Three major mesoplanktonic communities resolved by in situ imaging in the upper 500 m of the global ocean

Thelma Panaïotis, Marcel Babin, Tristan Biard, François Carlotti, Laurent Coppola, Lionel Guidi, Helena Hauss, Lee Karp-Boss, Rainer Kiko, Fabien Lombard, et al.

Three major mesoplanktonic communities resolved by *in situ* imaging in the upper 500 m of the global ocean

Thelma Panaïotis, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Marcel Babin, Takuvik International Research Laboratory (IRL 3376), CNRS and Université Laval, Québec, Qc, Canada
Tristan Biard, Laboratoire d'Océanologie et de Géosciences, Univ. Littoral Côte d'Opale, Univ. Lille, CNRS, IRD, Wimereux, France
François Carlotti, Mediterranean Institute of Oceanography, Aix-Marseille Université, Université de Toulon, CNRS, IRD, UMR 7294, Marseille, France
Laurent Coppola, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Lionel Guidi, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Helena Hauss, GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
Lee Karp-Boss, School of Marine Sciences, University of Maine, Orono, ME, USA
Rainer Kiko, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France & GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
Fabien Lombard, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Andrew MP McDonnell, Oceanography Department, University of Alaska Fairbanks, Fairbanks, AK, United States
Marc Picheral, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Andreas Rogge, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, D-27570 Bremerhaven, Germany
Anya M Waite, Ocean Frontier Institute and Department of Oceanography, Dalhousie University, Halifax, Canada
Lars Stemmann, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France
Jean-Olivier Irisson, Laboratoire d'Océanographie de Villefranche, Sorbonne Université, Villefranche-sur-Mer, France

Corresponding author: Thelma Panaïotis, thelma.panaiotis@imev-mer.fr

Abstract

Aim The distribution of mesoplankton communities have been poorly studied at global scale, especially from in situ instruments. This study aims to (1) describe the global distribution of mesoplankton communities in relation with their environment and (2) assess the ability of various environmental-based ocean regionalisations to explain the distribution of these communities.

Location Global ocean, 0 - 500 m depth.

Time period 2008 - 2019

Major taxa studied 28 groups of large mesoplanktonic and macroplanktonic organisms, covering Metazoa, Rhizaria and Cyanobacteria.

Methods From a global data set of 2500 vertical profiles making use of the Underwater Vision Profiler 5 (UVP5), an *in situ* imaging instrument, we studied the global distribution of large (> 600 µm) mesoplanktonic organisms. Among the 6.8 million imaged objects, 330,000 were large zooplanktonic organisms and phytoplankton colonies, the rest consisting of marine snow particles. Multivariate ordination (PCA) and clustering were used to describe patterns in community composition, while comparison with existing regionalisations was performed with regression methods (RDA).

Results Within the observed size range, epipelagic plankton communities were *Trichodesmium*-enriched in the intertropical Atlantic, Copepoda-enriched at high latitudes and in upwelling areas, and Rhizaria-enriched in oligotrophic areas. In the mesopelagic layer, Copepoda-enriched communities were also found at high latitudes and in the Atlantic Ocean, while Rhizaria-enriched communities prevailed in the Peruvian upwelling system and a few mixed communities were found elsewhere. The comparison between the distribution of these communities and a set of existing regionalisations of the ocean suggested that the structure of plankton communities described above is mostly driven by basin-level environmental conditions.

Main conclusions In both layers, three types of plankton communities emerged and seemed to be mostly driven by regional environmental conditions. This work sheds light on the role not only of metazoans, but also of unexpected large protists and cyanobacteria in structuring large mesoplankton communities.

Keywords biogeography, global ocean, in situ imagery, plankton communities, spatial distribution, *Trichodesmium*, Rhizaria, Copepoda

1 Introduction

Plankton are defined as organisms unable to swim against currents. This definition based on habitat rather than taxonomy thus results in a large taxonomic diversity (Caron et al., 2012; Ibarbalz et al., 2019), encompassing bacteria, protists, algae, as well as drifting animals. Therefore, planktonic organisms cover a large size range, spanning 6 orders of magnitude, from micrometre to metre (Lombard et al., 2019). Plankton supports oceanic food webs (Falkowski, 2012; Ware & Thomson, 2005) and plays a major role in biogeochemical cycles through the biological pump (Longhurst & Glen Harrison, 1989). As drifters, planktonic organisms are distributed worldwide but their distribution is shaped by the conditions of the water mass they are embedded in (Hays et al., 2005). Because these conditions strongly vary with latitude, both plankton diversity (Ibarbalz et al., 2019; Rombouts et al., 2009; Rutherford et al., 1999; Tittensor et al., 2010) and biomass (Ikeda, 1985) vary with latitude: high diversity and low biomass at low latitudes; low diversity and high biomass at high latitudes. Indeed, at low latitudes, higher temperatures increase the rate of metabolism, resulting in both shorter generation times and higher mutation rates, causing a higher diversity (Brown, 2014). Conversely, at high latitudes, more nutrients are available, enabling higher production and a higher biomass. Overall, plankton distribution is governed by both water mass movements (through dispersal) and environmental conditions, and the contribution of each effect might depend on the size of organisms (Sommeria-Klein et al., 2021; Villarino et al., 2018). Because they are sensitive to environmental conditions, planktonic organisms are also global change sentinels

(Beaugrand et al., 2009; Beaugrand et al., 2002; Hays et al., 2005). Studying plankton biography thus is relevant to understand anthropocene pelagic ecosystems.

In terrestrial biogeography, biomes rest on vegetation types, but also coincide with climatic zones and soil type distribution constraining plant growth (Brown & Lomolino, 1998). Compared to the terrestrial realm, assessing oceanic biogeography presents inherent difficulties: costly global scale offshore sampling, observing distribution varying in time and space in a three-dimensional and opaque environment. Early ocean biogeographies considered various taxa's distribution, including copepods, euphausiids, Rhizaria or phytoplankton (Spalding et al., 2012). Simultaneously, non-biological regionalisations were based on the physical environment: ocean currents, temperature, salinity, ice conditions (Spalding et al., 2012). Novel technologies, such as satellites, fostered ecological ocean geography: using surface chlorophyll *a* concentration computed from ocean colour (a proxy for phytoplankton concentration) new regionalisations emerged. However, most new approaches ignore organisms distribution: the 56 Longhurst Provinces (Longhurst, 1995, 2010) considered physical forcing (sea surface temperature, mixed layer depth...) as phytoplankton distribution regulator. Another widely used global synthetic regionalisation is based on latitudinal bands (Spalding et al., 2012). As explained above, it correlates with major environmental variables (temperature, light intensity...). Other regionalisations, the Word Marine Ecoregions (Spalding et al., 2007) or the Large Marine Ecosystems (Sherman, 2005) include biotic data, but focus on coastal areas only. In contrast, Costello et al., 2017 delineated marine biogeographic realms using the distribution of marine animals and plants, while Hofmann Elizondo et al., 2021 defined biomes using phytoplankton community composition as well as species co-occurrence patterns. Furthermore, these regionalisations focus on the epipelagic layer and are not suitable for less described deeper ocean, harder to sample and not necessarily linked to surface characteristics (Costello et al., 2018; Spalding et al., 2012). Few regionalisations targeted the mesopelagic: Reygondeau et al., 2018 suggested dividing it into 13 provinces, based on environmental variables, while defining the top and bottom boundaries of the mesopelagic layer dynamically, based on environmental conditions (light, density, carbon flux).

In summary, ocean biogeography was described through regionalisations, but mostly in the epipelagic layer or coastal areas. In comparison, widespread offshore areas, although crucial for biogeochemical cycles (Emerson et al., 1997) and the target of conservation through developing marine protected areas, were sampled much less, due to technical constraints. In a context of global warming, mesoplankton communities are facing changes in their abundance, size and diversity (Richardson, 2008), susceptible to affect higher trophic levels, with consequences for exploited marine resources such as fisheries. Thus, it is not only essential to describe mesoplankton communities at global scale, but also to assess the consistency between these and biogeochemistry-based regionalisations, typically used for the conservation and management of marine ecosystems (Rice et al., 2011).

Although a few organisms groups' spatial distribution – copepods (Beaugrand et al., 2013; Rombouts et al., 2009; Woodd-Walker et al., 2002), microorganisms (Fuhrman et al., 2008; Hörstmann et al., 2022; Li et al., 2004) or larger species assemblages (Brandão et al., 2021; Rutherford et al., 1999; Soviadan et al., 2022; Tittensor et al., 2010) – were described previously, consistency between these and biogeochemistry-based regionalisations remains unexplored.

Limited quantitative and basin-scale data about offshore planktonic organism distribution is available, especially for *in situ* imaging data. Traditional sampling tools (nets, pumps...) require lengthy taxonomic identification (Benfield et al., 2007), partly subjective (Culverhouse et al., 2003), therefore not scaling well to large spatiotemporal scales. Besides, they may damage fragile organisms (Remsen et al., 2004). New *in situ* cameras now image planktonic organisms in their natural environment and resolve their fine scale vertical distribution (Stemmann et al., 2008), while generating large datasets (Kiko et al.,

2022), homogenised by reviewing images (Irisson et al., 2022). These tools also allow studying fragile taxonomic groups: Rhizaria, whose contribution to global planktonic biomass was underestimated (Biard et al., 2016; Dennett et al., 2002; Drago et al., 2022). These approaches lack in taxonomic identification fineness, but compensate with identifications and data quantity consistency. Among these imaging systems, the Underwater Vision Profiler 5 (UVP5) images planktonic organisms and marine snow particles larger than 600 µm Equivalent Spherical Diameter (ESD) along vertical profiles (Picheral et al., 2010), therefore sampling large meso- and small macro-plankton, mostly comprising animals and some large phytoplankton colonies (this assemblage is referred as "plankton" in the following, for simplicity). Concentration and biovolume estimates from the UVP5 proved coherent with those from net samples for large (> 1 mm ESD) Arctic copepods, while smaller organisms were underestimated (Forest et al., 2012). Data from UVP5 was already used to estimate organic carbon vertical particle flux (Guidi et al., 2015) or study zooplankton distribution (Biard et al., 2016; Drago et al., 2022; Forest et al., 2012; Stemmann et al., 2008). Leveraging net data from the Tara Oceans expedition, Soviadan et al., 2022 detected a lower particle flux attenuation in Oxygen Minimum Zones (OMZ). UVP5 can also inform on planktonic organisms' individual traits, highlighting different size and activity patterns around Arctic ice melt zones (Vilgrain et al., 2021).

We study global scale plankton biogeography, leveraging several UVP5 datasets. We address the following questions: which types of large mesoplankton and macroplankton communities exist in the open ocean; how these communities differ between epipelagic and upper-mesopelagic layers; are these communities driven by rather large or small scale processes; do they correlate with environmental conditions? We first describe plankton communities structures; their relation to their immediate physical and biogeochemical environment. We then assess the ability of various physics and biogeochemistry-based regionalisations to describe these planktonic communities distribution.

2 Material and Methods

2.1 Data Collection

Data from multiple oceanographic campaigns (Figure S1, Table S1) (2008 - 2019) – when UVP5 vertical profiles were performed – was aggregated, creating a large dataset covering world's oceans. This *in situ* imaging system captures objects within an approximately 1 L volume, up to 20 Hz frequency during a CTD cast descending part (Conductivity, Temperature, Depth sensor) (Picheral et al., 2010). All objects larger than 100 µm ESD were measured for area and grey level. Images were saved for objects larger than ~600 µm ESD; this paper focuses on the latter part. In the following, we thus use the word "plankton" to refer to all the organisms detected by the UVP5, consisting mostly in mesozooplankton and macrozooplankton, as well as large phytoplankton colonies. The CTD provided temperature, salinity and also chlorophyll *a* fluorescence and oxygen profiles.

Satellite GlobColour[1] products completed the environmental dataset: averaged over one month on a 100 × 100 km area centred on UVP5's sampling site. Although this data resolution is low in terms of time and space, using higher resolution data provided a too great missing data proportion. Satellite data provided: surface chlorophyll *a* concentration; particulate backscattering coefficient (bbp); photosynthetically active radiation (PAR); diffuse attenuation coefficient (K_{dPAR}); particulate organic carbon (POC); particulate inorganic carbon (PIC). With these data, organism were associated with specific environmental conditions.

[1] http://globcolour.info

2.2 Data Processing

All UVP5 images entered the EcoTaxa web application (Picheral et al., 2017). They were classified as marine snow, artefact, badfocus, unidentified, or into several taxonomic groups. All objects were manually validated or corrected. Striving for consistency, a taxonomic sorting guide was published and circulated; difficult groups were reviewed by a single operator across all cruises. Human operators fully checked the resulting dataset, often several for each image. Data from 2500 fully validated profiles was retained (6.8 M objects, 330,000 classified as plankton). Differences in classification taxonomic depth among cruises caused some groups to be merged, obtaining a lower common denominator. Then, other groups were merged because they exhibited similar patterns in preliminary analysis (e.g., Copepoda and Copepoda-like), for a final list of 28 taxa (Figure 1). Finally, the same operator reviewed a random subsample of images from each final taxonomic group, with error rates at < 10% for all groups and < 2.5% for most (Drago et al., 2022). Although a normalised biomass size spectrum revealed an underestimation in the 600 µm - 1 mm size range, 80% of organisms in our dataset were over 1 mm in ESD, therefore accurately detected by UVP5. Furthermore, provided underestimation was consistent across taxonomic groups (we checked by inspecting dominant taxa's per-taxon size spectra), the community composition is little affected by the absolute concentrations underestimation.

