02
bac	*Helicotheca tamesis*	0.00	0.03	0.00	0.01	0.01	0.02	0.53	0.01
din	*Pyrocystis elegans*	0.01	0.00	0.00	0.03	0.00	0.02	0.51	0.02
bac	*Attheya septentrionalis*	0.02	0.18	0.05	0.00	0.03	0.05	0.08	0.19
bac	*Chaetoceros neglectus*	0.10	0.34	0.08	0.01	0.02	0.01	0.00	0.03
din	*Tripos geniculatus*	0.02	0.00	0.00	0.02	0.01	0.11	0.41	0.00
bac	*Coscinodiscus curvatulus*	0.00	0.31	0.00	0.00	0.02	0.01	0.22	0.00
bac	*Actinoptychus octonarius*	0.00	0.27	0.00	0.00	0.01	0.01	0.28	0.00
din	*Protoperidinium conicoides*	0.01	0.06	0.02	0.02	0.01	0.02	0.20	0.23
din	*Gonyaulax elongata*	0.01	0.09	0.02	0.01	0.01	0.01	0.01	0.40
bac	*Odontella longicruris*	0.01	0.03	0.02	0.02	0.01	0.02	0.24	0.20
din	*Ceratocorys bipes*	0.00	0.01	0.00	0.01	0.01	0.04	0.30	0.18
bac	*Chaetoceros constrictus*	0.03	0.01	0.01	0.02	0.01	0.12	0.33	0.01
din	*Prorocentrum arcuatum*	0.12	0.01	0.00	0.01	0.00	0.21	0.18	0.00
din	*Protoperidinium oblongum*	0.01	0.00	0.00	0.00	0.00	0.04	0.44	0.00
bac	*Eucampia antarctica*	0.08	0.33	0.02	0.00	0.00	0.02	0.00	0.05
din	*Protoperidinium pallidum*	0.01	0.03	0.00	0.01	0.01	0.05	0.39	0.01
bac	*Chaetoceros similis*	0.00	0.25	0.00	0.02	0.01	0.01	0.19	0.00
hap	*Phaeocystis pouchetii*	0.00	0.19	0.02	0.00	0.21	0.00	0.00	0.06
cry	*Leucocryptos marina*	0.00	0.02	0.00	0.01	0.01	0.01	0.42	0.00
din	*Protoperidinium excentricum*	0.00	0.00	0.00	0.01	0.01	0.02	0.42	0.00
bac	*Asteromphalus brookei*	0.00	0.00	0.00	0.01	0.00	0.01	0.44	0.00
bac	*Corethron hystrix*	0.01	0.09	0.01	0.02	0.00	0.02	0.09	0.21
bac	*Bacterosira bathyomphala*	0.02	0.28	0.01	0.00	0.02	0.03	0.00	0.08
din	*Diplopelta steinii*	0.00	0.00	0.00	0.00	0.00	0.04	0.39	0.00
din	*Dinophysis apicata*	0.00	0.06	0.00	0.02	0.02	0.02	0.32	0.00
bac	*Pleurosigma simonsenii*	0.01	0.17	0.01	0.00	0.00	0.07	0.11	0.05
din	*Amphidinium sphenoides*	0.06	0.01	0.04	0.03	0.03	0.14	0.03	0.06
din	*Tripos digitatus*	0.08	0.01	0.00	0.02	0.01	0.11	0.18	0.01
bac	*Fragilariopsis kerguelensis*	0.01	0.37	0.00	0.00	0.00	0.00	0.03	0.00

(*continued on next page*)

Table A.18 (*continued*)

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
din	*Amphisolenia bispinosa*	0.00	0.00	0.00	0.00	0.00	0.02	0.38	0.00
hap	*Calciosolenia granii*	0.00	0.01	0.00	0.03	0.04	0.03	0.27	0.01
bac	*Eupyxidicula turris*	0.04	0.00	0.01	0.01	0.02	0.12	0.14	0.05
bac	*Chaetoceros criophilus*	0.01	0.21	0.01	0.00	0.01	0.01	0.01	0.11
bac	*Asteromphalus parvulus*	0.00	0.35	0.00	0.00	0.02	0.00	0.00	0.00
din	*Tripos minutus*	0.03	0.08	0.01	0.01	0.15	0.06	0.00	0.03
din	*Gymnodinium aureolum*	0.02	0.03	0.00	0.03	0.08	0.15	0.02	0.03
bac	*Cylindrotheca fusiformis*	0.01	0.14	0.01	0.00	0.00	0.01	0.02	0.16
bac	*Porosira glacialis*	0.01	0.12	0.03	0.00	0.04	0.01	0.01	0.13
bac	*Fragilariopsis cylindrus*	0.00	0.34	0.00	0.00	0.00	0.00	0.00	0.00
din	*Kapelodinium vestifici*	0.02	0.18	0.00	0.00	0.00	0.07	0.06	0.00
din	*Tripos bucephalus*	0.00	0.08	0.00	0.02	0.23	0.00	0.00	0.00
din	*Lebessphaera urania*	0.00	0.02	0.12	0.00	0.20	0.00	0.00	0.00
bac	*Achnanthes longipes*	0.00	0.00	0.00	0.00	0.01	0.01	0.30	0.00
eug	*Eutreptiella gymnastica*	0.00	0.00	0.00	0.00	0.01	0.01	0.31	0.00
bac	*Proboscia truncata*	0.00	0.10	0.00	0.00	0.21	0.00	0.00	0.00
din	*Akashiwo sanguinea*	0.01	0.00	0.00	0.01	0.01	0.03	0.26	0.00
bac	*Asteroplanus karianus*	0.01	0.09	0.02	0.00	0.02	0.00	0.00	0.15
din	*Karlodinium veneficum*	0.03	0.01	0.01	0.01	0.02	0.07	0.01	0.13
bac	*Thalassiosira gracilis*	0.01	0.25	0.02	0.00	0.00	0.01	0.00	0.00
din	*Tripos platycornis*	0.01	0.07	0.01	0.01	0.13	0.03	0.03	0.01
bac	*Shionodiscus gracilis*	0.00	0.28	0.00	0.00	0.00	0.00	0.01	0.00
bac	*Thalassiosira mendiolana*	0.00	0.00	0.00	0.00	0.01	0.04	0.22	0.00
din	*Tripos lamellicornis*	0.01	0.07	0.01	0.00	0.15	0.02	0.00	0.02
din	*Protoperidinium grahamii*	0.00	0.01	0.00	0.00	0.01	0.02	0.24	0.00
din	*Tripos carnegiei*	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.00
bac	*Diatoma rhombica*	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00
din	*Dinophysis norvegica*	0.01	0.11	0.00	0.01	0.02	0.00	0.00	0.12
hap	*Phaeocystis antarctica*	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00
din	*Gyrodinium wulffii*	0.01	0.01	0.00	0.01	0.03	0.11	0.03	0.04
bac	*Fragilariopsis obliquecostata*	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00
din	*Polykrikos schwartzii*	0.01	0.07	0.00	0.01	0.13	0.02	0.00	0.01
bac	*Asteromphalus hyalinus*	0.00	0.22	0.00	0.00	0.00	0.00	0.02	0.00
bac	*Thalassiosira nordenskioeldii*	0.00	0.20	0.00	0.00	0.01	0.01	0.01	0.01
bac	*Navicula pelagica*	0.00	0.14	0.02	0.00	0.03	0.00	0.00	0.05
din	*Protoperidinium longispinum*	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.00
din	*Protoperidinium defectum*	0.00	0.23	0.00	0.00	0.01	0.00	0.00	0.00
bac	*Chaetoceros bulbosus*	0.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00
din	*Gymnodinium arcticum*	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.00
din	*Tripos bigelowii*	0.00	0.00	0.00	0.02	0.00	0.01	0.18	0.00
din	*Heterocapsa triquetra*	0.01	0.12	0.00	0.01	0.02	0.02	0.00	0.03
bac	*Fragilariopsis curta*	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00
bac	*Coscinodiscus concinnus*	0.00	0.12	0.00	0.01	0.02	0.03	0.02	0.00
din	*Ceratium arcticum*	0.01	0.12	0.00	0.01	0.01	0.01	0.00	0.02
chl	*Pterosperma vanhoeffenii*	0.01	0.04	0.01	0.00	0.01	0.01	0.00	0.09
bac	*Detonula confervacea*	0.00	0.06	0.00	0.00	0.01	0.02	0.08	0.00
din	*Gyrodinium prunus*	0.01	0.02	0.00	0.00	0.00	0.02	0.00	0.12
bac	*Chaetoceros borealis*	0.02	0.06	0.01	0.00	0.00	0.04	0.00	0.02
bac	*Melosira moniliformis*	0.00	0.06	0.00	0.01	0.01	0.01	0.05	0.02
din	*Protoperidinium mendiolae*	0.00	0.00	0.00	0.00	0.00	0.01	0.14	0.00
bac	*Pleurosigma nicobaricum*	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00
chl	*Pterosperma moebii*	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
bac	*Grammatophora angulosa*	0.01	0.00	0.00	0.00	0.00	0.04	0.10	0.00
bac	*Thalassiosira angustelineata*	0.00	0.03	0.00	0.01	0.01	0.02	0.09	0.00
din	*Tripos compressus*	0.00	0.03	0.00	0.00	0.07	0.01	0.01	0.00
bac	*Coscinodiscus wailesii*	0.01	0.03	0.00	0.00	0.01	0.02	0.04	0.01
din	*Dinophysis acuta*	0.00	0.10	0.01	0.00	0.01	0.00	0.00	0.00
din	*Azadinium caudatum*	0.00	0.03	0.00	0.01	0.05	0.02	0.00	0.00
bac	*Biddulphia alternans*	0.01	0.02	0.00	0.01	0.01	0.03	0.03	0.00
din	*Gyrodinium flagellare*	0.01	0.01	0.00	0.00	0.02	0.03	0.03	0.01
bac	*Coscinodiscus subbulliens*	0.00	0.05	0.00	0.01	0.01	0.01	0.02	0.00
din	*Protoperidinium peruvianum*	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.01
bac	*Ephemera planamembranacea*	0.00	0.06	0.00	0.00	0.01	0.00	0.00	0.00
din	*Protoperidinium ovatum*	0.01	0.03	0.00	0.00	0.00	0.02	0.00	0.02
chr	*Dinobryon balticum*	0.00	0.04	0.00	0.00	0.01	0.01	0.00	0.01
bac	*Thalassiosira minima*	0.01	0.00	0.00	0.01	0.01	0.02	0.03	0.00
din	*Karenia mikimotoi*	0.00	0.04	0.00	0.00	0.02	0.00	0.01	0.00
din	*Cochlodinium vinctum*	0.00	0.01	0.01	0.00	0.01	0.02	0.00	0.01
bac	*Chaetoceros cinctus*	0.00	0.02	0.00	0.00	0.01	0.02	0.00	0.01
bac	*Nitzschia frigida*	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
bac	*Fragilaria striatula*	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00

Table A.19
List of the top 100 most wide-spread species across our biomes. The number represent the relative area covered by a species on average across all months. The missing values indicate that the species is not among the top 100 for that respective biome. The average area coverage of all species is found in Table A.18. Marked in blue are the characteristic species mentioned in the text for each biome. The following abbreviations were used: Phy for phylum; cl for class; din for *Dinoflagellata*; bac for *Bacillariophyceae*; hap for *Haptophyta*; chl for *Chlorophyta*; cya for *Cyanobacteria*; dic for *Dictyochophyceae*; TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Phy/cl	Species	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
bac	*Actinoptychus octonarius*	NaN	27.26%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Actinoptychus senarius*	NaN	NaN	NaN	NaN	NaN	90.30%	97.64%	NaN
bac	*Asterolampra marylandica*	94.24%	NaN	94.12%	NaN	NaN	NaN	NaN	92.33%
bac	*Asteromphalus cleveanus*	95.36%	NaN	88.38%	NaN	46.12%	NaN	NaN	96.43%
bac	*Asteromphalus flabellatus*	NaN	NaN	NaN	NaN	NaN	85.48%	96.24%	NaN
bac	*Asteromphalus heptactis*	NaN	13.72%	NaN	NaN	NaN	81.89%	96.32%	NaN
bac	*Asteromphalus hyalinus*	NaN	22.21%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Asteromphalus parvulus*	NaN	35.42%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Attheya septentrionalis*	NaN	17.72%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Aulacoseira granulata*	NaN	NaN	71.91%	NaN	46.85%	NaN	NaN	NaN
bac	*Azpeitia nodulifera*	92.31%	NaN	77.37%	NaN	NaN	NaN	NaN	NaN
bac	*Bacteriastrum comosum*	93.90	NaN	86.99%	NaN	NaN	NaN	NaN	90.04%
bac	*Bacteriastrum delicatulum*	NaN	NaN	80.94%	NaN	NaN	NaN	NaN	93.79%
bac	*Bacteriastrum elegans*	NaN	NaN	80.80%	NaN	NaN	NaN	NaN	NaN
bac	*Bacteriastrum elongatum*	93.35%	NaN	79.49%	NaN	NaN	NaN	NaN	92.89%
bac	*Bacteriastrum furcatum*	97.07%	NaN	79.24%	NaN	NaN	NaN	NaN	97.69%
bac	*Bacteriastrum hyalinum*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	91.33%
bac	*Bacteriastrum mediterraneum*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	89.63%
bac	*Bacterosira bathyomphala*	NaN	27.71%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Cerataulina pelagica*	99.05%	NaN	94.91%	NaN	NaN	98.81%	97.68%	97.96%
bac	*Chaetoceros aequatorialis*	NaN	NaN	76.17%	NaN	NaN	NaN	NaN	99.49%
bac	*Chaetoceros affinis*	97.18%	NaN	74.36%	NaN	NaN	88.01%	93.00%	99.33%
bac	*Chaetoceros anastomosans*	97.28%	NaN	92.94%	NaN	42.86%	85.57%	NaN	99.97%
bac	*Chaetoceros atlanticus*	NaN	45.08%	NaN	NaN	NaN	NaN	95.65%	NaN
bac	*Chaetoceros brevis*	94.39%	NaN	NaN	NaN	NaN	88.76%	NaN	99.01%
bac	*Chaetoceros bulbosus*	NaN	23.75%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros coarctatus*	98.32%	NaN	86.47%	84.70%	NaN	98.83%	99.79%	93.32%
bac	*Chaetoceros compressus*	97.17%	NaN	94.20%	85.57%	NaN	84.50%	95.25%	92.11%
bac	*Chaetoceros concavicornis*	NaN	17.82%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros convolutus*	NaN	29.43%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros costatus*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	95.03%

(*continued on next page*)