Concentrations (L^{-1}) were computed per 5 m bins along each profile from object counts per class and imaged water volume. Concentration and biovolume were also computed for marine snow (objects > 600 µm identified as aggregates) and bulk particulate matter (all objects > 100 µm ESD imaged by the UVP, including both plankton and marine snow). Marine snow and bulk concentrations and biovolumes were considered environmental variables, as proxies for, respectively, organic matter amounts sinking from the upper layers to ocean depths and the overall trophic water mass state.

After removing abnormal values (codes for missing data, negative salinity. . .), profiles with more than 20% of missing data for any variable were ignored. All variables were linearly interpolated at a 1 m vertical resolution. Outliers were detected by computing the absolute deviation around a moving median along the profile (Leys et al., 2013) and removed. Smoothing was performed using a moving average. Potential density and apparent oxygen utilisation (AOU) were computed from temperature, salinity and oxygen concentration. Thermocline, halocline and pycnocline depths were calculated as depth of the largest variation in the relevant variable computed in a 5 m sliding window. The mixed layer depth (MLD) was computed at depth where density differed by more than 0.03 kg m^{-3} from the reference density in the 0-5 m surface layer (de Boyer Montégut et al., 2004). The deep chlorophyll maximum (DCM) and euphotic zone (Z_{eu}) depths were computed from the chlorophyll profile (Morel & Maritorena, 2001). The stratification index was computed as the difference in potential density between the surface and 250 m (below the pycnocline in most profiles). All 1 m precision profiles were binned over 5 m to match plankton data bins. Rare instances (1.7%) of missing satellite data were replaced by the corresponding variable average value.

Plankton and environmental data were averaged over two layers: epipelagic and upper mesopelagic. Instead of the commonly used fixed boundary (200 m) between these two layers (Costello & Breyer, 2017), we applied a dynamic definition. It was modified from Reygondeau et al., 2018 and meant to better represent the functional difference between the two layers. It was set as the deepest value among the mixed layer depth and the euphotic depth, hence delimiting the zone above which photosynthetic plankton activity is highest, resulting in an 88-metre median value for the epi-mesopelagic boundary (Q1 = 52 m, Q3 = 121 m, Figure S2). As numerous UVP5 profiles stopped at 500 m, we set the upper mesopelagic zone bottom at that depth. Any profile covering less than 80% of both layers' thickness was removed (3% and 29% of profiles for epipelagic and mesopelagic layers, respectively).

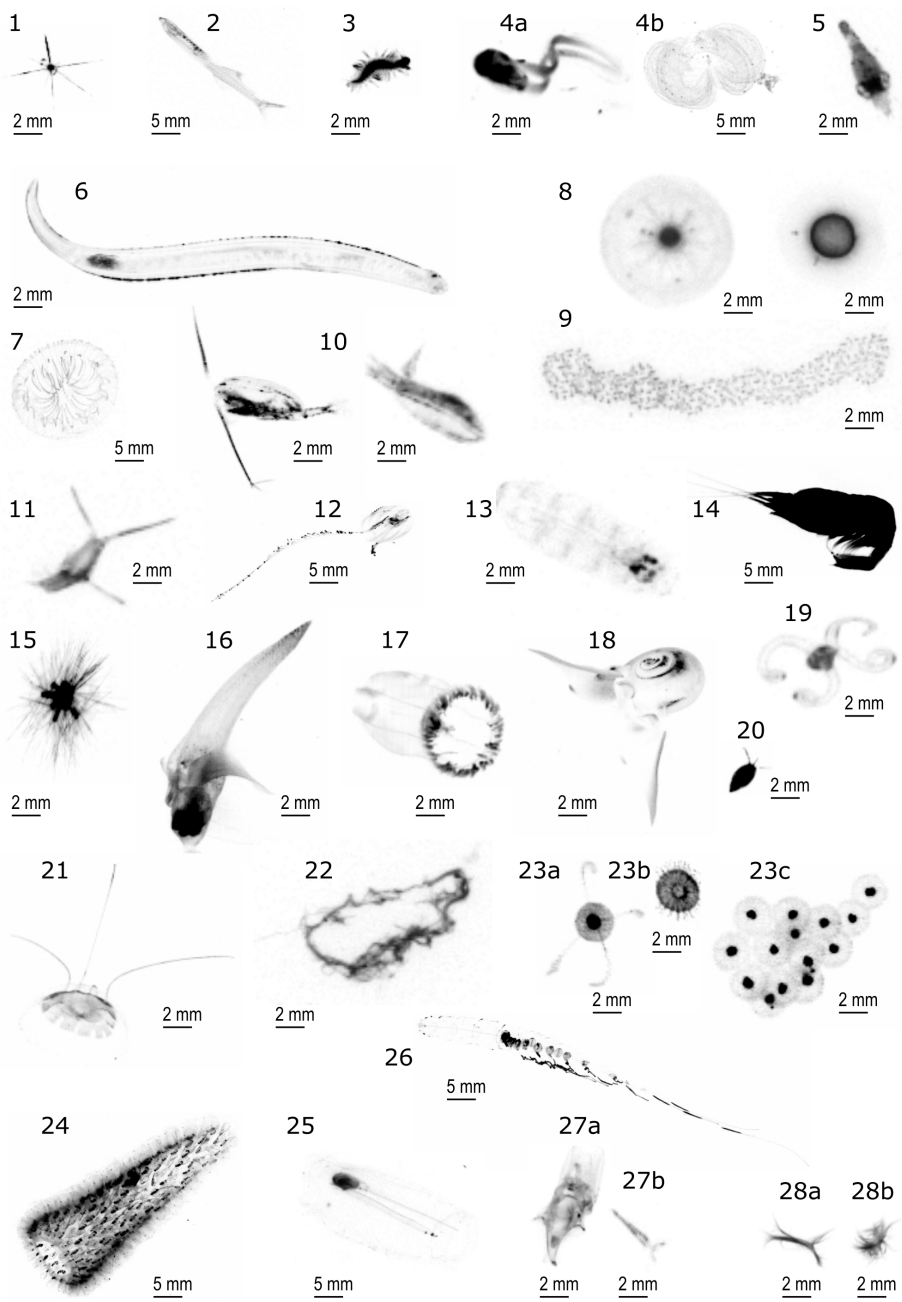

Figure 1: Examples of UVP5 images for selected taxonomic groups. 1: Acantharia, 2: Actinopterygii, 3: Annelida, 4: Appendicularia (4a: Appendicularia body, 4b: Appendicularia house), 5: Cephalopoda, 6: Chaetognatha, 7: Cnidaria others, 8: Collodaria, 9: colonial Collodaria, 10: Copepoda, 11: Crustacea others, 12: Ctenophora, 13: Doliolida, 14: Eumalacostraca, 15: Foraminifera, 16: Gymnosomata, 17: Hydrozoa others, 18: Limacinidae, 19: Mollusca others, 20: Ostracoda, 21: Narcomedusae, 22: Nostocales, 23: Phaeodaria (23a: Coelodendridae, 23b: Aulacantha, 23c: colonial Aulosphaeridae), 24: Pyrosoma, 25: Salpida, 26: Siphonophorae, 27: Thecosomata (27a: Cavoliniidae, 27b: Creseidae), 28: *Trichodesmium* (28a: tuft, 28b: puff).

2.3 Global Plankton Community Distribution Analysis

First, to synthetically describe plankton communities, a principal component analysis (PCA) was conducted on plankton concentration values. In each layer, the Hellinger transformation was applied to averaged plankton concentrations, as a compromise between absolute and relative concentrations, to focus on community composition dif-

ferences among profiles (Legendre & Legendre, 2012). This helped go beyond the well-known pattern of high latitudes higher concentrations and lower ones around the equator, hence optimising the taxonomic identification effort, while reducing very high abundances importance. Environmental variables were projected into the PCA space according to their correlation with plankton concentrations, after a log n+1 transformation for marine snow and bulk concentrations, to avoid over-representing some very high values. This allowed to visualise correlations directly on the PCA biplot and help with axes interpretation. Each profile's scores on the first five principal components (PC) helped perform a synoptic hierarchical agglomerative clustering (HAC), using the euclidean distance and Ward's criterion (Legendre & Legendre, 2012), in order to get a synoptic view of plankton communities. Using the PC scores, not the original data, preserved most of the variance, while removing noise. The resulting dendrogram (Figure S3) was separated into some main branches, based on inertia jumps, identifying broad plankton community types. Taxa proportions were computed for each plankton cluster.

Testing for potential diurnal and seasonal biases, we computed the variance portion in plankton community composition, explained by acquisition time or season. We used a redundancy analysis (RDA), with Hellinger-transformed concentrations as response variables, and a binary variable as explanatory, either day/night or productive/non-productive season, defined based on latitudes and sampling months (Table S2) (Lalli & Parsons, 1997). Diurnal effects were tested in both layers; while seasonality was not tested in the mesopelagic layer since seasonal changes are weaker at depth (Costello et al., 2018). In both cases, explained variance was computed as R^2.

2.4 Correspondence with Ocean Regionalisations

In a second step, we aimed to capture the link between plankton communities and their environment. To assess various regionalisations ability to capture processes driving plankton ecology, the part of variance in plankton community composition explained by each was computed. This was performed analogously to the circadian or seasonal effects test. In each layer, a RDA was performed, with Hellinger-transformed concentrations as response variables and a qualitative variable with regions from a given regionalisation as explanatory variable. To ensure adequate representativeness, each region in each regionalisation had to contain at least 25 profiles for inclusion. This limited the number of regions used for each regionalisation, but resulted in similar region numbers throughout all tested regionalisations: all contained 13 - 18 regions for the epipelagic and 10 - 12 regions regarding the mesopelagic zone (Table 1). This is important for comparison as the variance portion explained by a categorical variable often increases with the number of modalities.

To quantify how much variance is explainable by a categorical variable with a similar cardinality, a maximal model was built by computing a regionalisation on plankton concentrations themselves, with a similar groups number. We used the PCA on Hellinger-transformed concentration data, followed by a *k*-means clustering on the first five PCs. Here, the group's number was set as a middle ground between cardinalities of other regionalisations (epipelagic: 15; mesopelagic: 12). Note that in the epipelagic layer maximal model, the difference between 13 and 18 modalities is inconsequential: 13 modalities explain 55.7% of variance, while 18 modalities explain 59.2% (+3.5% of variance). A RDA was performed with this explanatory variable, and, since response and explanatory variables are built with the same data, this RDA captures the maximum part of explainable variance.

Other tested regionalisations included: Longhurst provinces (Longhurst, 2010), 10° latitudinal bands and mesopelagic provinces (Reygondeau et al., 2018) (tested for the mesopelagic layer only). Besides these regionalisations, often based on climatological averages, a regionalisation based on each profile's immediate environment was generated with a PCA performed on environmental data in both layers. This PCA included

layer wise data: CTD, UVP5 bulk and marine snow data; as well as data shared between layers: depth of clines and satellite data. The latter were also included in the mesopelagic PCA as phytoplankton present in the epipelagic layer can influence processes in the mesopelagic layer (Guidi et al., 2009; Hernández-León et al., 2020). Projections on the first 5 PCs were used to generate a *k*-means clustering. For comparison purposes, the number of modalities was set to be similar to the other regionalisations (13 for the epipelagic, 10 for the mesopelagic). Corresponding PCAs and maps can be found in supplementary (Figures S4, S5). Finally, we also compared these regionalisations to a null model: profiles were randomly grouped into a similar number of clusters. If the variance portion explained by a given regionalisation is similar to the null model portion, this regionalisation does not capture plankton community composition variations.

All analyses were conducted with R version 4.0.3 and the 'vegan' package version 2.5.7 (Oksanen et al., 2018).

3 Results

3.1 Spatial Distribution of Plankton Communities

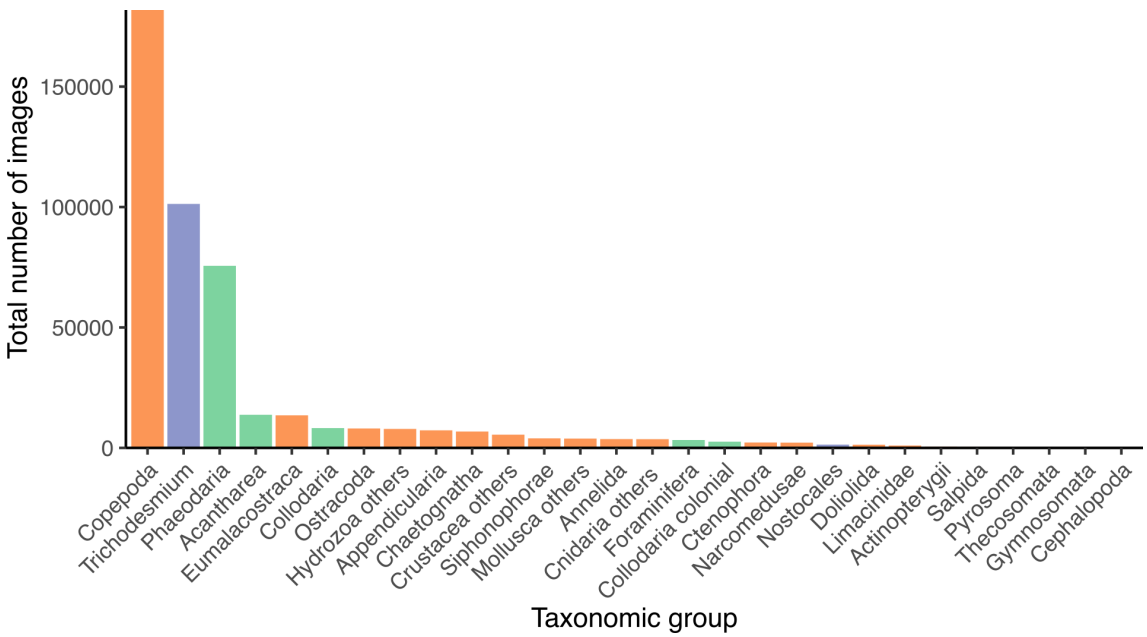

Figure 2: Dataset composition: total number of images per taxonomic group. Rhizaria (unicellular eukaryotes) are highlighted in green, Cyanobacteria are in purple and Metazoa in orange.

Within the UVP5 size range (600 µm to a few cm), more than 330,000 organisms were detected. Three groups dominated the number of individual images: Copepoda (metazoan), *Trichodesmium* (cyanobacteria) and Phaeodaria (Rhizaria subgroup, unicellular eukaryote) (Figure 2). Apart from Rhizaria, *Trichodesmium* and Nostocales (the latter both Cyanobacteria), all other imaged organisms belonged to the Metazoa kingdom.

3.1.1 Epipelagic Layer

To describe epipelagic plankton communities, 2517 profiles were included. The first two PCs captured 41.8% of variance. The first PC distinguished between *Trichodesmium*-rich communities and copepod-rich ones. *Trichosdesmium*-rich communities were associated with warm and stratified waters. The proportion of copepods was higher in cold, high

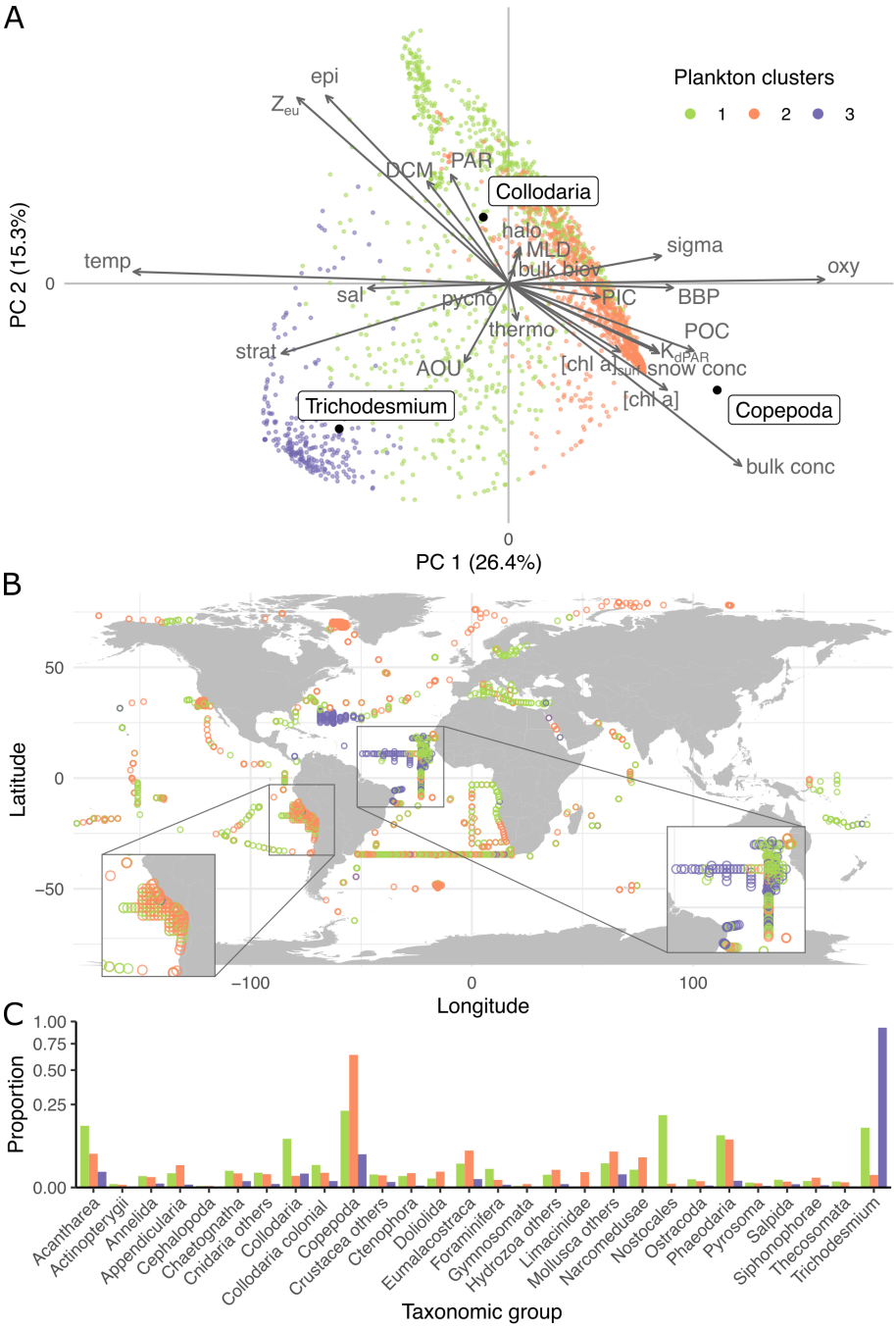

Figure 3: Plankton clusters within the epipelagic layer. (A) PCA performed on Hellinger-transformed plankton concentrations, illustrated by a biplot in scaling 2. Only taxa with a contribution higher than average are represented. Environment variables are projected as supplementary variables. Points represent profiles and are coloured according to the cluster defined by the HAC. (B) Map of epipelagic profiles, coloured as in A. (C) Relative composition of epipelagic plankton clusters. Note that the Y axis is square root transformed.

chlorophyll and particle-rich waters (Figure 3A). On the second PC, a third pole emerged, represented by Collodaria (Rhizaria), extant in oligotrophic waters (with deep DCM and Z_{eu}). The HAC dendrogram was separated into 3 clusters (Figure S3). Cluster 1 was characterised by a co-dominance of multiple Rhizaria groups (Acantharia, Collodaria and Phaeodaria), copepods, nostocales and *Trichodesmium* (Figure 3C), although Collodaria emerged as the third structuring ones in PCA space. This community type was widely

distributed in oceans but detected at lower frequencies in high latitudes (Figure 3B). Cluster 2, enriched in copepods, also had a widespread distribution but dominated the subpolar North Atlantic and Arctic shelf seas, and upwelling areas (California Current, Peruvian and Benguela upwellings). Cluster 3, enriched *Trichodesmium*, exists in the Atlantic Ocean's intertropical band.

3.1.2 Mesopelagic Layer

In the mesopelagic layer, 1747 profiles described plankton communities. The first two PCs captured a 39.6% variance. The first PC separated copepod-rich waters from waters with Phaeodaria (Figure 4A). Copepod-enriched waters were oxygen-rich, while Phaeodaria-rich waters showed high AOU and significant stratification of the water column top. On the second PC, a third Eumalacostraca pole emerged, associated with warmer, saltier, more oligotrophic waters. The HAC dendrogram was again split into 3 clusters (Figure S3). The first cluster, with a higher proportion of phaeodarians and fewer copepods (Figure 4C), was emblematic of the Peruvian upwelling, but also present in Mediterranean Sea and Pacific Ocean profiles. The Copepoda-enriched cluster 2 existed at high northern hemisphere latitudes; also present in the intertropical band and south of the Atlantic (Figure 4B). Finally, cluster 3 was not dominated by any taxon but had a diversified composition: Copepoda, Eumalacostraca, Foraminifera, Phaeodaria and *Trichodesmium*. Fewer profiles were in this cluster but broadly distributed.

3.1.3 Comparison Between Epipelagic and Mesopelagic layers

We compared each station's plankton community type in epipelagic and mesopelagic layers (Figure 5). Some stations, included in the epipelagic analysis, had no corresponding mesopelagic samples as the sea bottom or UVP maximum depth was too shallow. These stations are "NA" in Figure 5 mesopelagic row. Among the 1,122 profiles where the epipelagic part contained a mixed-type community (cluster 1 epipelagic), the mesopelagic layer contained a phaeodarian-enriched community (cluster 1 mesopelagic) in 49% of cases and a copepod-enriched community (cluster 2 mesopelagic) in 33% of cases. In the copepod-rich epipelagic community (cluster 2 epipelagic), most profiles displayed a mesopelagic copepod-enriched community (52%, cluster 2 mesopelagic). Finally, most profiles with a *Trichodesmium*-enriched epipelagic layer (cluster 3 epipelagic) had a phaeodarian-enriched mesopelagic community (75%, cluster 2 mesopelagic). Overall, this analysis highlights only an incomplete similarity between epipelagic and mesopelagic plankton communities' compositions, suggesting they are driven by different processes disjoining the surface and deep communities.

3.2 Representativity of Existing Ocean Regionalisations

In both layers, the clustering built on the plankton data itself (maximal model) explained about half of the variance in community composition (epipelagic: 56.7%; mesopelagic: 46.7%; Table 1). The limited amount of explained variance unsurprisingly confirms plankton communities' diversity cannot be summarised in only 12 to 15 groups, but also highlights that other regionalisations should be gauged relatively to these figures, not to 100% of variance explained. All regionalisations explained more variance than the null model (< 1% of variance explained in both layers). Thus, all regionalisations were relevant for explaining plankton communities' distribution. Among the tested regionalisations, Longhurst provinces explained more variance than others (epipelagic: 26.0%; mesopelagic: 13.2%), corresponding to about half of the explainable variance in the epipelagic layer and 1/3 in the mesopelagic zone. In the epipelagic, latitudinal bands explained similar variance to the local environment; in the mesopelagic, Reygondeau's mesopelagic provinces explained some more variance than the local environment and

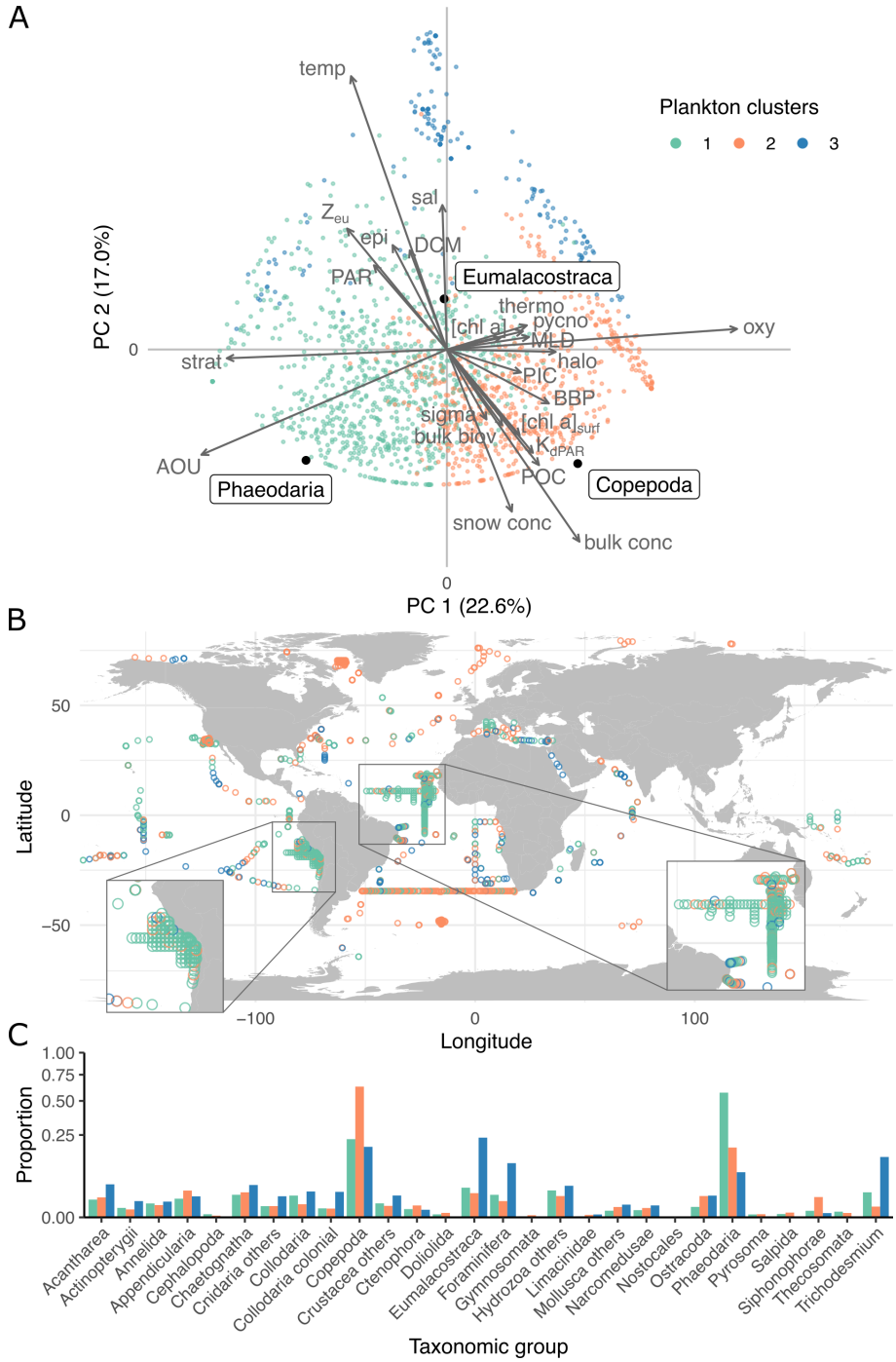

Figure 4: Plankton clusters within the mesopelagic layer. (A) PCA performed on Hellinger-transformed plankton concentrations, illustrated by a biplot in scaling 2. Only taxa with a contribution higher than average are represented. Environment variables are projected as supplementary variables. Points represent profiles and are coloured according to the cluster defined by the HAC. (B) Map of mesopelagic profiles, coloured as in A. (C) Relative composition of mesopelagic plankton clusters. Note that the Y axis is square root transformed.

latitudinal bands. In both layers, plankton community composition was better explained by biogeochemical-based provinces than by the immediate and local environment the plankton was sampled in.