Table A.19 (*continued*)

bac	*Chaetoceros criophilus*	NaN	21.36%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros curvisetus*	98.00%	NaN	98.05%	85.08%	34.54%	87.98%	95.60%	98.53%
bac	*Chaetoceros dadayi*	97.21%	NaN	82.68%	NaN	NaN	NaN	NaN	96.19%
bac	*Chaetoceros danicus*	NaN	NaN	82.81%	NaN	NaN	NaN	94.35%	90.16%
bac	*Chaetoceros decipiens*	94.72%	NaN	75.63%	NaN	NaN	NaN	NaN	92.03%
bac	*Chaetoceros dichaeta*	NaN	29.45%	NaN	NaN	NaN	NaN	NaN	99.53%
bac	*Chaetoceros didymus*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	97.25%
bac	*Chaetoceros diversus*	97.71%	NaN	93.07%	NaN	NaN	NaN	NaN	92.33%
bac	*Chaetoceros eibenii*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	93.82%
bac	*Chaetoceros furcellatus*	NaN	24.17%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros hyalochaetae*	96.74%	NaN	94.26%	96.69%	79.18%	91.93%	NaN	NaN
bac	*Chaetoceros laciniosus*	98.35%	21.93%	96.55%	91.79%	33.55%	NaN	NaN	99.49%
bac	*Chaetoceros lorenzianus*	95.64%	NaN	NaN	NaN	NaN	93.49%	96.82%	96.74%
bac	*Chaetoceros messanensis*	95.70%	NaN	85.71%	NaN	33.22%	NaN	NaN	97.91%
bac	*Chaetoceros neglectus*	NaN	34.47%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros paradoxus*	NaN	NaN	NaN	NaN	30.83%	NaN	NaN	98.65%
bac	*Chaetoceros pelagicus*	NaN	21.55%	NaN	NaN	NaN	NaN	85.61%	NaN
bac	*Chaetoceros pendulus*	95.68%	NaN	89.41%	NaN	37.06%	NaN	NaN	99.46%
bac	*Chaetoceros peruvianus*	99.44%	NaN	95.99%	88.73%	NaN	99.25%	99.46%	97.97%
bac	*Chaetoceros phaeoceros*	NaN	NaN	86.25%	NaN	37.15%	NaN	NaN	NaN
bac	*Chaetoceros pseudocurvisetus*	93.87%	NaN	75.14%	NaN	NaN	NaN	NaN	94.68%
bac	*Chaetoceros radicans*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	97.29%
bac	*Chaetoceros rostratus*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	90.85%
bac	*Chaetoceros saltans*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	94.21%
bac	*Chaetoceros seychellarum*	NaN	NaN	80.16%	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros similis*	NaN	25.03%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Chaetoceros simplex*	NaN	12.06%	NaN	NaN	36.89%	NaN	NaN	NaN
bac	*Chaetoceros subsecundus*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	97.13%
bac	*Chaetoceros tetrastichon*	94.85%	NaN	76.36%	NaN	29.23%	83.83%	NaN	97.90%
bac	*Chaetoceros vanheurckii*	NaN	NaN	73.86%	NaN	NaN	NaN	NaN	99.01%
bac	*Chaetoceros wighamii*	93.89%	NaN	89.68%	NaN	NaN	NaN	NaN	91.55%
bac	*Climacodium biconcavum*	NaN	NaN	78.12%	NaN	NaN	NaN	NaN	98.63%
bac	*Climacodium frauenfeldianum*	99.32%	NaN	96.97%	85.91%	38.87%	90.44%	NaN	99.11%
bac	*Climacosphenia moniligera*	94.87%	NaN	NaN	NaN	NaN	89.53%	89.74%	95.15%
bac	*Corethron pennatum*	NaN	60.73%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Coscinodiscus antarcticus*	NaN	22.70%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Coscinodiscus concinnus*	NaN	11.72%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Coscinodiscus curvatulus*	NaN	31.36%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Coscinodiscus granii*	NaN	NaN	NaN	NaN	NaN	NaN	96.47%	NaN
bac	*Coscinodiscus marginatus*	NaN	22.41%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Coscinodiscus radiatus*	NaN	NaN	NaN	NaN	NaN	88.33%	98.09%	NaN
bac	*Cylindrotheca fusiformis*	NaN	14.11%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Dactyliosolen antarcticus*	NaN	21.35%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Dactyliosolen fragilissimus*	NaN	NaN	78.83%	NaN	NaN	NaN	NaN	NaN
bac	*Detonula pumila*	96.23%	NaN	81.93%	NaN	NaN	88.37%	NaN	87.88%
bac	*Diatoma rhombica*	NaN	26.73%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Ditylum brightwellii*	NaN	NaN	74.24%	NaN	NaN	85.83%	NaN	NaN
bac	*Entomoneis paludosa*	NaN	31.44%	NaN	NaN	32.43%	NaN	NaN	NaN
bac	*Eucampia antarctica*	NaN	32.85%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Eucampia cornuta*	95.62%	18.08%	89.83%	NaN	31.58%	NaN	NaN	98.80%
bac	*Eucampia zodiacus*	NaN	NaN	NaN	NaN	NaN	NaN	87.44%	96.58%
bac	*Eucampia zoodiacus*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	98.68%

(*continued on next page*)

Table A.19 (*continued*)

bac	*Fragilariopsis curta*	NaN	21.94%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Fragilariopsis cylindrus*	NaN	33.59%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Fragilariopsis kerguelensis*	NaN	37.08%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Fragilariopsis obliquecostata*	NaN	25.00%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Fragilariopsis oceanica*	NaN	24.27%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Gossleriella tropica*	NaN	NaN	NaN	NaN	30.30%	NaN	NaN	89.10%
bac	*Guinardia cylindrus*	97.26%	11.96%	95.18%	88.30%	30.81%	NaN	NaN	97.36%
bac	*Guinardia delicatula*	NaN	NaN	NaN	NaN	NaN	88.03%	NaN	NaN
bac	*Guinardia flaccida*	95.56%	NaN	92.54%	85.53%	30.79%	NaN	NaN	91.01%
bac	*Guinardia striata*	95.27%	NaN	84.31%	NaN	37.95%	NaN	NaN	96.22%
bac	*Haslea wawrikae*	94.37%	NaN	84.52%	NaN	NaN	NaN	NaN	97.98%
bac	*Hemiaulus chinensis*	95.40%	NaN	87.29%	NaN	NaN	NaN	NaN	89.82%
bac	*Hemiaulus hauckii*	99.14%	NaN	96.39%	93.82%	NaN	86.64%	NaN	98.50%
bac	*Hemiaulus membranaceus*	95.36%	NaN	NaN	NaN	NaN	NaN	NaN	95.15%
bac	*Lauderia annulata*	NaN	NaN	NaN	NaN	NaN	94.06%	NaN	NaN
bac	*Leptocylindrus danicus*	99.47%	NaN	95.15%	95.61%	39.82%	88.04%	NaN	97.35%
bac	*Leptocylindrus minimus*	93.33%	NaN	NaN	NaN	NaN	NaN	NaN	99.34%
bac	*Licmophora lyngbyei*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	90.57%
bac	*Lioloma delicatulum*	NaN	NaN	NaN	NaN	NaN	NaN	89.66%	NaN
bac	*Lioloma pacificum*	NaN	NaN	NaN	NaN	NaN	NaN	89.48%	NaN
bac	*Mastogloia rostrata*	94.00%	NaN	96.19%	85.31%	NaN	NaN	NaN	92.76%
bac	*Membraneis challengeri*	NaN	41.55%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Meuniera membranacea*	94.68%	NaN	83.12%	NaN	NaN	NaN	NaN	99.52%
bac	*Navicula pelagica*	NaN	13.55%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Neocalyptrella robusta*	NaN	NaN	NaN	NaN	NaN	85.31%	96.29%	NaN
bac	*Nitzschia bicapitata*	98.27%	25.61%	93.68%	96.82%	75.16%	99.54%	98.00%	89.57%
bac	*Nitzschia longissima*	98.06%	25.89%	91.09%	90.83%	42.61%	88.69%	93.13%	99.12%
bac	*Nitzschia pacifica*	NaN	NaN	NaN	NaN	NaN	NaN	87.41%	NaN
bac	*Nitzschia sigma*	96.04%	NaN	81.74%	NaN	NaN	NaN	NaN	88.59%
bac	*Nitzschia tenuirostris*	NaN	NaN	NaN	NaN	NaN	82.46%	NaN	NaN
bac	*Odontella aurita*	NaN	45.67%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Planktoniella sol*	93.23%	NaN	NaN	NaN	NaN	93.22%	99.72%	NaN
bac	*Pleurosigma angulatum*	96.43%	NaN	98.08%	88.84%	64.53%	NaN	NaN	91.98%
bac	*Pleurosigma directum*	NaN	48.67%	NaN	NaN	NaN	87.15%	NaN	NaN
bac	*Pleurosigma elongatum*	NaN	NaN	96.54%	92.67%	70.67%	NaN	NaN	88.59%
bac	*Pleurosigma normanii*	95.39%	NaN	82.08%	87.31%	51.21%	89.95%	NaN	NaN
bac	*Pleurosigma simonsenii*	NaN	16.95%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Podosira stelligera*	NaN	28.12%	87.30%	NaN	78.37%	NaN	NaN	NaN
bac	*Porosira glacialis*	NaN	11.96%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Proboscia indica*	NaN	NaN	NaN	NaN	43.50%	NaN	NaN	NaN
bac	*Proboscia inermis*	NaN	48.20%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Pseudo-nitzschia pungens*	97.21%	NaN	81.30%	NaN	NaN	NaN	NaN	98.70%
bac	*Pseudosolenia calcar-avis*	99.88%	NaN	97.96%	97.86%	55.56%	94.15%	NaN	98.44%
bac	*Rhabdonema arcuatum*	NaN	24.11%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Rhizosolenia acuminata*	NaN	21.41%	NaN	NaN	49.76%	88.91%	96.63%	96.66%
bac	*Rhizosolenia bergonii*	NaN	NaN	NaN	NaN	62.68%	83.74%	98.73%	89.45%
bac	*Rhizosolenia castracanei*	94.53%	NaN	NaN	NaN	NaN	84.17%	95.82%	97.09%
bac	*Rhizosolenia chunii*	NaN	30.06%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Rhizosolenia formosa*	NaN	NaN	NaN	NaN	NaN	87.52%	86.22%	95.43%
bac	*Rhizosolenia setigera*	98.69%	NaN	91.43%	90.53%	53.98%	94.83%	98.72%	97.98%
bac	*Shionodiscus gracilis*	NaN	27.63%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Shionodiscus oestrupii*	96.26%	26.74%	85.02%	95.40%	47.68%	82.60%	NaN	97.66%

(*continued on next page*)

Table A.19 (*continued*)

bac	*Stellarima stellaris*	NaN	19.99%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Stephanopyxis nipponica*	NaN	29.05%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Stephanopyxis palmeriana*	NaN	NaN	NaN	NaN	NaN	NaN	89.44%	96.45%
bac	*Striatella unipunctata*	94.85%	NaN	NaN	90.94%	NaN	NaN	NaN	NaN
bac	*Thalassionema frauenfeldii*	96.18%	NaN	84.32%	NaN	NaN	NaN	NaN	97.36%
bac	*Thalassiosira aestivalis*	NaN	NaN	NaN	NaN	NaN	NaN	88.86%	NaN
bac	*Thalassiosira angulata*	NaN	36.79%	NaN	NaN	NaN	NaN	NaN	97.60%
bac	*Thalassiosira antarctica*	NaN	28.71%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira decipiens*	NaN	29.58%	NaN	NaN	NaN	NaN	NaN	97.71%
bac	*Thalassiosira eccentrica*	NaN	41.59%	NaN	NaN	NaN	96.77%	98.06%	92.77%
bac	*Thalassiosira gracilis*	NaN	24.70%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira gravida*	NaN	14.30%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira hyalina*	NaN	23.41%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira leptopus*	NaN	42.56%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira nordenskioeldii*	NaN	20.03%	NaN	NaN	NaN	NaN	NaN	NaN
bac	*Thalassiosira subtilis*	NaN	NaN	NaN	NaN	NaN	NaN	85.82%	NaN
bac	*Thalassiothrix heteromorpha*	NaN	NaN	NaN	NaN	NaN	NaN	NaN	93.02%
bac	*Trieres chinensis*	NaN	41.44%	NaN	NaN	47.63%	NaN	NaN	NaN
bac	*Trieres regia*	NaN	21.73%	NaN	NaN	NaN	NaN	NaN	NaN
chl	*Halosphaera viridis*	NaN	38.18%	NaN	NaN	NaN	NaN	NaN	NaN
chl	*Pterosperma moebii*	NaN	12.91%	NaN	NaN	NaN	NaN	NaN	NaN
dic	*Dictyocha fibula*	NaN	33.90%	NaN	NaN	NaN	NaN	NaN	NaN
dic	*Octactis speculum*	NaN	25.61%	NaN	NaN	NaN	NaN	NaN	NaN
cya	*Prochlorococcus*	99.21%	NaN	92.88%	98.10%	52.94%	97.14%	NaN	NaN
cya	*Synechococcus*	NaN	40.91%	NaN	NaN	31.12%	97.40%	NaN	96.46%
din	*Amphidinium acutissimum*	NaN	NaN	NaN	NaN	NaN	NaN	96.79%	NaN
din	*Amphisolenia bidentata*	NaN	NaN	NaN	NaN	NaN	93.48%	97.34%	NaN
din	*Blepharocysta splendor-maris*	NaN	NaN	NaN	86.79%	NaN	NaN	NaN	NaN
din	*Ceratium arcticum*	NaN	12.32%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Ceratium breve*	NaN	NaN	NaN	NaN	NaN	NaN	86.58%	NaN
din	*Ceratium contrarium*	99.19%	NaN	89.41%	94.60%	NaN	95.56%	NaN	97.29%
din	*Ceratium falcatiforme*	NaN	NaN	NaN	NaN	NaN	NaN	91.60%	NaN
din	*Ceratium falcatum*	NaN	NaN	NaN	NaN	NaN	83.18%	97.81%	NaN
din	*Ceratium gibberum*	NaN	NaN	NaN	NaN	NaN	NaN	95.28%	NaN
din	*Ceratium massiliense*	NaN	NaN	76.88%	98.85%	60.77%	NaN	NaN	NaN
din	*Ceratium setaceum*	98.35%	NaN	78.97%	98.32%	39.73%	97.55%	99.93%	NaN
din	*Ceratium symmetricum*	NaN	NaN	NaN	NaN	NaN	86.54%	98.59%	NaN
din	*Ceratium trichoceros*	NaN	NaN	80.06%	96.75%	63.35%	NaN	94.78%	NaN
din	*Ceratocorys horrida*	NaN	NaN	NaN	NaN	NaN	89.29%	99.84%	NaN
din	*Ceratoperidinium falcatum*	96.21%	NaN	NaN	NaN	37.26%	90.08%	NaN	NaN
din	*Cladopyxis brachiolata*	95.63%	NaN	77.18%	87.91%	NaN	88.06%	NaN	NaN
din	*Cochlodinium pupa*	NaN	NaN	NaN	NaN	39.19%	NaN	NaN	NaN
din	*Corythodinium tesselatum*	NaN	NaN	NaN	92.92%	NaN	NaN	NaN	NaN
din	*Dinophysis acuminata*	NaN	11.92%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Dinophysis argus*	NaN	NaN	NaN	NaN	NaN	NaN	94.51%	NaN
din	*Dinophysis caudata*	NaN	NaN	NaN	NaN	NaN	NaN	92.27%	NaN
din	*Dinophysis fortii*	93.89%	NaN	NaN	89.77%	NaN	NaN	NaN	97.58%
din	*Dinophysis ovum*	NaN	NaN	NaN	NaN	31.64%	NaN	NaN	NaN
din	*Dinophysis parvula*	96.49%	NaN	84.55%	96.81%	NaN	NaN	NaN	NaN
din	*Dinophysis schroederi*	NaN	NaN	78.06%	92.92%	43.27%	NaN	NaN	NaN
din	*Dinophysis schuettii*	98.74%	NaN	86.63%	99.01%	41.36%	95.70%	95.56%	NaN
din	*Dinophysis tripos*	NaN	NaN	NaN	NaN	48.29%	NaN	NaN	NaN

(*continued on next page*)