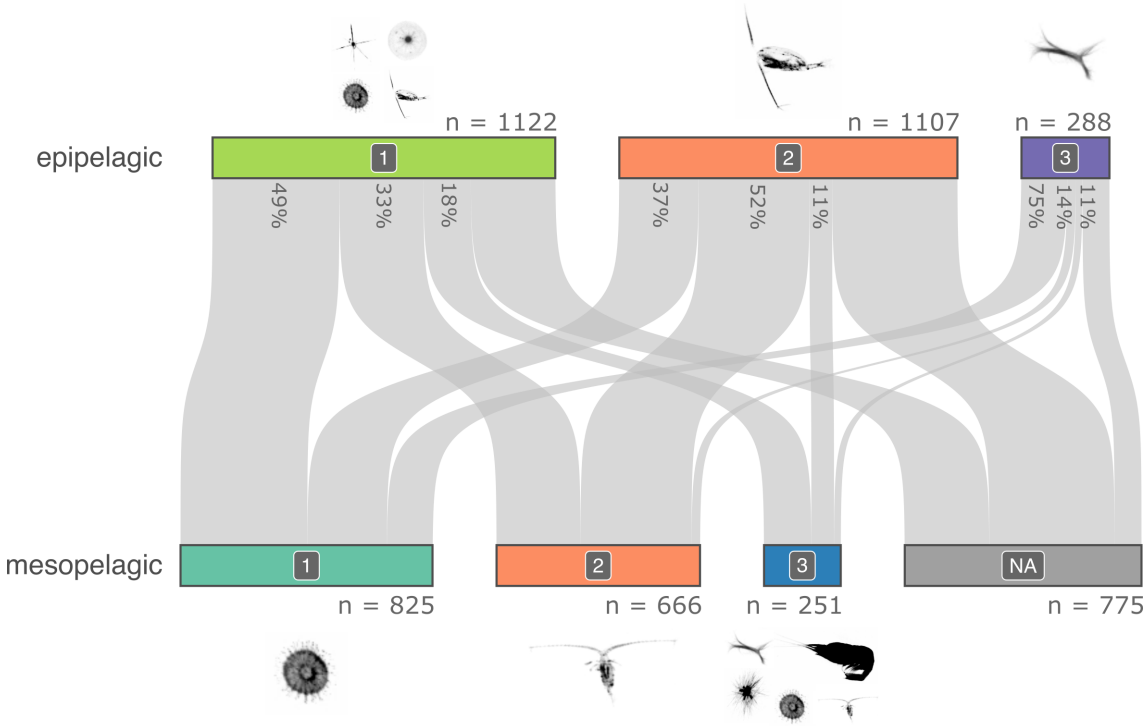

Figure 5: Comparison between epipelagic and mesopelagic plankton communities. Horizontal coloured bars represent profiles, split according to their plankton communities in each layer, as previously described in figures 3 and 4 (NA in the mesopelagic row represents stations with no mesopelagic portion). Grey bands show the correspondence of plankton communities between clusters in the epi and mesopelagic layer for each profile. Percentages show the repartition of epipelagic plankton communities in the mesopelagic layer, excluding profiles absent from the mesopelagic analysis (i.e., going in the NA band).

Table 1: Variance in community composition explained by different regionalisations for the epipelagic and mesopelagic layers. 2,203 and 1,193 profiles were included in the epipelagic and mesopelagic layers respectively. n = number of groups for each regionalisation.

Regionalisation	Epipelagic		Mesopelagic	
	n	R^2	n	R^2
Maximal model	15	0.567	12	0.467
Null model	15	0.007	11	0.007
Longhurst provinces	18	0.260	12	0.132
Latitude bands	13	0.173	11	0.104
Local environment	13	0.168	10	0.092
Mesopelagic provinces	-	-	11	0.116

4 Discussion

In summary, three meso/macro plankton communities' types were detected in both epipelagic and mesopelagic layers. Their composition was better explained by basin scale environmental conditions than by local ones. Below, we first briefly discuss methodological aspects to assess our results' robustness (these are discussed in more detail in the supplementary material) and then discuss our findings' consequences in the existing knowledge context.

4.1 Potential Biases

Plankton concentrations vary temporally, with seasons, but also between day and night because of the diel vertical migrations (DVM) (Lampert, 1989), conducted by many – mostly pluricellular – plankton taxa. Here, both diel and seasonal effects explained very little variance (Table S3), showing that they barely impact community composition and seemed negligible compared to spatial effects, although DVM was detected on raw concentrations (Figure S6).

Although UVP5 profiles were distributed unevenly in space (Figure S1) and in time (Figure S7), our analyses were robust to downsampling, showing that our results were not solely representative of oversampled areas (see procedure description in supplementary material). Moreover, UVP5 samples were shown to cover diverse enough environmental conditions, representative of worldwide oceans (Figures S8, S9 and text in supplementary material).

Finally, the UVP5's detection capabilities are limited, both in terms of size (> 600 µm for detection and > 1 mm for quantitative measurements) and taxonomy (only broad taxa can be identified). Thus, if poorly taxonomically resolved taxa (e.g. Copepoda, Eumalacostraca) could have been better resolved, the resulting lower level taxa might have emerged as structuring mesoplankton communities. Similarly, it cannot be excluded that the detection of smaller organisms could have revealed further structuring taxa (e.g. diatoms at high latitudes, Dutkiewicz et al., 2020). Indeed, smaller plankton are typically more abundant than larger organisms. These elements could also impact our findings regarding how the distribution of plankton communities can be explained by various regionalisations: assuming a strong geographic signature, this could have increased the ability of one given regionalisation in explaining the distribution of plankton communities.

4.2 Plankton Communities General Structure

4.2.1 Epipelagic Layer

Three plankton communities types emerged in the epipelagic layer, mostly driven by water masses' temperature and trophic statuses: copepod-enriched communities in cold and productive waters; *Trichodesmium*-enriched communities in warm waters; and mixed-type communities in oligotrophic ones.

Copepods' dominance in almost all pelagic ecosystems is already documented (Brandão et al., 2021; Ibarbalz et al., 2019; Rombouts et al., 2009; Soviadan et al., 2022), more so in the rich and productive Arctic waters (Brandão et al., 2021; Drago et al., 2022; Forest et al., 2008; Soviadan et al., 2022; Trudnowska et al., 2020). Similarly to other zooplankton, copepods' diversity decreases with latitude (Brandão et al., 2021; Ibarbalz et al., 2019) and conversely for size (Brun et al., 2016; Horne et al., 2016). Brandão et al., 2021 found a 400 - 500 µm median ESD for copepods between 60°N and 60°S. As UVP5 only detects copepods over 1 mm in ESD (Forest et al., 2012), tropical and temperate copepods' concentrations were likely underestimated but in various situations, conclusions on community composition are little affected, since all taxa concentrations are similarly underestimated near the UVP target size range limits. Large changes in the biogeography, community composition and diversity of calanoid copepods in the North Atlantic Ocean are detected, as a result of global warming (Beaugrand et al., 2009; Beaugrand et al., 2010). As copepods act as a trophic link between primary producers and higher trophic levels (Rombouts et al., 2009), these changes might prove detrimental to marine resources, like exploited fish stocks (Beaugrand et al., 2010). We show that, within the 600 µm range to a few cm, copepods are proportionally more abundant in polar waters and, to a lesser extent, in temperate ones, but clearly not in tropical ones.

Epipelagic plankton communities were also shaped by *Trichodesmium*, a filamentous cyanobacteria found in subtropical and tropical regions (Capone et al., 1997; Westberry

& Siegel, 2006) and previously observed with UVP5 in tropical waters (Guidi et al., 2012; Sandel et al., 2015). Given that phytoplanktonic organisms are rather uncommon in the size range observed by the UVP5 (> 600 µm) compared to zooplanktonic organisms, the role of *Trichodesmium* as structuring mesoplankton communities was quite unexpected. *Trichodesmium* contribute to primary production and fixation of atmospheric nitrogen (Capone et al., 1997) and can grow in nitrogen-limited environments, unlike other phytoplankton types (Westberry & Siegel, 2006). Their toxicity to several zooplankton species (Hawser et al., 1992) might contribute to the quasi-exclusion of other types of planktonic organisms in *Trichodesmium*-enriched communities, provided that both *Trichodesmium* and other organisms concentrations are sufficiently high to cause encounters.

Finally, the third cluster revealed a mixed plankton community, characterised by copepods' or *Trichodesmium*'s non-dominance. This cluster contained profiles from diverse environments with varying conditions, thus very diversified composition-wise. Its average composition (Figure 3) is skewed by high concentrations from a few profiles, therefore not representative of every profile' s composition aggregated in this cluster. Although Collodaria emerged as a structuring group in the PCA, it did not dominate the cluster's relative composition. Actually, six groups were found, accounting for 85% of the composition: Acantharia (Rhizaria), Collodaria (Rhizaria), Copepoda, Nostocales, Phaeodaria (Rhizaria) and *Trichodesmium*, highlighting the importance of various rhizarian groups, recently highlighted by other *in situ* imaging-based studies (Biard et al., 2016; Dennett et al., 2002). Acantharia and Collodaria are symbionts-bearing Rhizaria, widely distributed in oceans, but more abundant in tropical oligotrophic surface waters, where their symbionts are photosynthetically active (Suzuki & Not, 2015), hence coherent with our observations. Collodaria contributes significantly to the total organic matter in these environments (Biard et al., 2016; Suzuki & Not, 2015). Conversely, Phaeodaria are heterotrophic Rhizaria, lacking symbionts (Kling & Boltovskoy, 1999), flux-feeders and thus usually found below the epipelagic layer, where they feed on sinking particles (Biard & Ohman, 2020). However, some species are common in surface layers (Nakamura & Suzuki, 2015). Because Phaeodaria's mineral skeleton is made of silica, they can act as major biogenic silica exporters (Biard et al., 2018) and their distribution could be restricted by silica availability (Biard & Ohman, 2020; Nakamura & Suzuki, 2015). Finally, nostocales (*Aphanizomenon*, *Dolichospermum*, and *Nodularia*) were identified contextually (i.e., not just based on aspect on images), only in the Baltic Sea, and at very high concentrations. As a consequence, they account for 19% of the mixed-type community, though they were found in only a few profiles.

4.2.2 Mesopelagic Layer

In the mesopelagic layer, three types of plankton communities also emerged: copepod-enriched in cold and oxygenated waters, Phaeodaria-enriched in cold and oxygen-depleted waters, and, again, a mixed community in warmer waters.

In the latter, four groups accounted for 65% of organisms: Copepoda, Eumalacostraca, Foraminifera, and *Trichodesmium*. Foraminifera, heterotrophic rhizarians, feed on mesopelagic plankton and are typical of deep and rather poorly oxygenated waters (Biard & Ohman, 2020). Various organisms were identified within the broad Eumalacostraca group in our dataset, depending on the ecosystem sampled, resulting in a very heterogeneous group. Thus, Eumalacostraca are not representative of any typical environment, restricting ecological interpretations, and most likely under-sampled.

In cold waters, the plankton community type was linked to oxygen availability: a higher proportion of copepods was found when oxygen was available and Phaeodaria in water masses, with high AOU and low oxygen concentration, typical of Oxygen Minimum Zones (OMZ). Major OMZs are found below eastern-boundary upwelling systems, with particularly high primary production. Many planktonic taxa' concentrations, such as calanoid copepods', is reduced in OMZs (Auel & Verheye, 2007; Hoving et al., 2020;

Kiko et al., 2020; Soviadan et al., 2022). Associated with the low oxygen and low plankton concentrations, and possibly caused by it, the downward flux of particulate carbon is intense and less attenuated than elsewhere, resulting in high vertical export of the organic matter photosynthesis produces at the surface (Cavan et al., 2017; Engel et al., 2022). Conversely, Phaeodaria are typical of deep, low oxygenated waters (Biard & Ohman, 2020; Ikenoue et al., 2019) and already detected in OMZs (Hoving et al., 2020). Indeed, protists might prove more tolerant to hypoxia, as their passive feeding mode requires less oxygen than active feeding. In OMZs, they may therefore play a disproportionate role in the regulation of the vertical flux compared to elsewhere.

Although unexpected at first glance, the detection of *Trichodesmium* in the mesopelagic layer is consistent with previous observations (Benavides et al., 2022; Sellner, 1992; Walsby, 1978). Here, their presence in the mesopelagic layer could be partly explained by our dynamic definition of the epi-mesopelagic boundary: it started at shallower than 50 m for 25% of profiles, a depth where the presence of *Trichodesmium* would not be surprising. However, those were found mostly at higher latitudes (Figure S2), where *Trichodesmium* is absent. Its presence at a great depth could also result from downwelling and subduction events, bringing colonies to deeper waters (Guidi et al., 2012), or even simply represent dead colonies sinking down.

4.2.3 Comparison between Epipelagic and Mesopelagic Plankton Communities

Conditions in the epipelagic layer constrain those in the mesopelagic: i.e., the epipelagic phytoplankton type influences particle sizes in the mesopelagic (Guidi et al., 2009). Similarly, mesopelagic zooplankton biomass is conditioned by the net primary production in the euphotic layer, since it feeds on its remnants (Hernández-León et al., 2020). However, the results above show a low similarity of plankton communities between epipelagic and mesopelagic layers. Such a result is to be linked to the functional definition of the epi-mesopelagic boundary that was used, based on hydrological characteristics, so that different water masses and their potential inherent plankton community are not mixed together.

Where the plankton community is enriched in copepods and the environment productive in the epipelagic layer, the mesopelagic plankton community is usually copepod-rich as well. This is consistent with high secondary production in the mesopelagic, below subpolar surface waters hosting high primary production, which is not true at subtropical latitudes (Robinson et al., 2010). Below oligotrophic Rhizaria-rich epipelagic communities, the mesopelagic community was mostly Phaeodaria-enriched in the eastern tropical South Pacific OMZs but was copepod-enriched in the South Atlantic. The split between these two mesopelagic communities therefore seems to be driven by copepods' oxygen limitation in OMZs (Engel et al., 2022). In the South Atlantic gyre, both epipelagic (Signorini et al., 2015) and mesopelagic (Sutton et al., 2017) layers are considered as oligotrophic, consistent with a Rhizaria-enriched community in the epipelagic (Suzuki & Not, 2015). In the Peruvian upwelling system, the epipelagic community was mostly Rhizaria-enriched; surprising, since this environment is very productive (Ayón et al., 2008). Still, the presence of Rhizaria was already reported there by Santander Bueno, 1981, within a diverse zooplankton community over the region. This productive upwelling area drives an OMZ in deeper waters (Engel et al., 2022; Kiko & Hauss, 2019), which imposes a Phaeodaria-enriched community in the mesopelagic. Within *Trichodesmium*-rich stations, only equatorial Atlantic stations were sampled deep enough to be included in the epi/mesopelagic layers analysis. As *Trichodesmium* was almost absent from the mesopelagic layer there, profiles that were *Trichodesmium*-enriched in the epipelagic had to be distributed between the other two in the mesopelagic plankton communities, and most seen as Phaeodaria-rich.

The forcing environmental conditions associated with communities in both layers are not the same: oxygen plays a role at depth but less near the surface, while light structures

life near the surface but not in the mesopelagic layer. Besides, the conditions that remain structuring ones are less variable with increasing depth, leading to a more homogeneous habitat with depth (Costello & Breyer, 2017; Costello et al., 2018). This also shapes the plankton community, which becomes less spatially contrasted in the mesopelagic (Soviadan et al., 2022).

4.3 Plankton Communities Distribution was Driven by Regional Conditions

Among the regionalisations tested, the plankton communities distribution was better explained by Longhurst provinces (Longhurst, 2010), a regionalisation based on physical forcings as drivers of phytoplankton distribution, which might drive zooplankton communities too. Although plankton diversity and biomass follows a latitudinal gradient (Ikeda, 1985; Rombouts et al., 2009), mediated by joint effects of light availability and temperature, the regionalisation based on 10° latitudinal bands explained less variance than Longhurst provinces. In the mesopelagic, the regionalisation computed by (Reygondeau et al., 2018) – specifically for this layer and mainly based on annual climatologies of biogeochemical variables – did not explain plankton community distribution better than Longhurst provinces, even though it was supposed to be more appropriate. Yet, more importantly, all these basin-scale regionalisations explained plankton community distribution as well, or better, than a regionalisation based on local conditions, at sampling time. This suggests that plankton communities' spatial structure for our 28 taxonomic groups is driven by regional environmental conditions more than by very local and immediate conditions and processes. These findings agree with those of (Stemmann et al., 2008), who showed that mesopelagic macro-zooplankton communities were structured by large, basin-scale processes. Genomic analyses also underlined regional scales processes' importance in structuring plankton communities (Richter et al., 2020), especially for meso-zooplankton (Sommeria-Klein et al., 2021).

Among the many other existing regionalisations of the oceans, the one proposed by Hofmann Elizondo et al., 2021 and based on phytoplankton biogeography would have constituted an interesting point of comparison to assess how it correlates with mesoplankton communities. However, such comparison was not possible because of the low cardinality of this regionalisation. Still, Hofmann Elizondo et al., 2021's tropical biome matches well our *Trichodesmium*-enriched community in the Atlantic ocean, while the high latitude biome coincides well with communities rich in copepods.

5 Conclusion and Perspectives

In both layers, three plankton communities types emerged in the observed size range (> 600 µm) and seemed mostly driven by basin-level environmental conditions. Following on studies investigating plankton distribution and diversity across life kingdoms – from viruses to metazoans (de Vargas et al., 2015; Ibarbalz et al., 2019; Sunagawa et al., 2015) – this work highlights the role not only of metazoans, but also of unexpected large protists and cyanobacteria in structuring meso and megaplankton communities. This confirms underwater imaging relevance to reveal the importance of otherwise overlooked plankton groups, such as Rhizaria (Biard et al., 2016; Dennett et al., 2002).

The development of regionalisations – either based on biotic or abiotic data – for wide-ranging offshore areas is highly desired for conservation purposes, like the creation of protected marine areas. However, they should not be restricted to oceans' upper layers. Indeed, biological activity in the mesopelagic layer is key to mediate the flux of organic carbon from the surface and the deep seafloor is also heavily impacted by human activities (bottom fishing, seafloor mining...) (Watling et al., 2013). Therefore, oceans' deeper layers bio-geographies, currently rare, are also required to balance between human exploitation and ecological conservation.

References

Auel, H., & Verheye, H. M. (2007). Hypoxia tolerance in the copepod Calanoides carinatus and the effect of an intermediate oxygen minimum layer on copepod vertical distribution in the northern Benguela Current upwelling system and the Angola–Benguela Front. *Journal of Experimental Marine Biology and Ecology*, *352*(1), 234–243.

Ayón, P., Criales-Hernandez, M. I., Schwamborn, R., & Hirche, H.-J. (2008). Zooplankton research off Peru: A review. *Progress in Oceanography*, *79*(2), 238–255.

Beaugrand, G., Luczak, C., & Edwards, M. (2009). Rapid biogeographical plankton shifts in the North Atlantic Ocean. *Global Change Biology*, *15*(7), 1790–1803.

Beaugrand, G., Edwards, M., & Legendre, L. (2010). Marine biodiversity, ecosystem functioning, and carbon cycles. *Proceedings of the National Academy of Sciences*, *107*(22), 10120–10124.

Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A., & Edwards, M. (2002). Reorganization of North Atlantic Marine Copepod Biodiversity and Climate. *Science*, *296*(5573), 1692–1694.

Beaugrand, G., Rombouts, I., & Kirby, R. R. (2013). Towards an understanding of the pattern of biodiversity in the oceans. *Global Ecology and Biogeography*, *22*(4), 440–449.

Behrenfeld, M. J., & Boss, E. S. (2014). Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Annual Review of Marine Science*, *6*(1), 167–194.

Benavides, M., Bonnet, S., Le Moigne, F. A. C., Armin, G., Inomura, K., Hallstrøm, S., Riemann, L., Berman-Frank, I., Poletti, E., Garel, M., Grosso, O., Leblanc, K., Guigue, C., Tedetti, M., & Dupouy, C. (2022). Sinking Trichodesmium fixes nitrogen in the dark ocean. *The ISME Journal*, 1–8.

Benfield, M., Grosjean, P., Culverhouse, P., Irigolen, X., Sieracki, M., Lopez-Urrutia, A., Dam, H., Hu, Q., Davis, C., Hanson, A., Pilskaln, C., Riseman, E., Schulz, H., Utgoff, P., & Gorsky, G. (2007). RAPID: Research on Automated Plankton Identification. *Oceanography*, *20*(2), 172–187.

Biard, T., Krause, J. W., Stukel, M. R., & Ohman, M. D. (2018). The Significance of Giant Phaeodarians (Rhizaria) to Biogenic Silica Export in the California Current Ecosystem. *Global Biogeochemical Cycles*, *32*(6), 987–1004.

Biard, T., & Ohman, M. D. (2020). Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone revealed by in situ imaging. *Limnology and Oceanography*, *65*(11), 2583–2602.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., & Not, F. (2016). In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*, *532*(7600), 504–507.

Boyer, T. P., Garcia, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., Weathers, K. A., Baranova, O. K., Paver, C. R., Seidov, D., & Smolyar, I. V. (2018). World Ocean Atlas 2018. NOAA National Centers for Environmental Information. Dataset.

Brandão, M. C., Benedetti, F., Martini, S., Soviadan, Y. D., Irisson, J.-O., Romagnan, J.-B., Elineau, A., Desnos, C., Jalabert, L., Freire, A. S., Picheral, M., Guidi, L., Gorsky, G., Bowler, C., Karp-Boss, L., Henry, N., de Vargas, C., Sullivan, M. B., Stemmann, L., & Lombard, F. (2021). Macroscale patterns of oceanic zooplankton composition and size structure. *Scientific Reports*, *11*(1), 15714.

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, *41*(1), 8–22.

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography*. Sinauer Associates, Inc.

Brun, P., Payne, M. R., & Kiørboe, T. (2016). Trait biogeography of marine copepods – an analysis across scales. *Ecology Letters*, *19*(12), 1403–1413.

Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., & Carpenter, E. J. (1997). Trichodesmium, a Globally Significant Marine Cyanobacterium. *Science*, *276*(5316), 1221–1229.

Caron, D. A., Countway, P. D., Jones, A. C., Kim, D. Y., & Schnetzer, A. (2012). Marine protistan diversity. *Annual review of marine science*, *4*(1), 467–493.

Cavan, E. L., Trimmer, M., Shelley, F., & Sanders, R. (2017). Remineralization of particulate organic carbon in an ocean oxygen minimum zone. *Nature Communications*, *8*(1), 14847.

Costello, M. J., & Breyer, S. (2017). Ocean Depths: The Mesopelagic and Implications for Global Warming. *Current Biology*, *27*(1), R36–R38.

Costello, M. J., Tsai, P., Wong, P. S., Cheung, A. K. L., Basher, Z., & Chaudhary, C. (2017). Marine biogeographic realms and species endemicity. *Nature Communications*, *8*(1), 1057.

Costello, M. J., Basher, Z., Sayre, R., Breyer, S., & Wright, D. J. (2018). Stratifying ocean sampling globally and with depth to account for environmental variability. *Scientific Reports*, *8*(1), 1–9.

Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., & González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine indentification of dinoflagellates. *Marine Ecology Progress Series*, *247*, 17–25.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*, C12003.

Dennett, M. R., Caron, D. A., Michaels, A. F., Gallager, S. M., & Davis, C. S. (2002). Video plankton recorder reveals high abundances of colonial Radiolaria in surface waters of the central North Pacific. *Journal of Plankton Research*, *24*(8), 797–805.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N. L., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., … Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605.

Drago, L., Panaïotis, T., Irisson, J.-O., Babin, M., Biard, T., Carlotti, F., Coppola, L., Guidi, L., Hauss, H., Karp-Boss, L., Lombard, F., McDonnell, A. M. P., Picheral, M., Rogge, A., Waite, A. M., Stemmann, L., & Kiko, R. (2022). Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning. *Frontiers in Marine Science*, *9*.

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

Emerson, S., Quay, P., Karl, D., Winn, C., Tupas, L., & Landry, M. (1997). Experimental determination of the organic carbon flux from open-ocean surface waters. *Nature*, *389*(6654), 951–954.

Engel, A., Kiko, R., & Dengler, M. (2022). Organic Matter Supply and Utilization in Oxygen Minimum Zones. *Annual Review of Marine Science*, *14*(1), 355–378.

Falkowski, P. (2012). Ocean Science: The power of plankton. *Nature*, *483*(7387), S17–S20.

Forest, A., Stemmann, L., Picheral, M., Burdorf, L., Robert, D., Fortier, L., & Babin, M. (2012). Size distribution of particles and zooplankton across the shelf-basin system in southeast Beaufort Sea: Combined results from an Underwater Vision Profiler and vertical net tows. *Biogeosciences*, *9*(4), 1301–1320.

Forest, A., Sampei, M., Makabe, R., Sasaki, H., Barber, D. G., Gratton, Y., Wassmann, P., & Fortier, L. (2008). The annual cycle of particulate organic carbon export in Franklin Bay (Canadian Arctic): Environmental control and food web implications. *Journal of Geophysical Research: Oceans*, *113*(C3).

Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L., & Brown, J. H. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences*, *105*(22), 7774–7778.

Guidi, L., Calil, P. H. R., Duhamel, S., Björkman, K. M., Doney, S. C., Jackson, G. A., Li, B., Church, M. J., Tozzi, S., Kolber, Z. S., Richards, K. J., Fong, A. A., Letelier, R. M., Gorsky, G., Stemmann, L., & Karl, D. M. (2012). Does eddy-eddy interaction control surface phytoplankton distribution and carbon export in the North Pacific Subtropical Gyre? *Journal of Geophysical Research: Biogeosciences*, *117*(G2).

Guidi, L., Legendre, L., Reygondeau, G., Uitz, J., Stemmann, L., & Henson, S. A. (2015). A new look at ocean carbon remineralization for estimating deepwater sequestration. *Global Biogeochemical Cycles*, *29*(7), 1044–1059.

Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., Picheral, M., & Gorskya, G. (2009). Effects of phytoplankton community on production, size, and export of large aggregates: A world-ocean analysis. *Limnology and Oceanography*, *54*(6), 1951–1963.