Table A.19 (*continued*)

din	*Echinidinium delicatum*	NaN	NaN	NaN	85.53%	NaN	NaN	NaN	NaN
din	*Gonyaulax birostris*	NaN	NaN	NaN	87.91%	NaN	NaN	NaN	NaN
din	*Gonyaulax digitalis*	95.12%	NaN	NaN	87.96%	30.55%	92.01%	88.75%	NaN
din	*Gonyaulax membranacea*	95.17%	NaN	NaN	NaN	39.71%	94.60%	94.95%	NaN
din	*Gonyaulax monacantha*	NaN	NaN	NaN	96.80%	34.68%	NaN	NaN	NaN
din	*Gonyaulax pacifica*	NaN	NaN	NaN	NaN	NaN	NaN	96.25%	NaN
din	*Gonyaulax polygramma*	NaN	NaN	NaN	NaN	NaN	NaN	91.51%	NaN
din	*Gonyaulax scrippsae*	97.32%	NaN	92.58%	97.36%	45.18%	88.10%	NaN	NaN
din	*Gonyaulax spinifera*	99.53%	24.23%	86.68%	99.53%	46.50%	99.77%	99.75%	93.19%
din	*Gymnodinium arcticum*	NaN	22.57%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Gymnodinium marinum*	NaN	NaN	NaN	88.59%	42.03%	NaN	NaN	NaN
din	*Gymnodinium uberrimum*	NaN	31.29%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Gymnodinium wulffii*	NaN	42.79%	NaN	NaN	35.87%	NaN	NaN	NaN
din	*Gyrodinium fusiforme*	NaN	13.89%	NaN	NaN	NaN	88.48%	86.01%	NaN
din	*Gyrodinium spirale*	NaN	NaN	NaN	89.39%	42.65%	NaN	NaN	NaN
din	*Heterocapsa triquetra*	NaN	11.98%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Impagidinium aculeatum*	96.38%	30.75%	84.66%	96.25%	54.88%	92.40%	87.14%	NaN
din	*Impagidinium patulum*	96.84%	29.90%	89.06%	97.73%	62.82%	92.95%	NaN	NaN
din	*Impagidinium sphaericum*	93.76%	41.78%	90.87%	94.60%	55.07%	84.81%	86.08%	88.27%
din	*Kapelodinium vestifici*	NaN	17.74%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Lebouridinium glaucum*	NaN	NaN	NaN	NaN	NaN	85.20%	NaN	NaN
din	*Leonella granifera*	NaN	NaN	NaN	86.10%	NaN	88.62%	NaN	NaN
din	*Lingulodinium polyedra*	99.08%	NaN	79.58%	99.49%	38.16%	99.63%	100.00%	91.27%
din	*Mesoporos perforatus*	NaN	16.17%	NaN	NaN	31.41%	NaN	NaN	NaN
din	*Neoceratium hexacanthum*	NaN	NaN	NaN	NaN	33.02%	NaN	NaN	NaN
din	*Ornithocercus heteroporus*	95.26%	NaN	74.67%	87.33%	NaN	NaN	NaN	NaN
din	*Ornithocercus magnificus*	93.13%	NaN	NaN	87.84%	NaN	94.45%	99.39%	NaN
din	*Ornithocercus quadratus*	NaN	NaN	NaN	NaN	NaN	NaN	97.48%	NaN
din	*Ornithocercus thumii*	NaN	NaN	NaN	NaN	NaN	NaN	91.68%	NaN
din	*Oxytoxum caudatum*	NaN	NaN	NaN	88.76%	NaN	NaN	NaN	NaN
din	*Oxytoxum constrictum*	NaN	NaN	NaN	85.95%	58.77%	NaN	NaN	NaN
din	*Oxytoxum curvatum*	NaN	NaN	NaN	88.67%	48.67%	NaN	NaN	NaN
din	*Oxytoxum elegans*	NaN	NaN	NaN	NaN	29.40%	NaN	NaN	NaN
din	*Oxytoxum gracile*	NaN	NaN	NaN	90.68%	NaN	NaN	NaN	NaN
din	*Oxytoxum longiceps*	95.85%	NaN	NaN	85.25%	NaN	99.27%	97.79%	NaN
din	*Oxytoxum milneri*	97.43%	NaN	72.21%	96.65%	NaN	88.61%	NaN	NaN
din	*Oxytoxum parvum*	98.36%	NaN	NaN	97.68%	38.36%	94.14%	NaN	96.94%
din	*Oxytoxum sceptrum*	NaN	NaN	NaN	87.59%	NaN	NaN	NaN	NaN
din	*Oxytoxum scolopax*	99.44%	NaN	88.67%	99.84%	51.96%	98.79%	94.81%	97.32%
din	*Oxytoxum sphaeroideum*	NaN	NaN	NaN	88.68%	53.91%	89.82%	NaN	NaN
din	*Oxytoxum tesselatum*	NaN	NaN	77.83%	93.65%	49.84%	NaN	NaN	NaN
din	*Oxytoxum turbo*	NaN	NaN	88.34%	95.80%	43.87%	NaN	NaN	NaN
din	*Oxytoxum variabile*	96.59%	NaN	NaN	94.20%	56.66%	96.20%	92.83%	NaN
din	*Pentapharsodinium dalei*	NaN	29.47%	NaN	NaN	47.20%	96.06%	98.56%	NaN
din	*Phalacroma lens*	NaN	NaN	NaN	NaN	NaN	NaN	95.27%	NaN
din	*Phalacroma rotundatum*	NaN	31.11%	NaN	NaN	42.14%	NaN	NaN	NaN
din	*Podolampas bipes*	NaN	NaN	NaN	NaN	NaN	NaN	98.16%	NaN
din	*Podolampas elegans*	97.54%	NaN	87.75%	94.38%	31.42%	89.08%	90.33%	NaN
din	*Podolampas palmipes*	NaN	NaN	75.23%	96.51%	37.25%	85.54%	99.00%	NaN
din	*Podolampas spinifera*	99.07%	NaN	85.95%	98.36%	34.99%	96.98%	97.44%	92.24%
din	*Preperidinium meunieri*	NaN	14.85%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Pronoctiluca pelagica*	NaN	NaN	NaN	93.03%	NaN	88.31%	96.36%	NaN

(*continued on next page*)

Table A.19 (*continued*)

din	*Pronoctiluca spinifera*	NaN	NaN	NaN	85.75%	NaN	NaN	NaN	NaN
din	*Prorocentrum balticum*	95.65%	27.65%	NaN	NaN	NaN	88.82%	NaN	89.64%
din	*Prorocentrum cordatum*	NaN	58.21%	NaN	NaN	34.41%	NaN	NaN	NaN
din	*Prorocentrum dentatum*	99.16%	NaN	85.87%	99.17%	52.02%	88.31%	NaN	99.36%
din	*Prorocentrum lima*	98.70%	19.59%	90.62%	98.74%	44.84%	94.69%	97.93%	97.64%
din	*Prorocentrum mexicanum*	94.45%	NaN	NaN	NaN	NaN	98.95%	NaN	NaN
din	*Prorocentrum micans*	NaN	NaN	NaN	NaN	NaN	94.95%	95.24%	NaN
din	*Prorocentrum rostratum*	99.35%	NaN	NaN	98.87%	38.50%	98.29%	91.32%	97.01%
din	*Prorocentrum scutellum*	95.60%	NaN	95.35%	97.78%	67.06%	81.83%	NaN	97.77%
din	*Protoceratium reticulatum*	95.62%	32.51%	73.97%	92.54%	41.17%	93.20%	NaN	NaN
din	*Protoperidinium abei*	NaN	NaN	NaN	NaN	NaN	83.13%	NaN	NaN
din	*Protoperidinium bipes*	NaN	58.80%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Protoperidinium brevipes*	NaN	40.58%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Protoperidinium brochii*	NaN	NaN	NaN	NaN	NaN	NaN	98.41%	NaN
din	*Protoperidinium cerasus*	NaN	18.28%	NaN	NaN	NaN	NaN	NaN	93.15%
din	*Protoperidinium defectum*	NaN	23.13%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Protoperidinium depressum*	NaN	NaN	NaN	NaN	NaN	NaN	88.80%	NaN
din	*Protoperidinium divergens*	NaN	NaN	NaN	NaN	NaN	NaN	95.60%	NaN
din	*Protoperidinium elegans*	NaN	NaN	NaN	NaN	NaN	NaN	89.99%	NaN
din	*Protoperidinium grande*	NaN	NaN	NaN	NaN	NaN	NaN	98.77%	NaN
din	*Protoperidinium latidorsale*	93.81%	NaN	74.43%	NaN	NaN	98.16%	90.20%	NaN
din	*Protoperidinium longipes*	NaN	NaN	NaN	NaN	NaN	NaN	86.66%	NaN
din	*Protoperidinium mediterraneum*	96.47%	NaN	77.34%	93.41%	NaN	NaN	NaN	NaN
din	*Protoperidinium ovum*	NaN	NaN	NaN	92.63%	NaN	NaN	NaN	NaN
din	*Protoperidinium pedunculatum*	NaN	NaN	NaN	NaN	NaN	NaN	98.15%	NaN
din	*Protoperidinium punctulatum*	98.32%	NaN	NaN	88.64%	NaN	88.71%	NaN	96.34%
din	*Protoperidinium quarnerense*	NaN	NaN	NaN	NaN	NaN	NaN	95.45%	NaN
din	*Protoperidinium tuba*	NaN	NaN	NaN	92.40%	NaN	NaN	NaN	NaN
din	*Pyrocystis fusiformis*	NaN	NaN	NaN	NaN	NaN	NaN	98.82%	NaN
din	*Pyrocystis robusta*	NaN	NaN	NaN	NaN	NaN	87.02%	NaN	NaN
din	*Pyrophacus horologium*	NaN	NaN	NaN	NaN	NaN	NaN	90.56%	NaN
din	*Pyrophacus steinii*	NaN	NaN	NaN	NaN	NaN	NaN	96.08%	NaN
din	*Pyrophacus vancampoae*	NaN	NaN	NaN	93.21%	NaN	NaN	NaN	NaN
din	*Scrippsiella regalis*	NaN	NaN	89.96%	NaN	39.66%	84.55%	NaN	NaN
din	*Spiraulax kofoidii*	NaN	NaN	NaN	NaN	NaN	86.05%	93.47%	NaN
din	*Thoracosphaera heimii*	98.14%	NaN	78.15%	93.83%	NaN	97.90%	NaN	NaN
din	*Torodinium robustum*	NaN	13.15%	NaN	NaN	NaN	NaN	NaN	NaN
din	*Triadinium polyedricum*	NaN	NaN	NaN	NaN	NaN	NaN	95.98%	NaN
din	*Trinovantidinium applanatum*	NaN	NaN	73.19%	88.55%	61.10%	88.74%	NaN	NaN
din	*Tripos belone*	97.51%	NaN	74.74%	96.70%	NaN	97.04%	99.78%	88.61%
din	*Tripos candelabrum*	NaN	NaN	NaN	92.36%	48.53%	NaN	NaN	NaN
din	*Tripos carriensis*	96.73%	NaN	74.55%	97.57%	41.64%	93.45%	90.14%	NaN
din	*Tripos concilians*	NaN	NaN	NaN	NaN	NaN	NaN	86.93%	NaN
din	*Tripos contortus*	NaN	NaN	NaN	NaN	NaN	NaN	92.76%	NaN
din	*Tripos declinatus*	96.90%	NaN	NaN	95.20%	44.98%	NaN	NaN	88.14%
din	*Tripos deflexus*	NaN	NaN	NaN	NaN	NaN	82.27%	89.28%	NaN
din	*Tripos extensus*	100.00%	NaN	99.92%	99.94%	95.66%	100.00%	100.00%	98.16%
din	*Tripos hexacanthus*	NaN	NaN	NaN	NaN	46.60%	NaN	NaN	NaN
din	*Tripos karstenii*	96.62%	NaN	83.94%	93.78%	NaN	NaN	NaN	NaN
din	*Tripos kofoidii*	NaN	NaN	NaN	85.63%	NaN	87.67%	98.58%	NaN
din	*Tripos lunula*	NaN	NaN	NaN	NaN	NaN	NaN	90.63%	NaN
din	*Tripos pentagonus*	96.43%	16.17%	79.58%	94.20%	30.29%	89.86%	97.50%	99.50%

(*continued on next page*)

Table A.19 (*continued*)

din	*Tripos pulchellus*	94.51%	NaN	NaN	NaN	NaN	81.71%	99.28%	NaN
din	*Tripos ranipes*	NaN	NaN	NaN	NaN	NaN	NaN	89.95%	NaN
din	*Tripos strictus*	NaN	NaN	NaN	NaN	NaN	NaN	88.75%	NaN
din	*Tripos teres*	99.10%	NaN	97.51%	98.84%	62.06%	89.88%	NaN	95.34%
din	*Tripos trichoceros*	NaN	NaN	NaN	NaN	32.50%	NaN	NaN	NaN
din	*Tripos vultur*	NaN	NaN	NaN	NaN	NaN	NaN	99.60%	NaN
din	*Tryblionella compressa*	93.94%	15.81%	NaN	NaN	NaN	96.46%	NaN	NaN
hap	*Calciosolenia brasiliensis*	NaN	NaN	NaN	NaN	NaN	NaN	91.46%	NaN
hap	*Discosphaera tubifera*	95.77%	NaN	83.85%	98.39%	37.84%	NaN	NaN	88.27%
hap	*Florisphaera profunda*	NaN	12.29%	NaN	88.29%	NaN	81.75%	NaN	NaN
hap	*Gephyrocapsa caribbeanica*	NaN	18.01%	NaN	NaN	34.89%	NaN	NaN	NaN
hap	*Gephyrocapsa ericsonii*	NaN	16.63%	NaN	84.62%	36.93%	NaN	NaN	NaN
hap	*Gephyrocapsa muellerae*	NaN	NaN	NaN	NaN	35.45%	NaN	NaN	NaN
hap	*Gladiolithus flabellatus*	NaN	NaN	NaN	91.62%	31.77%	NaN	NaN	NaN
hap	*Hayaster perplexus*	NaN	NaN	NaN	91.37%	NaN	91.56%	87.03%	NaN
hap	*Helladosphaera cornifera*	NaN	NaN	NaN	86.12%	31.39%	NaN	NaN	NaN
hap	*Phaeocystis antarctica*	NaN	26.48%	NaN	NaN	NaN	NaN	NaN	NaN
hap	*Phaeocystis pouchetii*	NaN	18.59%	NaN	NaN	NaN	NaN	NaN	NaN
hap	*Reticulofenestra sessilis*	NaN	NaN	NaN	89.19%	NaN	NaN	NaN	NaN
hap	*Rhabdolithes claviger*	NaN	NaN	77.69%	92.18%	62.11%	NaN	NaN	NaN
hap	*Syracosphaera molischii*	NaN	NaN	NaN	NaN	45.87%	NaN	NaN	NaN
hap	*Syracosphaera nodosa*	NaN	NaN	NaN	95.39%	NaN	NaN	NaN	NaN
hap	*Syracosphaera prolongata*	NaN	NaN	NaN	NaN	35.06%	NaN	NaN	NaN
hap	*Umbellosphaera irregularis*	NaN	NaN	81.97%	96.08%	NaN	NaN	NaN	NaN
hap	*Umbellosphaera tenuis*	NaN	NaN	NaN	97.81%	37.99%	NaN	NaN	NaN
hap	*Umbilicosphaera hulburtiana*	NaN	NaN	NaN	98.63%	50.09%	NaN	NaN	NaN
hap	*Umbilicosphaera sibogae*	98.54%	NaN	98.75%	99.60%	62.40%	93.61%	NaN	98.53%