Hawser, S. P., O'Neil, J. M., Roman, M. R., & Codd, G. A. (1992). Toxicity of blooms of the cyanobacterium Trichodesmium to zooplankton. *Journal of Applied Phycology*, *4*(1), 79–86.

Hays, G. C., Richardson, A. J., & Robinson, C. (2005). Climate change and marine plankton. *Trends in Ecology & Evolution*, *20*(6), 337–344.

Hernández-León, S., Koppelmann, R., Fraile-Nuez, E., Bode, A., Mompeán, C., Irigoien, X., Olivar, M. P., Echevarría, F., Fernández de Puelles, M. L., González-Gordillo, J. I., Cózar, A., Acuña, J. L., Agustí, S., & Duarte, C. M. (2020). Large deep-sea zooplankton biomass mirrors primary production in the global ocean. *Nature Communications*, *11*(1), 6048.

Hofmann Elizondo, U., Righetti, D., Benedetti, F., & Vogt, M. (2021). Biome partitioning of the global ocean based on phytoplankton biogeography. *Progress in Oceanography*, *194*, 102530.

Horne, C. R., Hirst, A. G., Atkinson, D., Neves, A., & Kiørboe, T. (2016). A global synthesis of seasonal temperature–size responses in copepods. *Global Ecology and Biogeography*, *25*(8), 988–999.

Hörstmann, C., Buttigieg, P. L., John, U., Raes, E. J., Wolf-Gladrow, D., Bracher, A., & Waite, A. M. (2022). Microbial diversity through an oceanographic lens: Refining the concept of ocean provinces through trophic-level analysis and productivity-specific length scales. *Environmental Microbiology*, *24*(1), 404–419.

Hoving, H. J. T., Neitzel, P., Hauss, H., Christiansen, S., Kiko, R., Robison, B. H., Silva, P., & Körtzinger, A. (2020). In situ observations show vertical community structure of pelagic fauna in the eastern tropical North Atlantic off Cape Verde. *Scientific Reports*, *10*(1), 21798.

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

Ikeda, T. (1985). Metabolic rates of epipelagic marine zooplankton as a function of body mass and temperature. *Marine Biology*, *85*(1), 1–11.

Ikenoue, T., Kimoto, K., Okazaki, Y., Sato, M., Honda, M. C., Takahashi, K., Harada, N., & Fujiki, T. (2019). Phaeodaria: An Important Carrier of Particulate Organic Carbon in the Mesopelagic Twilight Zone of the North Pacific Ocean. *Global Biogeochemical Cycles*, *33*(8), 1146–1160.

Irisson, J.-O., Ayata, S.-D., Lindsay, D. J., Karp-Boss, L., & Stemmann, L. (2022). Machine Learning for the Study of Plankton and Marine Snow from Images. *Annual Review of Marine Science*, *14*(1), 277–301.

Kiko, R., Brandt, P., Christiansen, S., Faustmann, J., Kriest, I., Rodrigues, E., Schütte, F., & Hauss, H. (2020). Zooplankton-Mediated Fluxes in the Eastern Tropical North Atlantic. *Frontiers in Marine Science*, *7*.

Kiko, R., & Hauss, H. (2019). On the Estimation of Zooplankton-Mediated Active Fluxes in Oxygen Minimum Zone Regions. *Frontiers in Marine Science*, *6*.

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., Boss, E., Brandt, P., Carlotti, F., Christiansen, S., Coppola, L., de la Cruz, L., Diamond-Riquier, E., Durrieu de Madron, X., Elineau, A., Gorsky, G., Guidi, L., Hauss, H., Irisson, J.-O., … Stemmann, L. (2022). A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5. *Earth System Science Data Discussions*, 1–37.

Kling, S. A., & Boltovskoy, D. (1999). Radiolaria Phaeodaria. *South Atlantic Zooplankton*, *1*, 231–264.

Lalli, C. M., & Parsons, T. R. (1997). Phytoplankton and Primary Production. In *Biological Oceanography: An Introduction* (pp. 39–73). Elsevier.

Lampert, W. (1989). The Adaptive Significance of Diel Vertical Migration of Zooplankton. *Functional Ecology*, *3*(1), 21–27.

Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Li, W. K. W., Head, E. J. H., & Glen Harrison, W. (2004). Macroecological limits of heterotrophic bacterial abundance in the ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(11), 1529–1540.

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J.-B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O., Miloslavich, P., Lara-Lpez, A., Kudela, R., Lopes, R. M., Kiko, R., … Appeltans, W. (2019). Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Frontiers in Marine Science*, *6*.

Longhurst, A. R. (1995). Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, *36*(2), 77–167.

Longhurst, A. R. (2010). *Ecological geography of the sea*. Academic Press.

Longhurst, A. R., & Glen Harrison, W. (1989). The biological pump: Profiles of plankton production and consumption in the upper ocean. *Progress in Oceanography*, *22*(1), 47–123.

McGillicuddy, D. J., Anderson, L. A., Bates, N. R., Bibby, T., Buesseler, K. O., Carlson, C. A., Davis, C. S., Ewart, C., Falkowski, P. G., Goldthwait, S. A., Hansell, D. A., Jenkins, W. J., Johnson, R., Kosnyrev, V. K., Ledwell, J. R., Li, Q. P., Siegel, D. A., & Steinberg, D. K. (2007). Eddy/Wind Interactions Stimulate Extraordinary Mid-Ocean Plankton Blooms. *Science*, *316*(5827), 1021–1026.

Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, *106*(C4), 7163–7180.

Nakamura, Y., & Suzuki, N. (2015). Phaeodaria: Diverse Marine Cercozoans of World-Wide Distribution. In S. Ohtsuka, T. Suzaki, T. Horiguchi, N. Suzuki & F. Not (Eds.), *Marine Protists: Diversity and Dynamics* (pp. 223–249). Springer Japan.

Oksanen, J., Kindt, R., Pierre, L., O'Hara, B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Wagner, H., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2018). *Vegan: Community Ecology Package*.

Picheral, M., Colin, S., & Irisson, J.-O. (2017). *EcoTaxa, a tool for the taxonomic classification of images*. Retrieved November 13, 2020, from https://ecotaxa.obs-vlfr.fr/

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolu-

tion studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(9), 462–473.

Remsen, A., Hopkins, T. L., & Samson, S. (2004). What you see is not what you catch: A comparison of concurrently collected net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder data from the northeast Gulf of Mexico. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(1), 129–151.

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

Rice, J., Gjerde, K. M., Ardron, J., Arico, S., Cresswell, I., Escobar, E., Grant, S., & Vierros, M. (2011). Policy relevance of biogeographic classification for conservation and management of marine biodiversity beyond national jurisdiction, and the GOODS biogeographic classification. *Ocean & Coastal Management*, *54*(2), 110–122.

Richardson, A. J. (2008). In hot water: Zooplankton and climate change. *ICES Journal of Marine Science*, *65*(3), 279–295.

Richter, D. J., Watteaux, R., Vannier, T., Leconte, J., Frémont, P., Reygondeau, G., Maillet, N., Henry, N., Benoit, G., Silva, O. D., Delmont, T. O., Fernàndez-Guerra, A., Suweis, S., Narci, R., Berney, C., Eveillard, D., Gavory, F., Guidi, L., Labadie, K., … Jaillon, O. (2020). Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *bioRxiv*, 867739.

Robinson, C., Steinberg, D. K., Anderson, T. R., Arístegui, J., Carlson, C. A., Frost, J. R., Ghiglione, J.-F., Hernández-León, S., Jackson, G. A., Koppelmann, R., Quéguiner, B., Ragueneau, O., Rassoulzadegan, F., Robison, B. H., Tamburini, C., Tanaka, T., Wishner, K. F., & Zhang, J. (2010). Mesopelagic zone ecology and biogeochemistry – a synthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(16), 1504–1518.

Rombouts, I., Beaugrand, G., Ibaňez, F., Gasparini, S., Chiba, S., & Legendre, L. (2009). Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1670), 3053–3062.

Rutherford, S., D'Hondt, S., & Prell, W. (1999). Environmental controls on the geographic distribution of zooplankton diversity. *Nature*, *400*(6746), 749–753.

Sandel, V., Kiko, R., Brandt, P., Dengler, M., Stemmann, L., Vandromme, P., Sommer, U., & Hauss, H. (2015). Nitrogen Fuelling of the Pelagic Food Web of the Tropical Atlantic. *PLOS ONE*, *10*(6), e0131258.

Santander Bueno, H. (1981). The Zooplankton in an Upwelling Area off Peru. In *Coastal Upwelling* (pp. 411–416). American Geophysical Union (AGU).

Sellner, K. G. (1992). Trophodynamics of Marine Cyanobacteria Blooms. In *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs* (pp. 75–94). Springer Netherlands.

Sherman, K. (2005). The large marine ecosystem approach for assessment and management of ocean coastal waters. In T. Hennessey & J. Sutinen (Eds.), *Sustaining Large Marine Ecosystems: The Human Dimension* (pp. 3–16). Elsevier.

Signorini, S. R., Franz, B. A., & McClain, C. R. (2015). Chlorophyll variability in the oligotrophic gyres: Mechanisms, seasonality and trends. *Frontiers in Marine Science*, *2*.

Sommeria-Klein, G., Watteaux, R., Ibarbalz, F. M., Pierella Karlusich, J. J., Iudicone, D., Bowler, C., & Morlon, H. (2021). Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science*, *374*(6567), 594–599.

Soviadan, Y. D., Benedetti, F., Brandão, M. C., Ayata, S.-D., Irisson, J.-O., Jamet, J. L., Kiko, R., Lombard, F., Gnandi, K., & Stemmann, L. (2022). Patterns of mesozooplankton community composition and vertical fluxes in the global ocean. *Progress in Oceanography*, *200*, 102717.

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 & Coastal Management*, *60*, 19–30.

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.

Stemmann, L., Youngbluth, M. J., Robert, K., Hosia, A., Picheral, M., Paterson, H., Ibaňez, F., Guidi, L., Lombard, F., & Gorsky, G. (2008). Global zoogeography of fragile macrozooplankton in the upper 100-1000 m inferred from the underwater video profiler. *ICES Journal of Marine Science*, *65*(3), 433–442.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., D'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science (New York, N.Y.)*, *348*(6237), 1261359.

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., B\ lachowiak-Samo\ lyk, K., Robison, B. H., Gjerde, K. M., … Heino, M. (2017). A global biogeographic classification of the mesopelagic zone. *Deep Sea Research Part I: Oceanographic Research Papers*, *126*, 85–102.

Suzuki, N., & Not, F. (2015). Biology and Ecology of Radiolaria. In S. Ohtsuka, T. Suzaki, T. Horiguchi, N. Suzuki & F. Not (Eds.), *Marine Protists: Diversity and Dynamics* (pp. 179–222). Springer Japan.

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.

Trudnowska, E., Stemmann, L., Błachowiak-Samołyk, K., & Kwasniewski, S. (2020). Taxonomic and size structures of zooplankton communities in the fjords along the Atlantic water passage to the Arctic. *Journal of Marine Systems*, *204*, 103306.

Vilgrain, L., Maps, F., Picheral, M., Babin, M., Aubry, C., Irisson, J.-O., & Ayata, S.-D. (2021). Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. *Limnology and Oceanography*, *66*(4), 1155–1167.

Villarino, E., Watson, J. R., Jönsson, B., Gasol, J. M., Salazar, G., Acinas, S. G., Estrada, M., Massana, R., Logares, R., Giner, C. R., Pernice, M. C., Olivar, M. P., Citores, L., Corell, J., Rodríguez-Ezpeleta, N., Acuña, J. L., Molina-Ramírez, A., González-Gordillo, J. I., Cózar, A., … Chust, G. (2018). Large-scale ocean connectivity and planktonic body size. *Nature Communications*, *9*(1), 142.

Walsby, A. (1978). The properties and buoyancy-providing role of gas vacuoles in Trichodesmium Ehrenberg. *British Phycological Journal*, *13*(2), 103–116.

Ware, D. M., & Thomson, R. E. (2005). Bottom-Up Ecosystem Trophic Dynamics Determine Fish Production in the Northeast Pacific. *Science*, *308*(5726), 1280–1284.

Watling, L., Guinotte, J., Clark, M. R., & Smith, C. R. (2013). A proposed biogeography of the deep ocean floor. *Progress in Oceanography*, *111*, 91–112.

Westberry, T. K., & Siegel, D. A. (2006). Spatial and temporal distribution of Trichodesmium blooms in the world's oceans. *Global Biogeochemical Cycles*, *20*(4).

Woodd-Walker, R. S., Ward, P., & Clarke, A. (2002). Large-scale patterns in diversity and community structure of surface water copepods from the Atlantic Ocean. *Marine Ecology Progress Series*, *236*, 189–203.

Data Accessibility Statement

The code supporting this study is available at: https://github.com/ThelmaPana/planktyp. The dataset is made available on SEANOE: https://doi.org/10.17882/95809.