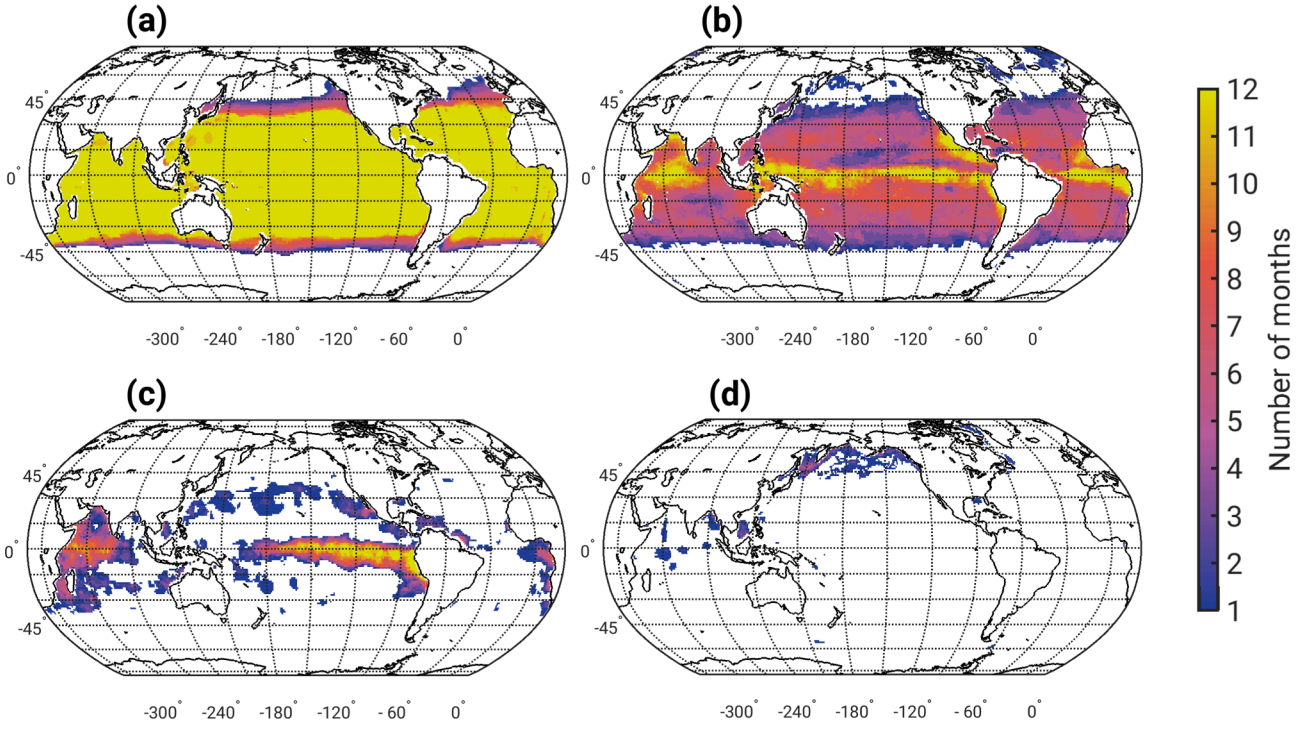


Fig. A.16. Comparison of the global area coverage of wide-spread and rare species. The continuous distribution from wide-spread to local species is shown with four representative species, reflecting the extreme ends of the spectrum. (a) Species *Tripos extensus* is the species with the largest mean area coverage, (b) species *Pyrophacus vancampoae* is the species with the 179th largest mean area coverage, (c) species *Heterodinium blackmanii* is the species with the 358th largest mean area coverage, (d) species *Fragilaria striatula* is the species with the least mean area coverage. The color bar shows the number of months the respective species is found in a $1^\circ \times 1^\circ$ pixel.

A.12. Indicator species

Fig. A.17.

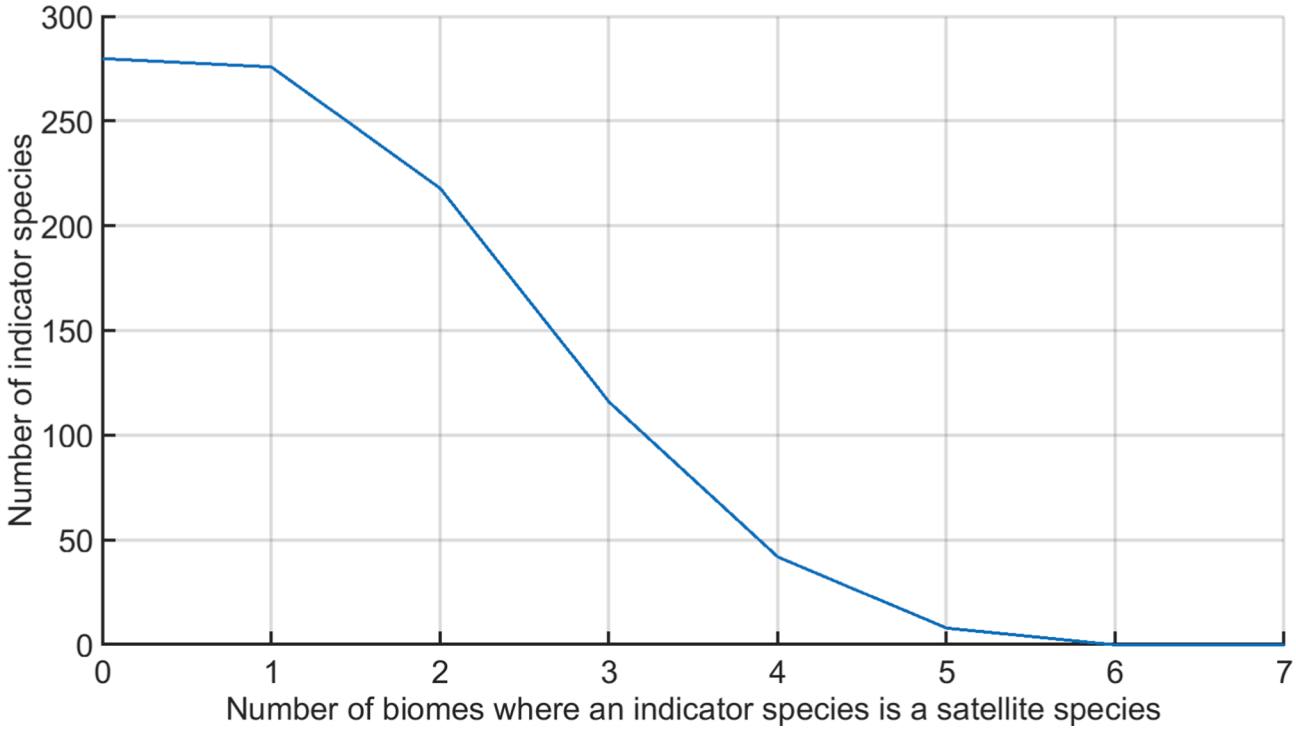


Fig. A.17. Number of times, an indicator species of a specific biome was a satellite species in another biome.

A.13. Species co-occurrences

Table A.20.

Table A.20
Co-occurrence metric for significant phytoplankton species pairs in all biomes. The co-occurrence metric is based on the text analysis algorithm by Dunning (1993). Positive values suggest a pair of species that co-occurs more often than expected from the individual presence frequencies of species across all months. Negative values suggest a pair of species that co-occurs less often than expected from the individual presence frequencies of species across all months. Species pairs are chosen as significant pairs if they have a positive value and the minimum area coverage between the species involved is above the 90*th* percentile in one biome, and the association value is negative or the minimum area coverage between the species involved is below the 10*th* percentile in all other biomes (Section 2.6.3). For each biome we highlighted in blue the respective significant pairs. The following abbreviations were used: Phy for phylum; cl for class; din for *Dinoflagellata*; bac for *Bacillariophyceae*; hap for *Haptophyta*; chl for *Chlorophyta*; dic for *Dictyochophyceae*; TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and TRopical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Phy/cl	(ID) Species 1	(ID) Species 2	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
bac-din	*(125) Eucampia cornuta*	*(278) Corythodinium tesselatum*	248.67	-40.41	0.00	-0.28	-0.05	-2.27	-8.67	-0.05
chl-din	*(233) Halosphaera viridis*	*(386) Prorocentrum cordatum*	0.23	128.82	-6.49	0.00	-8.87	11.60	0.18	-21.59
bac-bac	*(18) Attheya septentrionalis*	*(217) Thalassiosira antarctica*	-18.96	642.46	-0.21	0.00	-0.74	-15.91	-1.24	-0.12
bac-bac	*(35) Chaetoceros atlanticus*	*(121) Entomoneis paludosa*	-108.97	24.01	-6.72	0.00	-25.40	-1.29	-0.37	-11.55
bac-din	*(195) Rhabdonema arcuatum*	*(317) Gymnodinium uberrimum*	-30.17	878.84	-39.96	0.00	-0.03	-10.38	-8.06	5.19
bac-din	*(209) Stephanopyxis nipponica*	*(317) Gymnodinium uberrimum*	-97.53	883.05	-19.57	0.00	-0.16	-33.55	-6.52	-7.18
bac-din	*(209) Stephanopyxis nipponica*	*(399) Protoperidinium brevipes*	-371.00	16.92	-4.48	0.00	-15.95	-42.67	-21.70	-14.79
bac-bac	*(217) Thalassiosira antarctica*	*(223) Thalassiosira leptopus*	-72.89	5619.36	-81.14	-6.15	-10.25	-2.69	-0.51	-4.64
bac-chl	*(217) Thalassiosira antarctica*	*(233) Halosphaera viridis*	-53.94	4126.42	-4.28	0.00	-6.38	-4.25	-0.08	-35.97
bac-din	*(222) Thalassiosira hyalina*	*(317) Gymnodinium uberrimum*	-56.63	34.21	-104.40	0.00	-1.28	-8.17	-1.19	1.98
chl-din	*(233) Halosphaera viridis*	*(317) Gymnodinium uberrimum*	-8.31	2640.30	-31.83	0.00	-0.11	-0.77	-0.74	-6.65
bac-bac	*(35) Chaetoceros atlanticus*	*(217) Thalassiosira antarctica*	-677.30	3668.04	-261.13	-50.21	-71.11	-24.29	-5.85	-0.01
bac-bac	*(46) Chaetoceros convolutus*	*(217) Thalassiosira antarctica*	-373.99	197.53	-203.71	-8.52	-52.98	-28.86	-4.91	-0.24
bac-bac	*(174) Odontella aurita*	*(217) Thalassiosira antarctica*	-663.51	1835.57	-193.43	-18.07	-129.99	-20.39	-4.90	0.08
bac-bac	*(195) Rhabdonema arcuatum*	*(217) Thalassiosira antarctica*	-377.33	2203.00	-90.02	-0.47	-22.89	-90.60	-0.78	-0.77
din-din	*(329) Impagidinium aculeatum*	*(386) Prorocentrum cordatum*	-1.77	686.00	-0.16	-39.23	-0.04	-1.85	-101.68	-4.50
bac-bac	*(188) Proboscia inermis*	*(223) Thalassiosira leptopus*	-1.71	1578.31	0.00	-3.48	-0.42	-49.70	-57.86	-33.68
bac-chl	*(223) Thalassiosira leptopus*	*(233) Halosphaera viridis*	19.56	3484.58	0.00	0.00	-0.21	-2.44	29.34	-1.46
bac-din	*(223) Thalassiosira leptopus*	*(367) Pentapharsodinium dalei*	-691.12	11.59	0.00	-62.32	-24.12	-74.43	0.18	-0.77
bac-bac	*(60) Chaetoceros furcellatus*	*(217) Thalassiosira antarctica*	-20.06	758.37	0.00	-0.94	-12.02	6.71	-0.02	0.22
bac-din	*(60) Chaetoceros furcellatus*	*(317) Gymnodinium uberrimum*	-697.23	61.06	0.00	-10.43	-7.48	-4.84	-0.55	-59.73
bac-din	*(135) Fragilariopsis oceanica*	*(389) Prorocentrum lima*	-52.67	72.58	0.00	-119.15	-24.37	-16.26	-63.83	-0.43
bac-bac	*(161) Membraneis challengeri*	*(188) Proboscia inermis*	0.38	1503.61	0.00	-65.97	-3.58	-33.02	-110.13	-8.25
bac-chl	*(161) Membraneis challengeri*	*(233) Halosphaera viridis*	-20.87	6345.64	0.00	-25.01	-2.24	-12.66	-0.72	-2.66
bac-din	*(174) Odontella aurita*	*(317)Gymnodinium uberrimum*	-626.91	312.90	-27.73	-125.77	-0.83	-10.71	-0.48	0.00
bac-din	*(174) Odontella aurita*	*(389) Prorocentrum lima*	-58.51	11.44	0.00	-7.14	-40.57	-6.50	-2.63	-0.12
bac-din	*(188) Proboscia inermis*	*(398) Protoperidinium bipes*	-0.58	1374.13	0.00	-1.71	-7.34	-1.00	-7.79	-4.55
bac-din	*(217) Thalassiosira antarctica*	*(330) Impagidinium patulum*	-0.59	1383.75	0.00	-116.93	-60.50	-11.18	0.72	2.53
bac-din	*(217) Thalassiosira antarctica*	*(367) Pentapharsodinium dalei*	-62.53	243.49	0.00	-40.14	-250.99	-17.52	0.05	0.67
bac-din	*(217) Thalassiosira antarctica*	*(389) Prorocentrum lima*	-9.03	2577.30	0.00	-85.94	-3.29	-77.92	0.10	0.12
chl-din	*(233) Halosphaera viridis*	*(398)Protoperidinium bipes*	6.33	2435.37	0.00	-0.65	-1.05	-1.08	0.02	-9.07
bac-din	*(35) Chaetoceros atlanticus*	*(317) Gymnodinium uberrimum*	-257.06	1862.73	-49.01	-47.61	-1.55	1.67	0.53	0.00
bac-din	*(46) Chaetoceros convolutus*	*(317) Gymnodinium uberrimum*	-121.03	25.82	-40.28	-87.12	-0.80	0.02	-17.15	0.00
bac-bac	*(121) Entomoneis paludosa*	*(179) Pleurosigma directum*	-214.61	313.34	-54.38	-0.02	-12.20	-66.92	-166.92	0.00
bac-din	*(121) Entomoneis paludosa*	*(331) Impagidinium sphaericum*	-42.44	55.32	-1.75	-9.94	-31.10	-99.99	-301.31	0.00
bac-din	*(161) Membraneis challengeri*	*(398) Protoperidinium bipes*	-263.17	2332.83	-0.17	6.40	-1.00	-3.59	-0.94	0.00
bac-din	*(215) Thalassiosira angulata*	*(317) Gymnodinium uberrimum*	-86.72	1622.28	-30.45	-66.70	-1.44	-0.23	-0.67	0.00
bac-din	*(218) Thalassiosira decipiens*	*(317) Gymnodinium uberrimum*	-352.71	498.00	-38.19	-7.45	-1.08	-5.64	-2.83	0.00
bac-din	*(219) Thalassiosira eccentrica*	*(317) Gymnodinium uberrimum*	-68.29	644.72	-16.41	-29.44	-0.88	0.89	0.07	0.00
bac-din	*(223) Thalassiosira leptopus*	*(317) Gymnodinium uberrimum*	-47.47	3189.12	-19.72	-1.24	-0.47	1.21	-0.05	0.00
bac-din	*(223) Thalassiosira leptopus*	*(331) Impagidinium sphaericum*	-10.95	96.07	-0.33	-2.00	-0.89	-6.62	-6.28	0.00
dic-din	*(237) Dictyocha fibula*	*(317) Gymnodinium uberrimum*	-356.56	624.30	-50.63	-68.47	-1.19	1.50	-5.81	0.00
dic-din	*(237) Dictyocha fibula*	*(389) Prorocentrum lima*	-82.64	21.04	-28.50	-171.41	-63.37	-94.74	-61.23	0.00
din-din	*(317) Gymnodinium uberrimum*	*(398) Protoperidinium bipes*	-1220.05	4286.93	-3.86	6.46	-0.05	-14.49	-1.15	0.00
bac-bac	*(56) Chaetoceros dichaeta*	*(121) Entomoneis paludosa*	-190.81	1160.16	-459.66	-4.91	-0.14	-34.88	-1.79	0.00
bac-din	*(113) Dactyliosolen antarcticus*	*(367) Pentapharsodinium dalei*	-26.80	5.44	-38.79	0.34	-427.66	2.56	0.36	0.00
bac-bac	*(121) Entomoneis paludosa*	*(161) Membraneis challengeri*	-183.96	381.92	-110.45	-0.90	-31.93	-5.80	-7.71	0.00
bac-din	*(121) Entomoneis paludosa*	*(386) Prorocentrum cordatum*	-2.60	26.88	-5.25	8.07	-4.09	-28.55	-0.30	0.00
bac-din	*(185) Podosira stelligera*	*(320) Gyrodinium fusiforme*	-621.84	268.91	-352.65	-233.51	-46.90	-52.54	1.43	0.00
bac-din	*(230) Trieres chinensis*	*(367) Pentapharsodinium dalei*	-123.43	1.42	-65.64	-0.43	-303.54	-25.93	-14.49	0.00
bac-bac	*(99) Corethron pennatum*	*(184) Pleurosigma simonsenii*	-17.56	123.14	-17.46	0.00	-1.76	-206.40	-59.03	0.00
bac-bac	*(161) Membraneis challengeri*	*(209) Stephanopyxis nipponica*	-269.41	1780.57	-131.26	0.00	-0.13	-4.59	-1.42	0.00
bac-bac	*(184) Pleurosigma simonsenii*	*(185) Podosira stelligera*	-0.39	1.09	-63.80	0.00	-17.85	-118.45	2.04	0.00

(*continued on next page*)

Table A.20 (*continued*)

bac-din	*184) Pleurosigma simonsenii*	*(398) Protoperidinium bipes*	-6.76	481.80	-2.15	0.00	-0.01	-73.03	-10.64	0.00
bac-bac	*(188) Proboscia inermis*	*(215) Thalassiosira angulata*	-2.67	681.72	-6.62	0.00	-9.02	-121.74	-6.33	0.00
bac-bac	*(188) Proboscia inermis*	*(218) Thalassiosira decipiens*	-4.70	353.11	-12.32	0.00	-8.23	-262.46	-40.64	0.00
bac-dic	*(188) Proboscia inermis*	*(237) Dictyocha fibula*	-0.42	568.38	0.28	0.00	0.00	-84.02	-18.80	0.00
bac-din	*(181) Pleurosigma elongatum*	*(310) Gymnodinium agiliforme*	-79.29	182.36	-21.09	0.00	-0.01	-1.98	-17.08	0.01
bac-din	*(185) Podosira stelligera*	*(310) Gymnodinium agiliforme*	-75.34	125.46	-5.90	0.00	-25.04	-7.50	-3.07	1.25
bac-din	*(21) Bacteriastrum comosum*	*(330) Impagidinium patulum*	-0.67	-0.13	9.98	0.00	-21.98	-18.18	-2.32	-0.05
din-din	*(329) Impagidinium aculeatum*	*(471) Tripos carriensis*	-9.83	-123.37	0.00	61.73	-4.41	-9.01	-49.92	-3.33
bac-hap	*(168) Nitzschia longissima*	*(527) Rhabdolithes claviger*	-0.02	-278.19	-189.25	0.00	72.37	-141.74	-17.80	-1.78
bac-bac	*(168) Nitzschia longissima*	*(207) Shionodiscus oestrupii*	-1.33	-67.34	-41.03	0.00	0.08	-96.43	-8.82	-0.04
bac-din	*(168) Nitzschia longissima*	*(462) Trinovantidinium applanatum*	-12.99	-7.15	-6.95	0.00	42.87	-18.65	-48.29	-1.03
bac-hap	*(9) Asterolampra marylandica*	*(527) Rhabdolithes claviger*	-21.68	-1.07	-93.80	0.00	0.13	-149.29	-14.43	0.00
din-din	*(367) Pentapharsodinium dalei*	*(471) Tripos carriensis*	-14.06	-0.33	0.00	-3.40	-62.86	12.87	-0.76	-95.07
din-din	*(248) Amphidinium acutissimum*	*(367) Pentapharsodinium dalei*	-67.08	-33.11	0.00	-3.67	-0.51	4.66	-0.30	-1.71
bac-din	*(219) Thalassiosira eccentrica*	*(471) Tripos carriensis*	-0.57	-109.23	-63.09	-2.73	-57.52	0.27	-3.58	0.00
bac-din	*(49) Chaetoceros curvisetus*	*(367) Pentapharsodinium dalei*	-13.37	-0.72	-3.89	0.00	-0.16	-13.95	17.34	-0.83
bac-din	*(51) Chaetoceros danicus*	*(367) Pentapharsodinium dalei*	-184.36	-3.77	-152.97	0.00	-22.31	-22.98	35.94	-0.82
bac-bac	*(51) Chaetoceros danicus*	*(196) Rhizosolenia acuminata*	-3.16	-24.35	-7.48	0.00	-18.44	-0.07	39.82	-1.36
bac-din	*(51) Chaetoceros danicus*	*(389) Prorocentrum lima*	-32.62	-1.96	-40.74	0.00	-0.17	-24.43	27.55	-0.10

A.14. Environmental conditions in our eight biomes

Fig. A.18 and A.19
Table A.21.

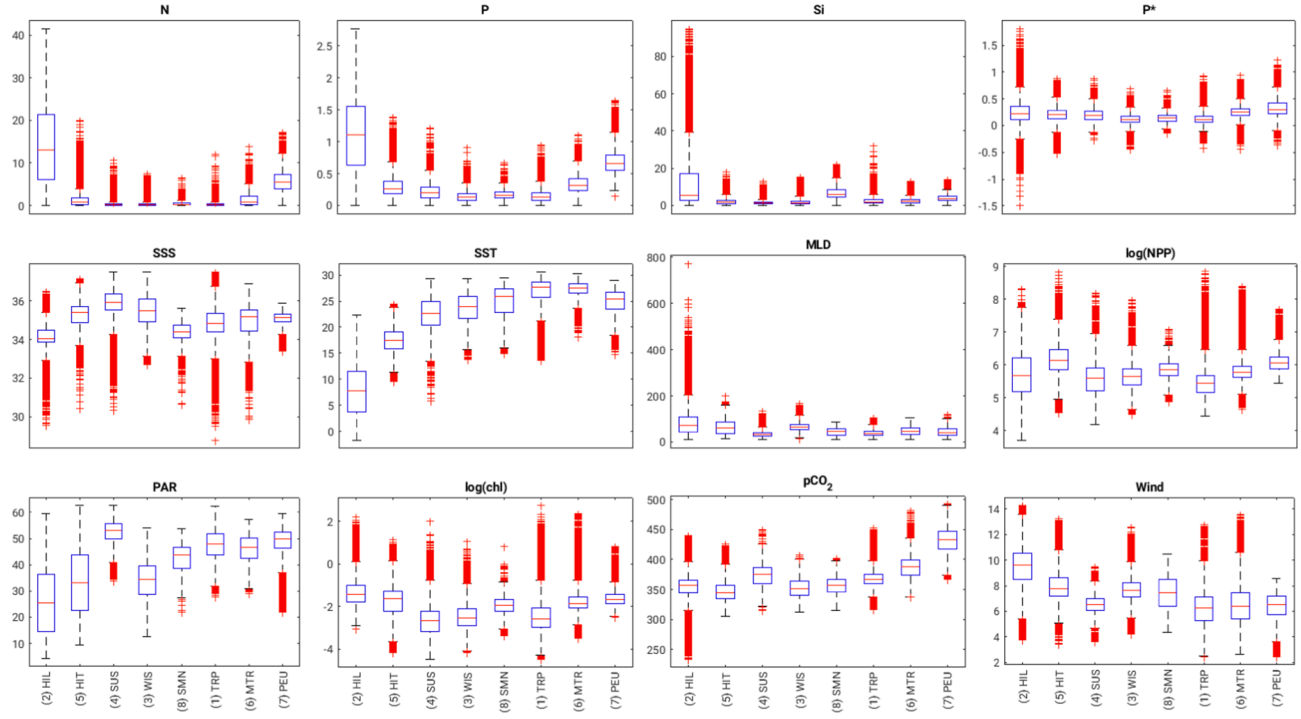


Fig. A.18. Environmental conditions found in each of our biomes across the twelve months together analyzed jointly, shown for individual environmental variables separately. The biomes are sorted according to their average latitude starting from high latitudes to the equator. The units of the environmental variables are: N, P, Si and P*/N* in $\frac{\mu mol}{L}$; SSS in *PSU*; SST in $^{\circ}C$; MLD in *m*; NPP in $\frac{mg\,C}{m^2 d}$; PAR in $\frac{Einstein}{m^2 d}$; chl in $\frac{mg}{m^3}$; pCO_2 in μatm; wind in $\frac{m}{s}$. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

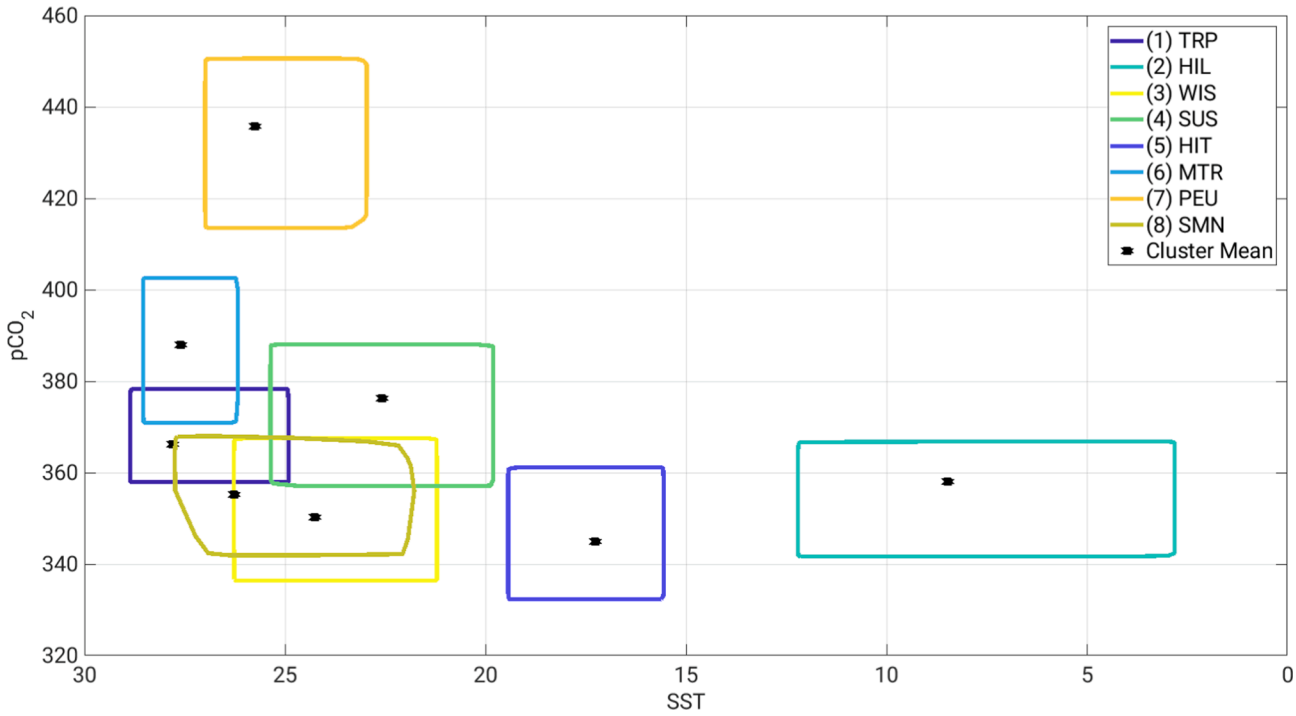


Fig. A.19. Two-dimensional projection of the pCO_2 and SST values found within our eight biomes across all twelve months analyzed jointly. The different colors denote the different biomes. The black dot shows the mean biome values in the two dimensions. Envelopes show the convex hull covering the interquartile range of pCO_2 and SST for each individual biome. The units of the environmental variables are: SST in $^{\circ}C$; pCO_2 in μatm. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Table A.21
Median and interquartile range of the environmental conditions found in each of our biomes across the twelve months together analyzes jointly. The units of the environmental variables are: N, P, Si and P*/N* in $\frac{\mu mol}{L}$; SSS in *PSU*; SST in $^{\circ}C$; MLD in m; NPP in $\frac{mgC}{m^2 d}$; PAR in $\frac{Einstein}{m^2 d}$; chl in $\frac{mg}{m^3}$; pCO_2 in μatm; wind in $\frac{m}{s}$. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

	TRP	HIL	WIS	SUS	HIT	MTR	PEU	SMN
N	0.14 ± 0.26	13.12 ± 15.3	0.12 ± 0.26	0.13 ± 0.32	0.74 ± 1.49	0.76 ± 1.98	5.55 ± 3.3	0.28 ± 0.45
P	0.13 ± 0.12	1.1 ± 0.91	0.13 ± 0.12	0.2 ± 0.17	0.27 ± 0.2	0.32 ± 0.18	0.66 ± 0.23	0.16 ± 0.1
Si	1.92 ± 1.81	5.6 ± 14.61	1.47 ± 1.57	1.16 ± 1.03	1.72 ± 1.92	2.08 ± 1.63	3.74 ± 2.46	6.02 ± 4.2
P*	0.12 ± 0.12	0.22 ± 0.25	0.12 ± 0.12	0.18 ± 0.16	0.2 ± 0.16	0.25 ± 0.12	0.29 ± 0.21	0.14 ± 0.1
SSS	34.83 ± 0.93	34.05 ± 0.62	35.48 ± 1.19	35.9 ± 0.84	35.42 ± 0.81	35.16 ± 1.08	35.14 ± 0.41	34.38 ± 0.63
SST	27.62 ± 2.99	7.83 ± 7.78	23.94 ± 4.1	22.7 ± 4.57	17.44 ± 3.12	27.5 ± 1.86	25.42 ± 3.32	25.96 ± 4.58
MLD	36 ± 18.94	71.19 ± 64.4	65.53 ± 24.63	31.41 ± 16.45	61.01 ± 49.58	48.87 ± 30.88	41.44 ± 30.81	46.51 ± 31.72
logNPP	5.45 ± 0.52	5.67 ± 1.03	5.65 ± 0.49	5.6 ± 0.7	6.14 ± 0.61	5.79 ± 0.34	6.05 ± 0.35	5.86 ± 0.38
PAR	47.96 ± 8.06	25.32 ± 21.91	34.28 ± 10.87	53.03 ± 5.9	33.13 ± 21.26	46.56 ± 7.75	49.78 ± 6.2	43.64 ± 7.98
logCHL	-2.57 ± 0.89	-1.44 ± 0.76	-2.55 ± 0.79	-2.67 ± 0.98	-1.65 ± 0.95	-1.86 ± 0.52	-1.67 ± 0.41	-1.97 ± 0.56
pCO2	366.77 ± 15.21	356.84 ± 20.33	351.87 ± 24.38	375.16 ± 25.99	344.87 ± 23.2	388.49 ± 24.32	433.35 ± 29.52	357.16 ± 21.26
wind	6.23 ± 1.88	9.61 ± 2.04	7.61 ± 1.1	6.54 ± 0.92	7.75 ± 1.44	6.38 ± 2.09	6.54 ± 1.4	7.47 ± 2.09

A.15. Comparison between our biomes and previous partitionings

Fig. A.20.