Acknowledgements

We are grateful for the ship time provided by the respective institutions and programs and we would like to thank all scientists, officers and crew of the various R/V which took part in UVP5 data collection. We are thankful to Léo Lacour for providing satellite data and to Laetitia Drago for taxonomy checks. GlobColour data (http://globcolour.info) used in this study has been developed, validated, and distributed by ACRI-ST, France. We acknowledge the MOOSE program (Mediterranean Ocean Observing System for the Environment) coordinated by CNRS-INSU and the Research Infrastructure ILICO (CNRS-IFREMER). TP's doctoral fellowship was granted by the French Ministry of Higher Education, Research and Innovation (3500/2019). JOI acknowledges support by the Belmont Forum grant #ANR-18-BELM-0003-01. RK and LS received support from the European Union project TRIATLAS (European Union Horizon 2020 program, grant agreement 817578). RK furthermore acknowledges support via a Make Our Planet Great Again grant from the French National Research Agency (ANR) within the Programme d'Investissements d'Avenir #ANR-19-MPGA-0012 and funding from the Heisenberg Programme of the German Science Foundation #KI 1387/5-1. AR was funded by the PACES II (Polar Regions and Coasts in a Changing Earth System) program of the Helmholtz Association and the INSPIRES program of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. LS was supported by the CNRS/Sorbonne University Chair VISION to initiate the global observation. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement #862923.

Author Contributions

TP, JOI and LS conceived the study; MB, TB, FC, LC, LG, HH, LKB, RK, FL, AMD, MP, AR, AW, LS contributed to data acquisition and image validation; TP conducted the analyses under the supervision of JOI and LS; TP wrote the manuscript; all authors contributed to the discussion of the results, supported manuscript preparation and approved the final submitted manuscript.

Supplementary Material

Here we present the extra analyses showing that our results are robust to various potential biases.

Seasonal and Diel Cycles Effects Many – mostly pluricellular – plankton taxa conduct diel vertical migrations (DVM) (Lampert, 1989). Yet, our analysis showed no significant effect of this migration on plankton community composition (i.e., relative concentrations) (Table S3), in line with previous findings (Soviadan et al., 2022). To further investigate the circadian effect separately from the seasonal effect, we conducted an additional test on pairs of profiles performed both during day and night at the same location (geographic distance < 2 km, time distance < 24h). For the 172 pairs of such profiles existing in our dataset, we compared raw concentrations in the epipelagic layer at day and night for the four most abundant taxa (*Trichodesmium*, Copepoda, Phaeodaria and Acantharea), using a paired a Wilcoxon-Mann-Whitney test after removing double zeros. This revealed statistical differences for all four tested taxa (Figure S6).

Both phytoplankton and zooplankton concentrations also vary seasonally: spring (and possibly autumn) blooms cause an increase in productivity and plankton concentration (Behrenfeld & Boss, 2014). But plankton may also bloom outside seasonal blooms, due to favourable conditions following water mass displacements (McGillicuddy et al., 2007). These sudden events, restricted spatially and temporally, are called intermittent blooms. For example, *Trichodesmium* can bloom locally in tropical and subtropical oceans (Westberry & Siegel, 2006). Colonies formed during these events can be detected by the UVP5. However, although seasonality affects absolute concentrations, our results suggested a negligible effect of season on community composition.

Briefly, both diel and seasonal effects were detected on absolute concentrations. Their non-significance in our analysis was therefore due to Hellinger's transformation, focusing our analyses on relative rather than absolute concentrations (Legendre & Legendre, 2012). With such focus on community composition and at the broad taxonomic level studied, the large-scale geographical effect dominated over seasonal and diel cycles.

Sampling Effort Heterogeneity UVP5 profiles were distributed unevenly: some areas were sampled intensively (California Current, Peruvian upwelling, Mediterranean Sea), others were rarely visited (Indian Ocean, Southern Ocean, Figure S1). Moreover, sampling was heterogeneous in time too: high latitudes were not visited during winter months (Figure S7).

To make sure that our results are not solely representative of oversampled areas, we conducted our analyses on a subsample of our data. Focusing on the epipelagic layer, variograms computed on the concentrations of Copepoda, *Trichodesmium* and Collodaria showed a scale of autocorrelation around 1000 km. Thus, a maximum of 20 profiles were selected in squares of 10° by 10° (~1000 km × 1000 km), for a total of 1388 selected profiles. These 1388 profiles were used to perform a subset PCA and build a factorial space in which all 2517 profiles from the epipelagic layer were projected. Projections on PC1 and PC2 of all 2517 profiles were extracted and compared to projections obtained from the PCA performed with all profiles. This resulted in good correlations (PC1: $R^2 = 0.97$, $p < 0.001$; PC2: $R^2 = 0.90$, $p < 0.001$), showing that our analyses are robust to downsampling.

Furthermore, our analysis does not explicitly consider location or date, only each sample's community and environmental conditions; so the relevant question is: does UVP sampling cover environmental conditions representative of global scale variance? For this, we compared conditions distribution at UVP samples' locations to the same variables distribution at global scale. Of course, simultaneous worldwide *in situ* observations are not available. Instead, we used annual climatologies on a 1° grid from World

Ocean Atlas (WOA) (Boyer et al., 2018) for important water characteristics: temperature, salinity, and oxygen. We first checked that those climatologies were representative of the *in situ* conditions at locations sampled by UVP5, over the epipelagic and mesopelagic layers previously defined; this was the case since correlations were good (all $R^2 > 0.84$, except for AOU in the epipelagic: $R^2 = 0.35$, Figure S8). Then, we compared each variable distributions from the WOA data at UVP5 profiles' locations vs. worldwide (Figure S9), for two depth layers (0 - 200 m; 200 - 500 m), since the above dynamic boundary could not be computed from WOA data. Distributions were similar, showing UVP samples covered diverse enough environmental conditions, representative of worldwide oceans.

Figures

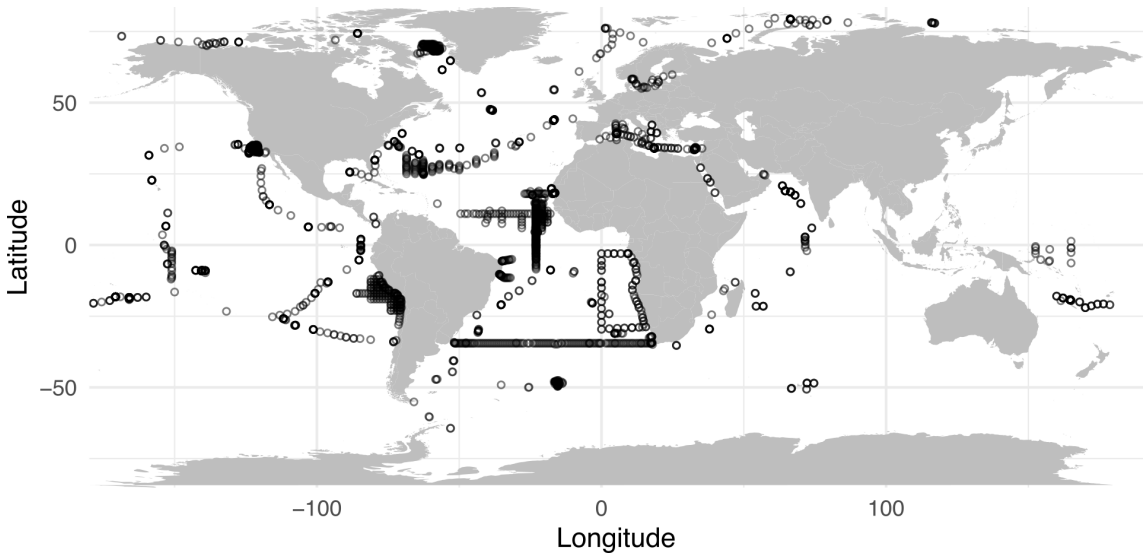

Figure S1: World map of included stations (whether in the epipelagic or mesopelagic layer).

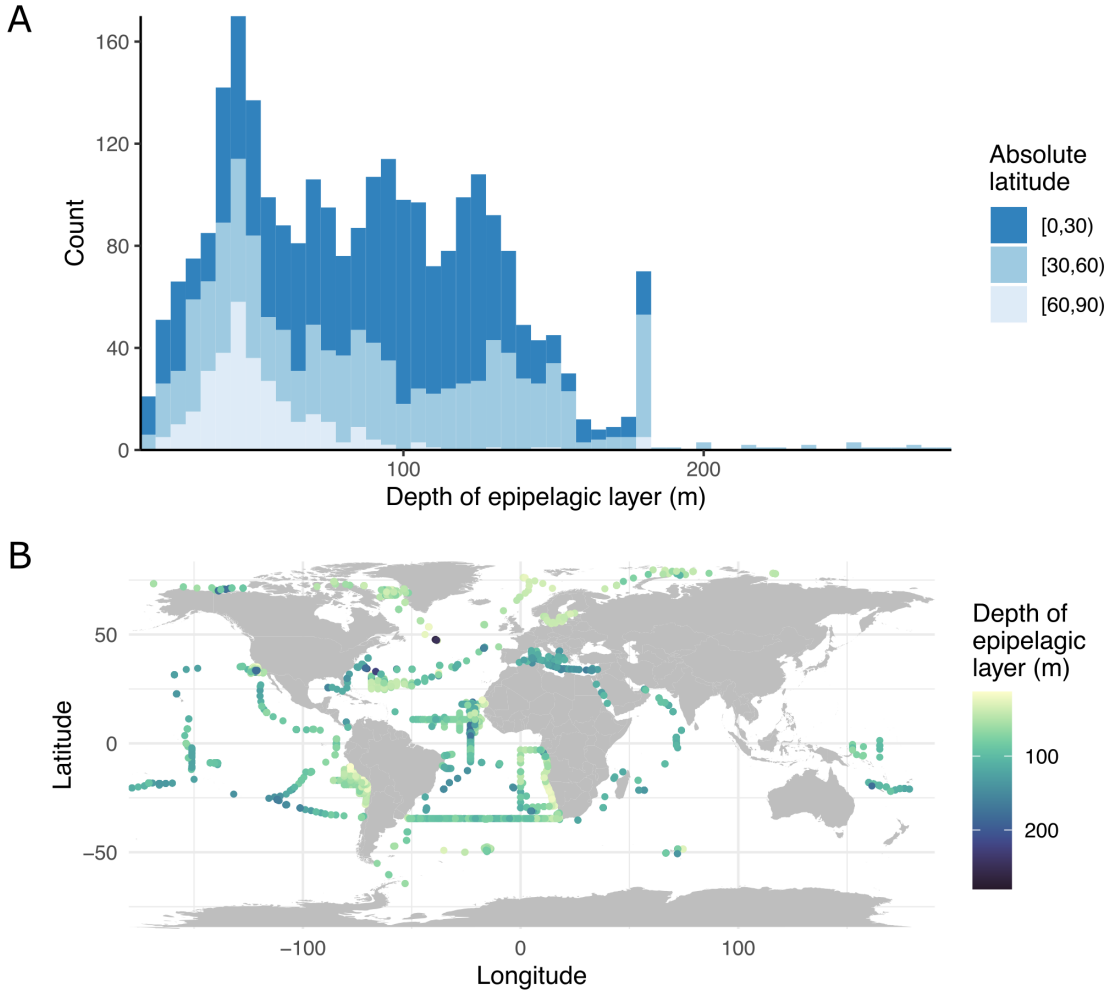

Figure S2: Depth of the dynamic epi-mesopelagic boundary, computed as the deepest value among the mixed layer depth and the euphotic depth. (A) Histogram of the epipelagic layer depth per 30° of absolute latitude bands. The peak at 180 m highlights cases of euphotic depth at 180 m and shallower mixed layer depth. (B) World map of the epipelagic layer depth.

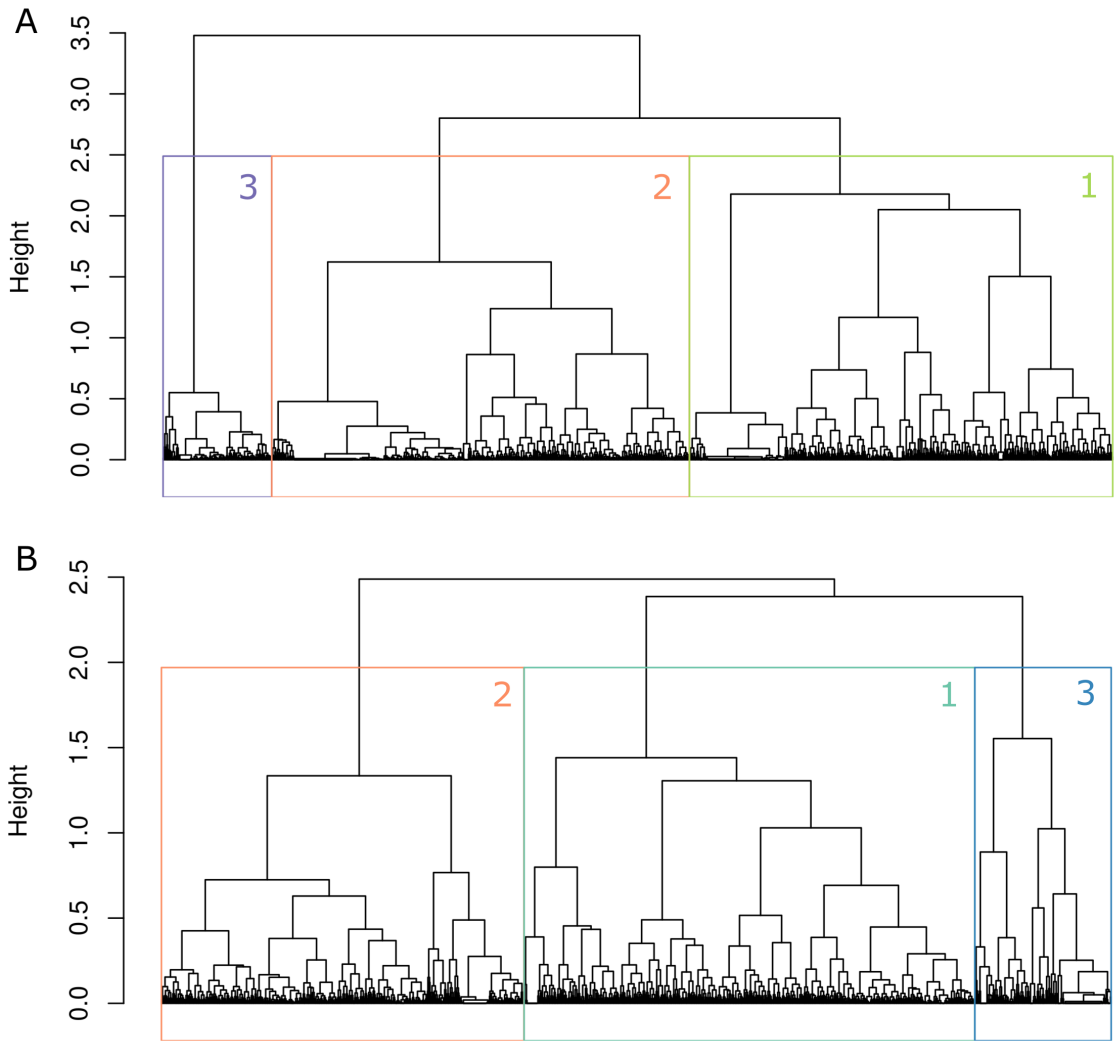

Figure S3: HAC dendrograms based on the first five principal components of profiles projection in the Hellinger-transformed plankton PCA data, for (A) epipelagic and (B) mesopelagic layers. Generated clusters are shown in the same colours and numbers as they appear on figures 3, 4 and 5.