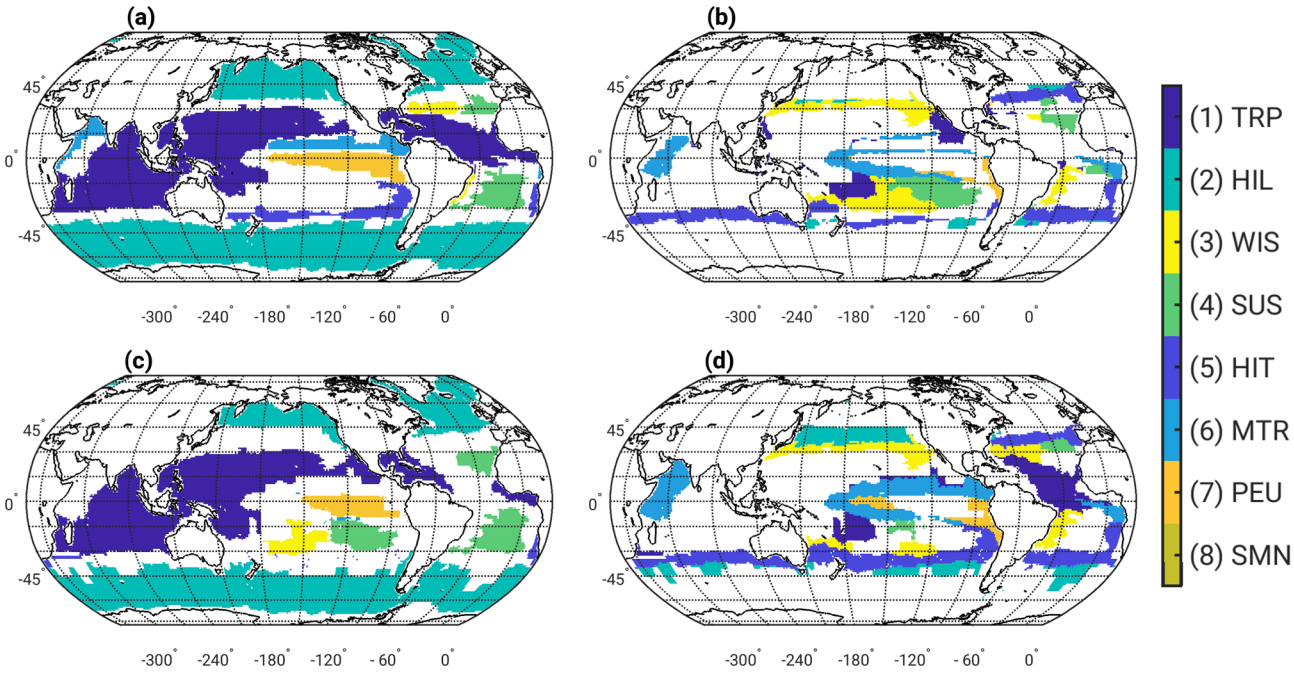


Fig. A.20. The overlaps between biomes presented in this study (derived from 536 phytoplankton species) and the ocean partitionings of Longhurst (2007) or the regions defined by Costello et al. (2017), respectively. (a) Shows the regions where we found a spatial overlap between our biomes and the 57 provinces defined by Longhurst, and (b) shows the areas where we did not find such an overlap between our biomes and the provinces. (c) Shows the regions where we found a spatial overlap between our biomes and bioregions defined by Costello et al. (2017), and (d) shows the areas where we did not find such an overlap. The following abbreviations were used: TRP for TRoPical biome; HIL for HIgh Latitude biome; WIS for WInter Subtropical biome; SUS for SUmmer Subtropical biome; HIT for HIgh latitude Transition biome; MTR for Monsoon and Tropical biome; PEU for Pacific Equatorial Upwelling biome; SMN for Seasonal MoNsoon.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.pocean.2021.102530.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscipl. Rev. Comput. Stat. 2 (4), 433–459. https://doi.org/10.1002/wics.101.

Aggarwal, C.C., 2015. Data classification. In: Data Mining. Springer International Publishing, pp. 285–344. https://doi.org/10.1007/978-3-319-14142-8_10.

Aggarwal, C.C., Hinneburg, A., Keim, D.A., 2001. On the surprising behavior of distance metrics in high dimensional space. In: Database Theory — ICDT 2001. Springer, Berlin, Heidelberg, pp. 420–434. https://doi.org/10.1007/3-540-44503-x_27.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43 (6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x.

Angermeier, P.L., Karr, J.R., 2019. Ecological health indicators. In: Encyclopedia of Ecology. Elsevier, pp. 391–401. https://doi.org/10.1016/b978-0-12-409548-9.10926-1.

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K., Gombos, D., 2011. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bull. Am. Meteorol. Soc. 92 (2), 157–174. https://doi.org/10.1175/2010bams2946.1.

Bailey, R.G., 1998. Ecoregions: the ecosystem geography of the oceans and continents, vol. 36. American Library Association. https://doi.org/10.5860/choice.36-1591.

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. Proc. Nat. Acad. Sci. 113 (11), 2964–2969. https://doi.org/10.1073/pnas.1519080113.

Baumann, K.-H., Saavedra-Pellitero, M., Böckel, B., Ott, C., 2016. Morphometry, biogeography and ecology of calcidiscus and umbilicosphaera in the south atlantic. Revue de Micropaléontologie 59 (3), 239–251. https://doi.org/10.1016/j.revmic.2016.03.001.

Beale, M.H., Hagan, M.T., Demuth, H.B., 2017. Neural network toolbox user's guide. The Mathworks Inc.

Beaugrand, G., 2009. Decadal changes in climate and ecosystems in the north atlantic ocean and adjacent seas. Deep Sea Res. Part II 56 (8–10), 656–673. https://doi.org/10.1016/j.dsr2.2008.12.022.

Beaugrand, G., Edwards, M., Brander, K., Luczak, C., Ibanez, F., 2008. Causes and projections of abrupt climate-driven ecosystem shifts in the north atlantic. Ecol. Lett. 11 (11), 1157–1168.

Behrenfeld, M.J., Falkowski, P.G., 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. 42 (1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001 https://doi.org/10.4319/lo.1997.42.1.0001.

Bellman, R.E., 2015. Adaptive control processes: a guided tour, vol. 2045. Princeton University Press. https://books.google.ch/books?id=iwbWCgAAQBAJ.

Benedetti, F., Guilhaumon, F., Adloff, F., Ayata, S.-D., 2018. Investigating uncertainties in zooplankton composition shifts under climate change scenarios in the mediterranean sea. Ecography 41 (2), 345–360 https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02434.

Biller, S.J., Berube, P.M., Dooley, K., Williams, M., Satinsky, B.M., Hackl, T., Hogle, S.L., Coe, A., Bergauer, K., Bouman, H.A., Browning, T.J., Corte, D.D., Hassler, C., Hulston, D., Jacquot, J.E., Maas, E.W., Reinthaler, T., Sintes, E., Yokokawa, T., Chisholm, S.W., 2018. Marine microbial metagenomes sampled across space and time. Sci. Data 5 (1). https://doi.org/10.1038/sdata.2018.176 https://doi.org/10.1038/sdata.2018.176.

Boltovskoy, D., Correa, N., 2016. Biogeography of radiolaria polycystina (protista) in the world ocean. Prog. Oceanogr. 149, 82–105. https://doi.org/10.1016/j.pocean.2016.09.006.

Bopp, L., Aumont, O., Cadule, P., Alvain, S., Gehlen, M., 2005. Response of diatoms distribution to global warming and potential implications: A global model study. Geophys. Res. Lett. 32 (19) https://doi.org/10.1029/2005gl023653.

Boyce, D.G., Lewis, M.R., Worm, B., 2010. Global phytoplankton decline over the past century. Nature 466 (7306), 591–596. https://doi.org/10.1038/nature09268.

Boyd, P.W., Strzepek, R., Fu, F., Hutchins, D.A., 2010. Environmental control of open-ocean phytoplankton groups: Now and in the future. Limnol. Oceanogr. 55 (3), 1353–1376. https://doi.org/10.4319/lo.2010.55.3.1353.

Bracher, A., Taylor, M.H., Taylor, B., Dinter, T., Röttgers, R., Steinmetz, F., 2015. Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Sci. 11 (1), 139–158. https://doi.org/10.5194/os-11-139-2015.

Breece, M.W., Fox, D.A., Dunton, K.J., Frisk, M.G., Jordaan, A., Oliver, M.J., 2016. Dynamic seascapes predict the marine occurrence of an endangered species: Atlantic SturgeonAcipenser oxyrinchus oxyrinchus. Methods Ecol. Evol. 7 (6), 725–733. https://doi.org/10.1111/2041-210x.12532.

Breiman, L., 2004. Consistency for a simple model of random forests. Statistical department. University of California at Berkeley. Technical Report (670).

Briggs, J.C., Bowen, B.W., 2011. A realignment of marine biogeographic provinces with particular reference to fish distributions. J. Biogeogr. 39 (1), 12–30. https://doi.org/10.1111/j.1365-2699.2011.02613.x https://doi.org/10.1111/j.1365-2699.2011.02613.x.

Briones, E.E., Rice, J., Ardron, J., 2009. Global open oceans and deep seabed (goods) biogeographic classification. UNESCO, IOC, p. 54.

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

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Change Biol. 16 (4), 1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x.

Buitenhuis, E.T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S.C., Leblanc, K., Quéré, C.L., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R., Swan, C., 2013. MAREDAT: towards a world atlas of MARine ecosystem DATa. Earth Syst. Sci. Data 5 (2), 227–239. https://doi.org/10.5194/essd-5-227-2013.

Cabré, A., Shields, D., Marinov, I., Kostadinov, T.S., 2016. Phenology of size-partitioned phytoplankton carbon-biomass from ocean color remote sensing and CMIP5 models. Front. Mar. Sci. 3. URL https://doi.org/10.3389/fmars.2016.00039.

Caldow, C., Monaco, M.E., Pittman, S.J., Kendall, M.S., Goedeke, T.L., Menza, C., Kinlan, B.P., Costa, B.M., 2015. Biogeographic assessments: A framework for information synthesis in marine spatial planning. Mar. Policy 51, 423–432. https://doi.org/10.1016/j.marpol.2014.07.023.

Cardoso, S.J., Nabout, J.C., Farjalla, V.F., Lopes, P.M., Bozelli, R.L., Huszar, V.L.M., Roland, F., 2017. Environmental factors driving phytoplankton taxonomic and functional diversity in amazonian floodplain lakes. Hydrobiologia 802 (1), 115–130. https://doi.org/10.1007/s10750-017-3244-x.

Cermeño, P., Falkowski, P.G., 2009. Controls on diatom biogeography in the ocean. Science 325(5947), 1539–1541. URL https://doi.org/10.1126/science.1174159.

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. Mar. Ecol. Prog. Ser. 488, 1–9. URL https://doi.org/10.3354/meps10443.

Cermeño, P., Teixeira, I.G., Branco, M., Figueiras, F.G., Marañón, E., 2014. Sampling the limits of species richness in marine phytoplankton communities. J. Plankton Res. 36 (4), 1135–1139. URL https://doi.org/10.1093/plankt/fbu033.

Cermeño, P., Chouciño, P., Fernández-Castro, B., Figueiras, F.G., Marañön, E., Marrasé, C., Mouriño-Carballido, B., Pérez-Lorenzo, M., Rodríguez-Ramos, T., Teixeira, I.G., Vallina, S.M., 2016. Marine primary productivity is driven by a selection effect. Frontiers in Marine Science 3, 173. URL https://www.frontiersin.org/article/10.3389/fmars.2016.00173.

Chaffron, S., Delage, E., Budinich, M., Vintache, D., Henry, N., Nef, C., Ardyna, M., Zayed, A., Junger, P., Galand, P., Lovejoy, C., Murray, A., Sarmento, H., Acinas, S., Babin, M., Iudicone, D., Jaillon, O., Karsenti, E., Wincker, P., Karp-Boss, L., Sullivan, M., Bowler, C., de Vargas, C., and, D.E., 2020. Environmental vulnerability of the global ocean plankton community interactome. URL Preprint at https://doi.org/10.1101/2020.11.09.375295.

Clarke, K.R., 1993. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 18 (1), 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x.

Clements, F., 1917. The development and structure of biotic communities. J. Ecol. 5, 120–121.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20 (1), 37–46. https://doi.org/10.1177/001316446002000104.

Costello, M.J., Tsai, P., Wong, P.S., Cheung, A.K.L., Basher, Z., Chaudhary, C., 2017. Marine biogeographic realms and species endemicity. Nat. Commun. 8 (1) https://doi.org/10.1038/s41467-017-01121-2.

Cotti-Rausch, B.E., Lomas, M.W., Lachenmyer, E.M., Goldman, E.A., Bell, D.W., Goldberg, S.R., Richardson, T.L., 2016. Mesoscale and sub-mesoscale variability in phytoplankton community composition in the sargasso sea. Deep Sea Res. Part I 110, 106–122. https://doi.org/10.1016/j.dsr.2015.11.008.

Crowder, L.B., 2006. SUSTAINABILITY: Resolving mismatches in u.s. ocean governance. Science 313 (5787), 617–618.

Cunningham, P., 2008. Dimension Reduction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 91–112. https://doi.org/10.1007/978-3-540-75171-7_4.

Davis, C.O., Kavanaugh, M., Letelier, R., Bissett, W.P., Kohler, D., 2007. Spatial and spectral resolution considerations for imaging coastal waters. In: Frouin, R.J. (Ed.), Coastal Ocean Remote Sensing. SPIE. https://doi.org/10.1117/12.734288.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet siimple technique for ecological data analysis. Ecology 81 (11), 3178–3192 https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9658(2000)081%5B3178%3ACARTAP%5D2.0.CO%3B2.

Decelle, J., Probert, I., Bittner, L., Desdevises, Y., Colin, S., de Vargas, C., Gali, M., Simo, R., Not, F., 2012. An original mode of symbiosis in open ocean plankton. Proc. Nat. Acad. Sci. 109 (44), 18000–18005. https://doi.org/10.1073/pnas.1212303109.

De Monte, S., Soccodato, A., Alvain, S., d'Ovidio, F., 2013. Can we detect oceanic biodiversity hotspots from space? ISME J. 7 (10), 2054–2056. https://doi.org/10.1038/ismej.2013.72.

Deutsch, C., Sarmiento, J.L., Sigman, D.M., Gruber, N., Dunne, J.P., 2007. Spatial coupling of nitrogen inputs and losses in the ocean. Nature 445 (7124), 163–167. https://doi.org/10.1038/nature05392.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348 (6237), 1261605. https://science.sciencemag.org/content/348/6237/1261605.

Devred, E., Sathyendranath, S., Platt, T., 2007. Delineation of ecological provinces using ocean colour radiometry. Mar. Ecol. Prog. Ser. 346, 1–13. https://doi.org/10.3354/meps07149.

Dodge, J.D., Marshall, H.G., 1994. Biogeographic analysis of the armored planktonic dinoflagellate ceratium in the north atlantic and adjacent seas. J. Phycol. 30 (6), 905–922. https://doi.org/10.1111/j.0022-3646.1994.00905.x.

Doney, S.C., 2010. The growing human footprint on coastal and open-ocean biogeochemistry. Science 328 (5985), 1512–1516. https://doi.org/10.1126/science.1185198.

Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., Galindo, H.M., Grebmeier, J.M., Hollowed, A.B., Knowlton, N., et al., 2011. Climate change impacts on marine ecosystems. URL https://doi.org/10.1146/annurev-marine-041911-111611.

Dunning, T., 1993. Accurate methods for the statistics of surprise and coincidence. Comput. Linguist. 19 (1), 61–74. http://dl.acm.org/citation.cfm?id=972450.972454.

Ekman, S., 1953. Zoogeography of the Sea, vol. 3. URL https://doi.org/10.1093/aibsbulletin/3.2.17-e.

Ellison, A.M., 2019. Foundation species, non-trophic interactions, and the value of being common. iScience 13, 254–268. UR. https://doi.org/10.1016/j.isci.2019.02.020.

Endo, H., Ogata, H., Suzuki, K., 2018. Contrasting biogeography and diversity patterns between diatoms and haptophytes in the central pacific ocean. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-29039-9.

Evert, S., 2009. Corpora and collocations. Corpus Linguistics: Int. Handbook 2, 1212–1248.

Falkowski, P.G., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science 281 (5374), 200–206. https://doi.org/10.1126/science.281.5374.200.

Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., Taylor, F., 2004. The evolution of modern eukaryotic phytoplankton. Science 305 (5682), 354–360.

Fay, A.R., McKinley, G.A., 2014. Global open-ocean biomes: mean and temporal variability. Earth Syst. Sci. Data 6 (2), 273–284. https://doi.org/10.5194/essd-6-273-2014.

Fendereski, F., Vogt, M., Payne, M.R., Lachkar, Z., Gruber, N., Salmanmahiny, A., Hosseini, S.A., 2014. Biogeographic classification of the caspian sea. Biogeosciences 11 (22), 6451–6470. http://www.biogeosciences.net/11/6451/2014/.

Ficetola, G.F., Mazel, F., Thuiller, W., 2017. Global determinants of zoogeographical boundaries. Nat. Ecol. Evol. 1 (4) https://doi.org/10.1038/s41559-017-0089.

Field, C.B., 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.

Gamfeldt, L., Roger, F., 2017. Revisiting the biodiversity–ecosystem multifunctionality relationship. Nat. Ecol. Evol. 1 (7) https://doi.org/10.1038/s41559-017-0168.

Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Baranova, O., Zweng, M., Reagan, J., Johnson, D., 2013. World Ocean Atlas 2013. In: Levitus, S. (Ed.), Dissolved Inorganic Nutrients (phosphate, nitrate, silicate), vol. 4, A. Mishonov Technical Ed. NOAA Atlas NESDIS 76, 25 pp.

Gibson, D.J., Ely, J.S., Collins, S.L., 1999. The core-satellite species hypothesis provides a theoretical basis for grime's classification of dominant, subordinate, and transient species. J. Ecol. 87(6), 1064–1067. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2745.1999.00424.x.

Goebel, N.L., Edwards, C.A., Follows, M.J., Zehr, J.P., 2014. Modeled diversity effects on microbial ecosystem functions of primary production, nutrient uptake, and remineralization. Ecology 95 (1), 153–163. https://doi.org/10.1890/13-0421.1.

Gregor, L., Lebehot, A.D., Kok, S., Scheel Monteiro, P.M., 2019. A comparative assessment of the uncertainties of global surface ocean co_2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall? Geosci. Model Develop. 12 (12), 5113–5136. https://www.geosci-model-dev.net/12/5113/2019/.

Griffiths, H.J., Barnes, D.K.A., Linse, K., 2009. Towards a generalized biogeography of the southern ocean benthos. J. Biogeogr. 36 (1), 162–177. https://doi.org/10.1111/j.1365-2699.2008.01979.x.

Gruber, N., 2011. Warming up, turning sour, losing breath: ocean biogeochemistry under global change. Philos. Trans. Roy. Soc. Math. Phys. Eng. Sci. 369 (1943), 1980–1996. https://doi.org/10.1098/rsta.2011.0003.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J.R., et al., 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature. https://doi.org/10.1038/nature16942.

Guiry, M., Guiry, G., 2017. Algaebase. http://www.algaebase.org.

Halpern, B.S., Longo, C., Scarborough, C., Hardy, D., Best, B.D., Doney, S.C., Katona, S.K., McLeod, K.L., Rosenberg, A.A., Samhouri, J.F., 2014. Assessing the health of the

us west coast with a regional-scale application of the ocean health index. Plos One 9 (6), e98995. https://doi.org/10.1371/journal.pone.0098995.
Hanski, I., 1982. Dynamics of regional distribution: the core and satellite species hypothesis. Oikos 210–221. https://doi.org/10.2307/3544021.
Hardman-Mountford, N.J., Hirata, T., Richardson, K.A., Aiken, J., 2008. An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. Remote Sens. Environ. 112 (8), 3341–3352. https://doi.org/10.1016/j.rse.2008.02.016.
Hastie, T., Tibshirani, R., 2017. Generalized additive models. Generalized Additive Models. Routledge 136–173. https://doi.org/10.1201/9780203753781-6.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning, Vol. 1. Springer, New York. https://doi.org/10.1007/b94608.
Hattam, C., Atkins, J.P., Beaumont, N., Brger, T., Bhnke-Henrichs, A., Burdon, D., de Groot, R., Hoefnagel, E., Nunes, P.A., Piwowarczyk, J., Sastre, S., Austen, M.C., 2015. Marine ecosystem services: Linking indicators to their classification. Ecol. Ind. 49, 61–75. https://doi.org/10.1016/j.ecolind.2014.09.026.
Higgins, S.I., Buitenwerf, R., Moncrieff, G.R., 2016. Defining functional biomes and monitoring their change globally. Glob. Change Biol. 22 (11), 3583–3593. https://doi.org/10.1111/gcb.13367.
Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., Fabre, P.-H., Graham, C.H., Graves, G.R., Jønsson, K.A., Nogués-Bravo, D., Wang, Z., Whittaker, R.J., Fjeldså, J., Rahbek, C., 2012. An update of wallace's zoogeographic regions of the world. Science 339 (6115), 74–78. https://doi.org/10.1126/science.1228282.
IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. URL www.climatechange2013.org.
Irwin, A.J., Oliver, M.J., 2009. Are ocean deserts getting larger? Geophys. Res. Lett. 36 (18) https://doi.org/10.1029/2009gl039883.
Irwin, A.J., Nelles, A.M., Finkel, Z.V., 2012. Phytoplankton niches estimated from field data. Limnol. Oceanogr. 57 (3), 787–797.
Jain, A.K., Murty, M.N., Flynn, P.J., Sep. 1999. Data clustering: A review. ACM Comput. Surv. 31 (3), 264–323. https://doi.org/10.1145/331499.331504.
Jönsson, B.F., Watson, J.R., 2016. The timescales of global surface-ocean connectivity. Nat. Commun. 7 (1) https://doi.org/10.1038/ncomms11239.
Kavanaugh, M.T., Hales, B., Saraceno, M., Spitz, Y.H., White, A.E., Letelier, R.M., 2014. Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. Prog. Oceanogr. 120, 291–304. https://doi.org/10.1016/j.pocean.2013.10.013.
Kavanaugh, M.T., Oliver, M.J., Chavez, F.P., Letelier, R.M., Muller-Karger, F.E., Doney, S. C., 2016. Seascapes as a new vernacular for pelagic ocean monitoring, management and conservation. ICES J. Mar. Sci. 73 (7), 1839–1850. https://doi.org/10.1093/icesjms/fsw086.
Keith, S.A., Baird, A.H., Hughes, T.P., Madin, J.S., Connolly, S.R., 2013. Faunal breaks and species composition of indo-pacific corals: the role of plate tectonics, environment and habitat distribution. Proc. Roy. Soc. B: Biol. Sci. 280 (1763), 20130818. https://doi.org/10.1098/rspb.2013.0818.
Kiviluoto, K., 1996. Topology preservation in self-organizing maps. In: Neural Networks, 1996, IEEE International Conference on, vol. 1. IEEE, pp. 294–299. https://doi.org/10.1109/icnn.1996.548907.
Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78 (9), 1464–1480. https://doi.org/10.1109/5.58325.
Kohonen, T., 2001. The basic SOM. In: Self-Organizing Maps. Springer, Berlin, Heidelberg, pp. 105–176.
Kramer, C.Y., 1956. Extension of multiple range tests to group means with unequal numbers of replications. Biometrics 12 (3), 307. https://doi.org/10.2307/3001469.
Kulbicki, M., Parravicini, V., Bellwood, D.R., Arias-Gonzàlez, E., Chabanet, P., Floeter, S. R., Friedlander, A., McPherson, J., Myers, R.E., Vigliola, L., Mouillot, D., 2013. Global biogeography of reef fishes: A hierarchical quantitative delineation of regions. PLoS ONE 8 (12), e81847. https://doi.org/10.1371/journal.pone.0081847.
Lai, F., Jutfelt, F., Nilsson, G.E., 2015. Altered neurotransmitter function in CO2-exposed stickleback (gasterosteus aculeatus): a temperate model species for ocean acidification research. Conservation. Physiology 3 (1), cov018. https://doi.org/10.1093/conphys/cov018.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159. https://doi.org/10.2307/2529310.
Landschützer, P., Gruber, N., Bakker, D.C.E., Schuster, U., 2014. Recent variability of the global ocean carbon sink. Glob. Biogeochem. Cycles 28 (9), 927–949. https://doi.org/10.1002/2014GB004853.
Layden, O.K., 2020. okomarov/schemaball. URL https://www.github.com/okomarov/schemaball.
Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E., Marshall, H., Peloquin, J., Piontkovski, S., Poulton, A.J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., van Leeuwe, M.A., Varela, M., Widdicombe, C., Yallop, M., 2012. A global diatom database – abundance, biovolume and biomass in the world ocean. Earth Syst. Sci. Data 4 (1), 149–165. https://doi.org/10.5194/essd-4-149-2012.
Lévy, M., Franks, P.J.S., Smith, K.S., 2018. The role of submesoscale currents in structuring marine ecosystems. Nat. Commun. 9 (1) https://doi.org/10.1038/s41467-018-07059-3.
Lewison, R., Hobday, A.J., Maxwell, S., Hazen, E., Hartog, J.R., Dunn, D.C., Briscoe, D., Fossette, S., O'Keefe, C.E., Barnes, M., Abecassis, M., Bograd, S., Bethoney, N.D., Bailey, H., Wiley, D., Andrews, S., Hazen, L., Crowder, L.B., 2015. Dynamic ocean management: Identifying the critical ingredients of dynamic approaches to ocean resource management. Bioscience 65 (5), 486–498. https://doi.org/10.1093/biosci/biv018.
Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A.M., Coppola, L., Cornejo-Castillo, F.M., d'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.-J., Guidi, L., Lara, E., Pesant, S., Royo-Llonch, M., Salazar, G., Sánchez, P., Sebastian, M., Souffreau, C., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L., Weissenbach, J., Wincker, P., Acinas, S.G., Sunagawa, S., Bork, P., Sullivan, M.B., Karsenti, E., Bowler, C., de Vargas, C., Raes, J., 2015. Determinants of community structure in the global plankton interactome. Science 348 (6237). https://science.sciencemag.org/content/348/6237/1262073.
Litchman, E., 2007. Resource competition and the ecological success of phytoplankton. In: Evolution of Primary Producers in the Sea. Elsevier, pp. 351–375. https://doi.org/10.1016/b978-012370518-1/50017-5.
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. J. Ecol. 103 (6), 1384–1396. https://doi.org/10.1111/1365-2745.12438.
Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Baranova, O.K., Zweng, M.M., Paver, C.R., Reagan, J.R., Johnson, D.R., Hamilton, M., Seidov, D., 2013. World Ocean Atlas 2013, In: Levitus, S. (Ed.), Temperature, vol. 1, A. Mishonov Technical Ed. NOAA Atlas NESDIS 73, 40 pp.
Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W.H., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Romac, S., Shalchian-Tabrizi, K., Simon, N., Stoeck, T., Santini, S., Siano, R., Wincker, P., Zingone, A., Richards, T.A., de Vargas, C., Massana, R., 2014. Patterns of rare and abundant marine microbial eukaryotes. Curr. Biol. 24 (8), 813–821. https://doi.org/10.1016/j.cub.2014.02.050.
Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. Prog. Oceanogr. 36 (2), 77–167. https://doi.org/10.1016/0079-6611(95)00015-1.
Longhurst, A.R., 2007. Ecological geography of the sea. Academic Press. URL https://doi.org/10.1016/b978-0-12-455521-1.x5000-1.
Lyons, K.G., Brigham, C.A., Traut, B., Schwartz, M.W., 2005. Rare species and ecosystem functioning. Conserv. Biol. 19 (4), 1019–1024. https://doi.org/10.1111/j.1523-1739.2005.00106.x.
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. Proc. Nat. Acad. Sci. 113 (11), E1516–E1525. https://doi.org/10.1073/pnas.1509523113.
Manning, P., van der Plas, F., Soliveres, S., Allan, E., Maestre, F.T., Mace, G., Whittingham, M.J., Fischer, M., 2018. Redefining ecosystem multifunctionality. Nat. Ecol. Evol. 2 (3), 427–436. https://doi.org/10.1038/s41559-017-0461-7.
Maxwell, S.M., Hazen, E.L., Lewison, R.L., Dunn, D.C., Bailey, H., Bograd, S.J., Briscoe, D.K., Fossette, S., Hobday, A.J., Bennett, M., Benson, S., Caldwell, M.R., Costa, D.P., Dewar, H., Eguchi, T., Hazen, L., Kohin, S., Sippel, T., Crowder, L.B., 2015. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. Mar. Policy 58, 42–50. https://doi.org/10.1016/j.marpol.2015.03.014.
Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., W++est, R.O., Zimmermann, N.E., Elith, J., 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography 37 (12), 1267–1281. URL https://doi.org/10.1111/ecog.00845.
Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50 (2), 159–179. https://doi.org/10.1007/bf02294245.
Moisan, T.A., Rufty, K.M., Moisan, J.R., Linkswiler, M.A., 2017. Satellite observations of phytoplankton functional type spatial distributions, phenology, diversity, and ecotones. Front. Mar. Sci. 4. URL https://doi.org/10.3389/fmars.2017.00189.
Mojica, K.D.A., van de Poll, W.H., Kehoe, M., Huisman, J., Timmermans, K.R., Buma, A. G.J., van der Woerd, H.J., Hahn-Woernle, L., Dijkstra, H.A., Brussaard, C.P.D., 2015. Phytoplankton community structure in relation to vertical stratification along a north-south gradient in the northeast atlantic ocean. Limnol. Oceanogr. 60 (5), 1498–1521. https://doi.org/10.1002/lno.10113.
Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the kappa statistic. Ecol. Model. 62 (4), 275–293. https://doi.org/10.1016/0304-3800(92)90003-w.
Montégut, C.d., 2004. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. J. Geophys. Res. 109 (C12).
Moorthi, S.D., Countway, P.D., Stauffer, B.A., Caron, D.A., 2006. Use of quantitative real-time PCR to investigate the dynamics of the red tide dinoflagellate lingulodinium polyedrum. Microb. Ecol. 52 (1), 136–150. https://doi.org/10.1007/s00248-006-9030-3.
Morán, X.A.G., López-Urrutia, Á., Calvo-Díaz, A., Li, W.K.W., 2010. Increasing importance of small phytoplankton in a warmer ocean. Glob. Change Biol. 16 (3), 1137–1144. https://doi.org/10.1111/j.1365-2486.2009.01960.x.
NASA, G., 2018a. Sea-viewing wide field-of-view sensor (seawifs) chlorophyll data; 2018 reprocessing. nasa ob.daac, greenbelt, md, usa. https://oceancolor.gsfc.nasa.gov/data/10.5067/ORBVIEW-2/SEAWIFS/L3B/CHL/2018/.
NASA, G., 2018b. Sea-viewing wide field-of-view sensor (seawifs) photosynthetically available radiation data; 2018 reprocessing. nasa ob.daac, greenbelt, md, usa. https://oceancolor.gsfc.nasa.gov/data/10.5067/ORBVIEW-2/SEAWIFS/L3B/PAR/2018/.
Navarro, G., Alvain, S., Vantrepotte, V., Huertas, I., 2014. Identification of dominant phytoplankton functional types in the mediterranean sea based on a regionalized

remote sensing approach. Remote Sens. Environ. 152, 557–575. https://doi.org/10.1016/j.rse.2014.06.029.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. J. Roy. Stat. Soc. Ser. A (Gen.) 135 (3), 370.
O'Brien, C., Peloquin, J., Vogt, M., Heinle, M., Gruber, N., Ajani, P., Andruleit, H., Arístegui, J., Beaufort, L., Estrada, M., et al., 2013. Global marine plankton functional type biomass distributions: coccolithophores. Earth Syst. Sci. Data 5 (2), 259–276.
O'Brien, C.J., Vogt, M., Gruber, N., 2016. Global coccolithophore diversity: Drivers and future change. Prog. Oceanogr. 140, 27–42. https://doi.org/10.1016/j.pocean.2015.10.003.
Ocean Productivity, 2017. Custom products. URL http://www.science.oregonstate.edu/ocean.productivity/custom.php.
Oliver, M.J., 2004. Bioinformatic approaches for objective detection of water masses on continental shelves. J. Geophys. Res. 109 (C7) https://doi.org/10.1029/2003jc002072.
Oliver, M.J., Irwin, A.J., 2008. Objective global ocean biogeographic provinces. Geophys. Res. Lett. 35 (15) https://doi.org/10.1029/2008gl034238.
Padisák, J., 1994. Identification of relevant time-scales in non-equilibrium community dynamics: conclusions from phytoplankton surveys. New Zealand J. Ecol. 169–176. http://www.jstor.org/stable/24066771.
Partensky, F., Hess, W.R., Vaulot, D., 1999. Prochlorococcus, a marine photosynthetic prokaryote of global significance. Microbiol. Mol. Biol. Rev. 63 (1), 106–127. https://doi.org/10.1128/mmbr.63.1.106-127.1999.
Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197. https://doi.org/10.1890/07-2153.1.
Ptacnik, R., Solimini, A.G., Andersen, T., Tamminen, T., Brettum, P., Lepistö, L., Willén, E., Rekolainen, S., Mar 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. Proc. Nat. Acad. Sci. 105 (13), 5134–5138. https://doi.org/10.1073/pnas.0708328105.
Racault, M.-F., Quéré, C.L., Buitenhuis, E., Sathyendranath, S., Platt, T., 2012. Phytoplankton phenology in the global ocean. Ecol. Ind. 14 (1), 152–163. https://doi.org/10.1016/j.ecolind.2011.07.010.
Reid, P.C., Johns, D.G., Edwards, M., Starr, M., Poulin, M., Snoeijs, P., 2007. A biological consequence of reducing arctic ice cover: arrival of the pacific diatom neodenticula seminae in the north atlantic for the first time in 800 000 years. Glob. Change Biol. 13 (9), 1910–1921. https://doi.org/10.1111/j.1365-2486.2007.01413.x.
Reygondeau, G., Maury, O., Beaugrand, G., Fromentin, J.M., Fonteneau, A., Cury, P., 2011. Biogeography of tuna and billfish communities. J. Biogeogr. 39 (1), 114–129. https://doi.org/10.1111/j.1365-2699.2011.02582.x.
Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., Maury, O., 2013. Dynamic biogeochemical provinces in the global ocean. Glob. Biogeochem. Cycles 27 (4), 1046–1058. https://doi.org/10.1002/gbc.20089.
Ricotta, C., 2002. Bridging the gap between ecological diversity indices and measures of biodiversity with shannon's entropy: comment to izsák and papp. Ecol. Model. 152 (1), 1–3. http://www.sciencedirect.com/science/article/pii/S030438 0001004689.
Ricotta, C., 2005. Through the jungle of biological diversity. Acta Biotheoretica 53 (1), 29–38.
Righetti, D., Vogt, M., Gruber, N., Psomas, A., Zimmermann, N.E., 2019a. 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., Zimmermann, N.E., Gruber, N., 2019b. Phytobase: A global synthesis of open ocean phytoplankton occurrences. Earth Syst. Sci. Data Discuss. 2019, 1–39. https://www.earth-syst-sci-data-discuss.net/essd-2019159/.
Ringelberg, J.J., Zimmermann, N.E., Weeks, A., Lavin, M., Hughes, C.E., 2020. Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. Glob. Ecol. Biogeogr. 29 (7), 1100–1113. https://doi.org/10.1111/geb.13089.
Rivero-Calle, S., Gnanadesikan, A., Castillo, C.E.D., Balch, W.M., Guikema, S.D., 2015. Multidecadal increase in north atlantic coccolithophores and the potential role of rising CO2. Science 350 (6267), 1533–1537. https://doi.org/10.1126/science.aaa8026.
Sal, S., Löpez-Urrutia, Irigoien, X., Harbour, D.S., Harris, R.P., 2013. Marine microplankton diversity database. Ecology 94(7), 1658–1658. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/13-0236.1.
Salvador, S., Chan, P., 2004. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: 16th IEEE International Conference on Tools with Artificial Intelligence. IEEE Comput. Soc. https://doi.org/10.1109/ictai.2004.50.
Sarmento, H., Montoya, J.M., Vázquez-Domínguez, E., Vaqué, D., Gasol, J.M., 2010. Warming effects on marine microbial food web processes: how far can we go when it comes to predictions? Philos. Trans. Roy. Soc. B: Biol. Sci. 365 (1549), 2137–2149. https://doi.org/10.1098/rstb.2010.0045.
Sarmiento, J.L., Gruber, N., 2006. Ocean Biogeochemical Dynamics. Princeton University Press. URL https://doi.org/10.1515/9781400849079.
Sarmiento, J.L., Slater, R., Barber, R., Bopp, L., Doney, S.C., Hirst, A., Kleypas, J., Matear, R., Mikolajewicz, U., Monfray, P., et al., 2004. Response of ocean ecosystems to climate warming. Glob. Biogeochem. Cycles 18 (3). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GB002134.
Sathyendranath, S., Aiken, J., Alvain, S., Barlow, R., Bouman, H., Bracher, A., Brewin, R., Bricaud, A., Brown, C.W., Ciotti, A.M., Clementson, L.A., Craig, S.E., Devred, E., Hardman-Mountford, N., Hirata, T., Hu, C., Kostadinov, T.S., Lavender, S., Loisel, H., Moore, T.S., Morales, J., Mouw, C.B., Nair, A., Raitsos, D., Roesler, C., Shutler, J.D., Sosik, H.M., Soto, I., Stuart, V., Subramaniam, A., Uitz, J., 2014. Phytoplankton functional types from space. International Ocean-Colour Coordinating Group, Dartmouth, Nova Scotia, B2Y 4A2, Canada, pp. 1–156.
Schimel, J., 1995. Ecosystem consequences of microbial diversity and community structure. In: Ecological Studies. Springer, Berlin, Heidelberg, pp. 239–254.
Schott, F.A., Xie, S.-P., McCreary, J.P., 2009. Indian ocean circulation and climate variability. Rev. Geophys. 47 (1). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007RG000245.
Schwartz, M.W., Brigham, C.A., Hoeksema, J.D., Lyons, K.G., Mills, M.H., van Mantgem, P., 2000. Linking biodiversity to ecosystem function: implications for conservation ecology. Oecologia 122 (3), 297–305. https://doi.org/10.1007/s004420050035.
Ser-Giacomi, E., Zinger, L., Malviya, S., Vargas, C.D., Karsenti, E., Bowler, C., Monte, S. D., 2018. Ubiquitous abundance distribution of non-dominant plankton across the global ocean. Nat. Ecol. Evol. 2 (8), 1243–1249. https://doi.org/10.1038/s41559-018-0587-2.
Silva de Miranda, P.L., Oliveira-Filho, A.T., Pennington, R.T., Neves, D.M., Baker, T.R., Dexter, K.G., 2018. Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical south america. Glob. Ecol. Biogeogr. 27 (8), 899–912. https://doi.org/10.1111/geb.12749.
Sinha, S., Singh, T.N., Singh, V.K., Verma, A.K., 2009. Epoch determination for neural network by self-organized map (SOM). Comput. Geosci. 14 (1), 199–206. https://doi.org/10.1007/s10596-009-9143-0.
Smith, V.H., Foster, B.L., Grover, J.P., Holt, R.D., Leibold, M.A., deNoyelles, F., 2005. Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. Proc. Nat. Acad. Sci. 102 (12), 4393–4396. https://www.pnas.org/content/102/12/4393.
Smith, N., Kessler, W.S., Cravatte, S., Sprintall, J., Wijffels, S., Cronin, M.F., Sutton, A., Serra, Y.L., Dewitte, B., Strutton, P.G., Hill, K., Sen Gupta, A., Lin, X., Takahashi, K., Chen, D., Brunner, S., 2019. Tropical pacific observing system. Front. Mar. Sci. 6, 31. https://www.frontiersin.org/article/10.3389/fmars.2019.00031.
Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. BioScience 57(7), 573–583. URL https://doi.org/10.1641/b570707.
Spalding, M.D., Agostini, V.N., Rice, J., Grant, S.M., 2012. Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. Ocean Coast. Manage. 60, 19–30. https://doi.org/10.1016/j.ocecoaman.2011.12.016.
Sreeush, M.G., Valsala, V., Pentakota, S., Prasad, K.V.S.R., Murtugudde, R., 2018. Biological production in the indian ocean upwelling zones –part 1: refined estimation via the use of a variable compensation depth in ocean carbon models. Biogeosciences 15 (7), 1895–1918. https://doi.org/10.5194/bg-15-1895-2018.
Strom, S.L., 2008. Microbial ecology of ocean biogeochemistry: A community perspective. Science 320 (5879), 1043–1045. https://doi.org/10.1126/science.1153527.
Sugar, C.A., James, G.M., 2003. Finding the number of clusters in a dataset. J. Am. Stat. Assoc. 98 (463), 750–763. https://doi.org/10.1198/016214503000000666.
Sutton, T.T., Clark, M.R., Dunn, D.C., Halpin, P.N., Rogers, A.D., Guinotte, J., Bograd, S. J., Angel, M.V., Perez, J.A.A., Wishner, K., Haedrich, R.L., Lindsay, D.J., Drazen, J. C., Vereshchaka, A., Piatkowski, U., Morato, T., Błachowiak-Samołyk, K., Robison, B. H., Gjerde, K.M., Pierrot-Bults, A., Bernal, P., Reygondeau, G., Heino, M., 2017. A global biogeographic classification of the mesopelagic zone. Deep Sea Res. Part I 126, 85–102. https://doi.org/10.1016/j.dsr.2017.05.006.
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.
Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466 (7310), 1098–1101. https://doi.org/10.1038/nature09329.
Townsend, C.R., Begon, M., Harper, J.L., 2008. Essentials of ecology, third ed. Blackwell Publishing, Malden, MA.
Townsend, M., Davies, K., Hanley, N., Hewitt, J.E., Lundquist, C.J., Lohrer, A.M., 2018. The challenge of implementing the marine ecosystem service concept. Front. Mar. Sci. 5 https://doi.org/10.3389/fmars.2018.00359.
Vallina, S.M., Follows, M.J., Dutkiewicz, S., Montoya, J.M., Cermeño, P., Loreau, M., 2014. Global relationship between phytoplankton diversity and productivity in the ocean. Nat. Commun. 5(1). https://doi.org/10.1038/ncomms5299.
van der Spoel, S., 1994. The basis for boundaries in pelagic biogeography. Prog. Oceanogr. 34 (2–3), 121–133. https://doi.org/10.1016/0079-6611(94)90005-1.
Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Trans. Neural Netw. 11 (3) https://doi.org/10.1109/72.846731.
Vichi, M., Allen, J.I., Masina, S., Hardman-Mountford, N.J., 2011. The emergence of ocean biogeochemical provinces: A quantitative assessment and a diagnostic for model evaluation. Global Biogeochem. Cycles 25 (2).
Vidal, D.E., Horne, A.J., 2003. Inheritance of mercury tolerance in the aquatic oligochaete tubifex tubifex. Environ. Toxicol. Chem. 22 (9), 2130. https://doi.org/10.1897/02-407.
Villar, E., Farrant, G.K., Follows, M., Garczarek, L., Speich, S., Audic, S., Bittner, L., Blanke, B., Brum, J.R., Brunet, C., Casotti, R., Chase, A., Dolan, J.R., d'Ortenzio, F., Gattuso, J.-P., Grima, N., Guidi, L., Hill, C.N., Jahn, O., Jamet, J.-L., Le Goff, H., Lepoivre, C., Malviya, S., Pelletier, E., Romagnan, J.-B., Roux, S., Santini, S., Scalco, E., Schwenck, S.M., Tanaka, A., Testor, P., Vannier, T., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Boss, E., de Vargas, C., Gorsky, G., Ogata, H., Pesant, S., Sullivan, M.B., Sunagawa, S., Wincker, P., Karsenti, E., Bowler, C., Not, F., Hingamp, P., Iudicone, D., 2015. Environmental characteristics of agulhas rings affect interocean

plankton transport. Science 348 (6237). https://science.sciencemag.org/content/348/6237/1261447.
Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., Bachelerie, P., Rosnet, T., Pelletier, E., Sunagawa, S., Hingamp, P., 2018. The ocean gene atlas: exploring the biogeography of plankton genes online. Nucleic Acids Res. 46 (W1), W289–W295. https://doi.org/10.1093/nar/gky376.
Villarino, E., Watson, J.R., Jönsson, B., Gasol, J.M., Salazar, G., Acinas, S.G., Estrada, M., Massana, R., Logares, R., Giner, C.R., et al., 2018. Large-scale ocean connectivity and planktonic body size. Nat. Commun. 9 (1), 1–13. https://doi.org/10.1038/s41467-017-02535-8.
Vincent, F., Bowler, C., 2020. Diatoms are selective segregators in global ocean planktonic communities. mSystems 5 (1).
Vogt, M., O'Brien, C., Peloquin, J., Schoemann, V., Breton, E., Estrada, M., Gibson, J., Karentz, D., Leeuwe, M.A.V., Stefels, J., Widdicombe, C., Peperzak, L., 2012. Global marine plankton functional type biomass distributions: Phaeocystis spp. Earth Syst. Sci. Data 4 (1), 107–120. https://doi.org/10.5194/essd-4-107-2012.
Voigt, W., Perner, J., Davis, A.J., Eggers, T., Schumacher, J., Bährmann, R., Fabian, B., Heinrich, W., Köhler, G., Lichter, D., Marstaller, R., Sander, F.W., 2003. Trophic levels are differentially sensitive to climate. Ecology 84 (9), 2444–2453. https://doi.org/10.1890/02-0266.
Wahl, A., Gries, S.T., 2018. Multi-word expressions: A novel computational approach to their bottom-up statistical extraction. In: Lexical Collocation Analysis. Springer International Publishing, pp. 85–109. https://doi.org/10.1007/978-3-319-92582-0_5.
Wallace, A.R., 1876. The geographical distribution of animals: with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface, vol. 1. Cambridge University Press.
Waters, J.M., Wernberg, T., Connell, S.D., Thomsen, M.S., Zuccarello, G.C., Kraft, G.T., Sanderson, J.C., West, J.A., Gurgel, C.F., 2010. Australia's marine biogeography revisited: Back to the future? Austral Ecol. 35 (8), 988–992. https://doi.org/10.1111/j.1442-9993.2010.02114.x.
Weber, M., Teeling, H., Huang, S., Waldmann, J., Kassabgy, M., Fuchs, B.M., Klindworth, A., Klockow, C., Wichels, A., Gerdts, G., Amann, R., Glöckner, F.O., Dec 2010. Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. ISME J. 5 (5), 918–928. https://doi.org/10.1038/ismej.2010.180.
Westberry, T., Behrenfeld, M.J., Siegel, D.A., Boss, E., 2008. Carbon-based primary productivity modeling with vertically resolved photoacclimation. Global Biogeochem. Cycles 22 (2).
Whittaker, Robert, 1970. Current concepts in biology series. In: Communities and ecosystems. Collier-Macmillan, London.
Whittaker, K.A., Rynearson, T.A., 2017. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. Proc. Nat. Acad. Sci. 114 (10), 2651–2656. https://doi.org/10.1073/pnas.1612346114.
Wilks, D.S., 2011. Statistical methods in the atmospheric sciences, International geophysics series, 3rd ed., vol. 100. Academic Press, Amsterdam.
Xu, R., II, D.W., 2005. Survey of clustering algorithms. IEEE Trans. Neural Netw. 16(3), 645–678. URL https://doi.org/10.1109/tnn.2005.845141.
Yang, Q., Blanco, N.E., Hermida-Carrera, C., Lehotai, N., Hurry, V., Strand, Å., 2020. Two dominant boreal conifers use contrasting mechanisms to reactivate photosynthesis in the spring. Nat. Commun. 11 (1) https://doi.org/10.1038/s41467-019-13954-0.
Zaki, M.J., Meira Jr, W., 2014. Classification assessment. In: Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, pp. 548–584. https://doi.org/10.1017/CBO9780511810114.023.
Zhao, Q., Basher, Z., Costello, M.J., 2019. Mapping near surface global marine ecosystems through cluster analysis of environmental data. Ecol. Res. https://doi.org/10.1111/1440-1703.12060.
Zweng, M., Reagan, J., Antonov, J., Locarnini, R., Mishonov, A., Boyer, T., Garcia, H., Baranova, O., Johnson, D., Seidov, D., Biddle, M., 2013. World Ocean Atlas 2013. In: Levitus, S. (Ed.), vol. 2, A. Mishonov Technical Ed. NOAA Atlas NESDIS 74, 39 pp.