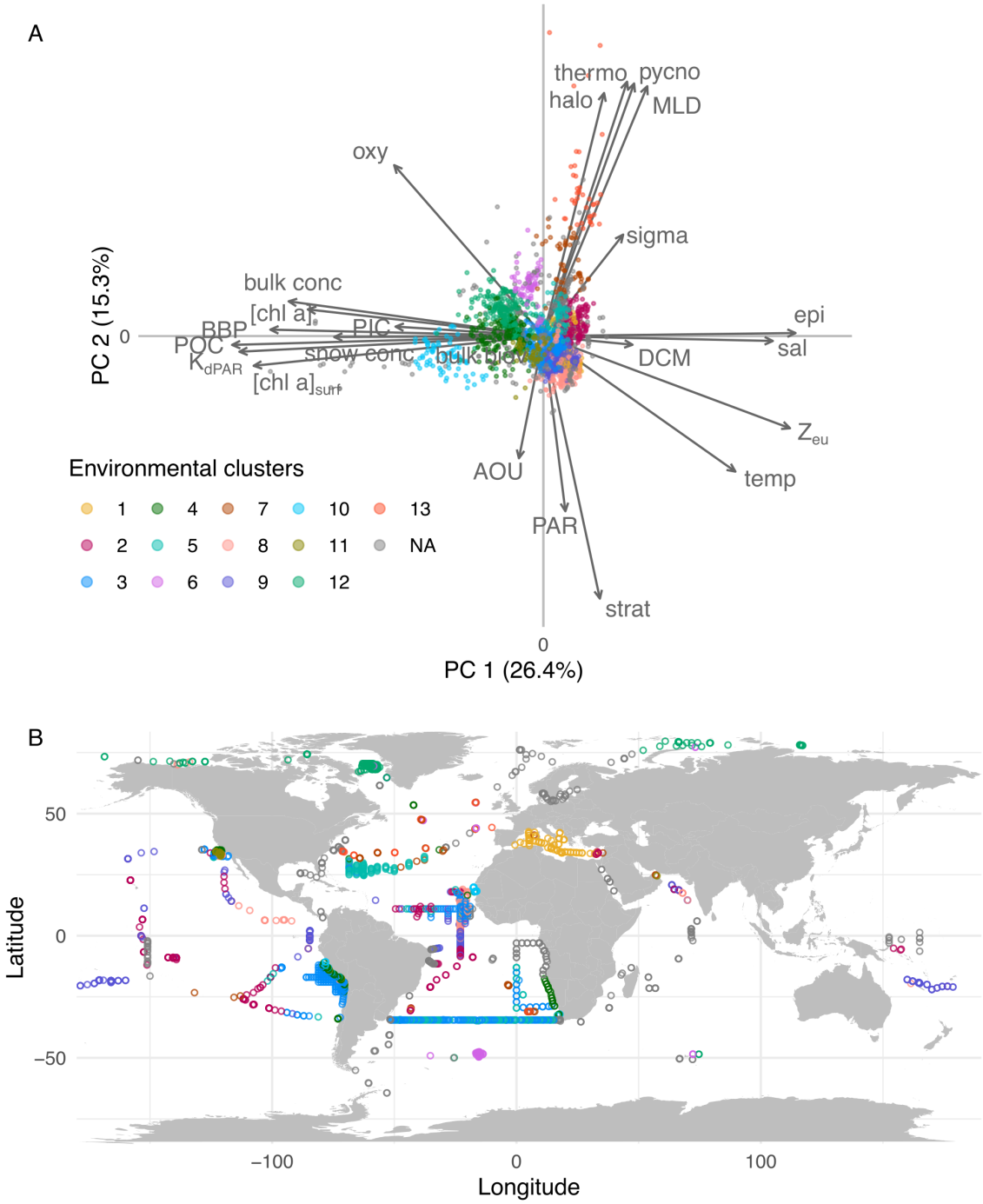

Figure S4: Local environment clustering for the epipelagic layer. (A) PCA performed on environmental variables, illustrated by a biplot in scaling 2. Points represent profiles and are coloured according to the cluster defined by the *k*-means algorithm. NA represents profiles that could not be associated wuth a cluster with more than 25 profiles. (B) Map of epipelagic profiles, coloured as in A.

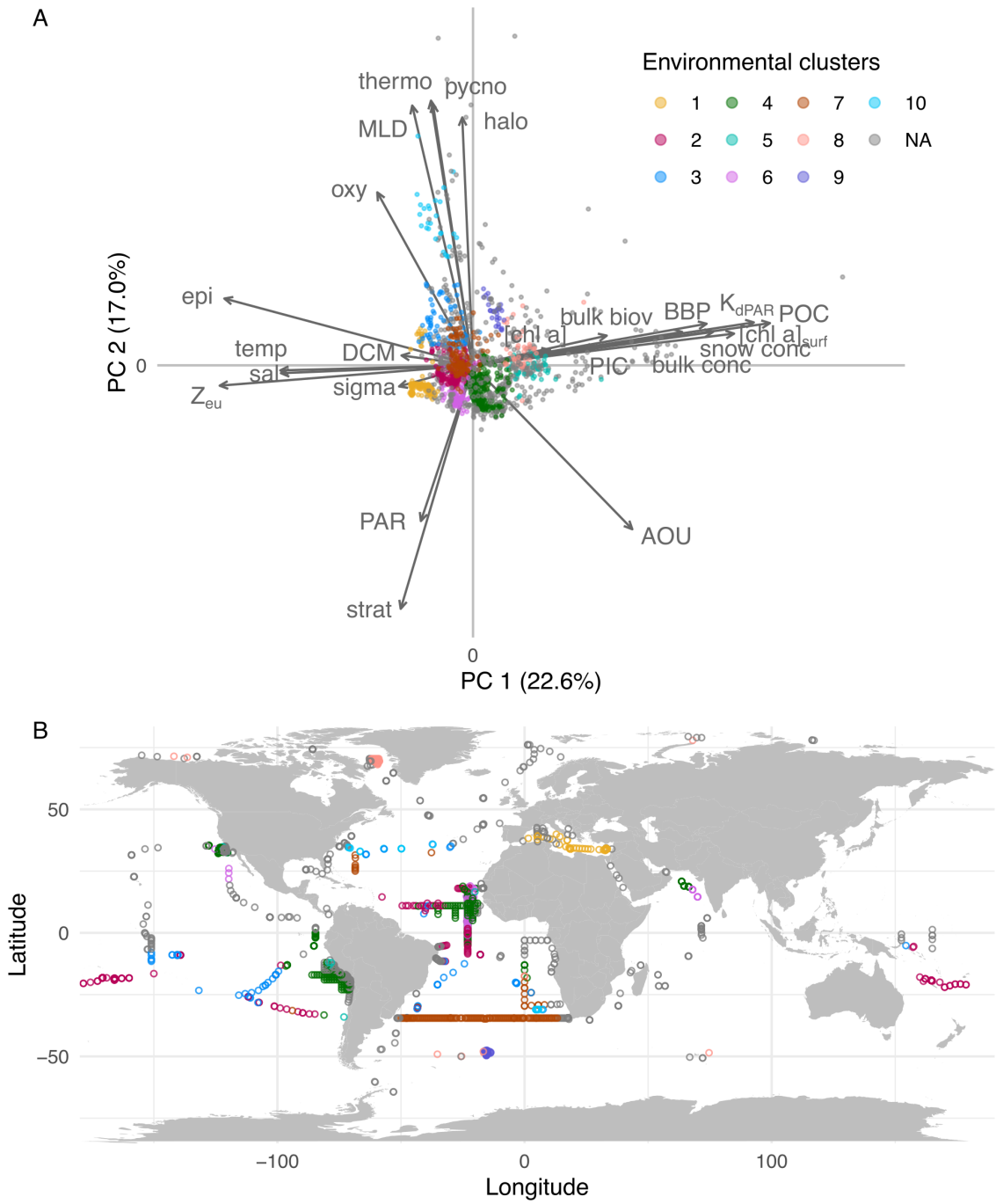

Figure S5: Local environment clustering for the mesopelagic layer. (A) PCA performed on environmental variables, illustrated by a biplot in scaling 2. Points represent profiles and are coloured according to the cluster defined by the *k*-means algorithm. NA represents profiles that could not be associated with a cluster with more than 25 profiles. (B) Map of mesopelagic profiles, coloured as in A.

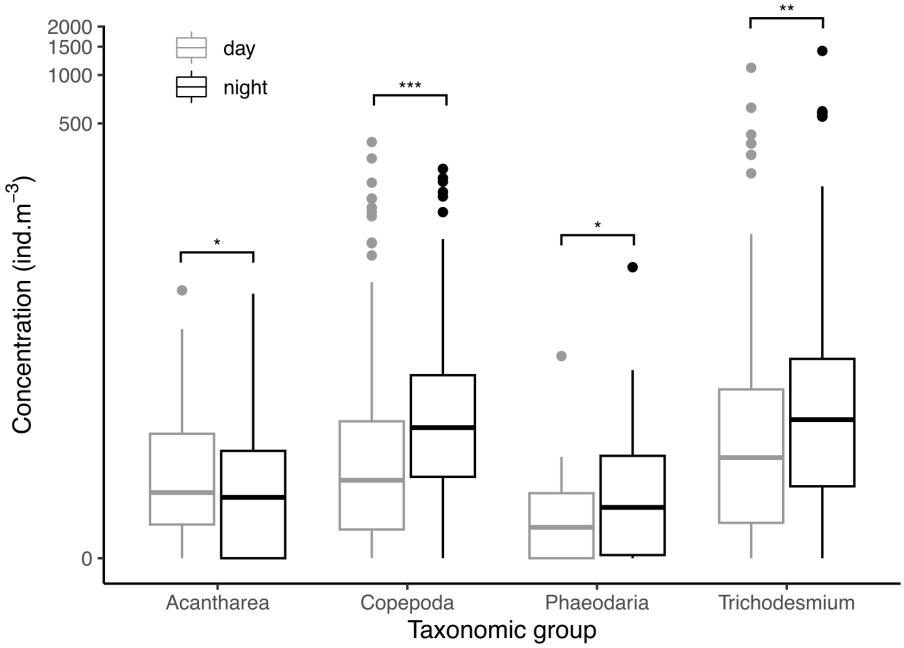

Figure S6: Average epipelagic concentration of the four most abundant taxa in 172 pairs of day/night stations. Stations were paired according to both geographical distance (< 2 km) and time (< 24 h). Note that the Y axis is log-transformed. Differences were tested with a paired Wilcoxon-Mann-Whitney test. * = 0.05, ** = 0.01, *** = 0.001.

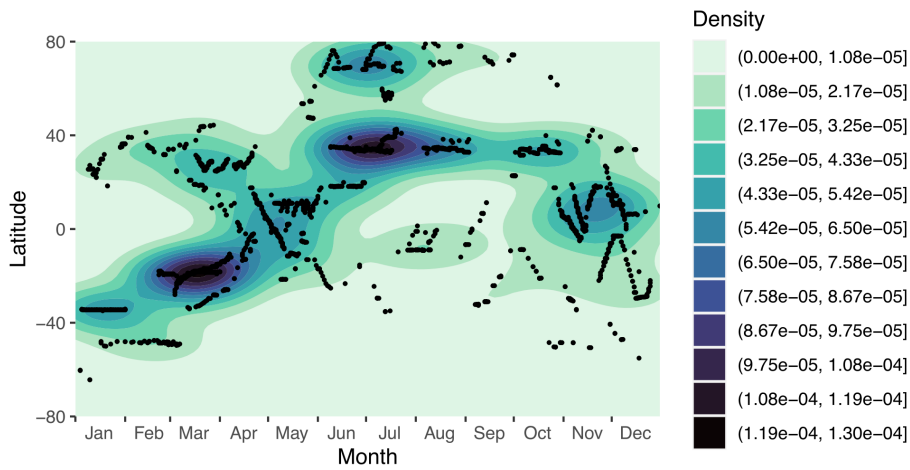

Figure S7: Time versus latitude Hovmöller diagram of sampled stations.

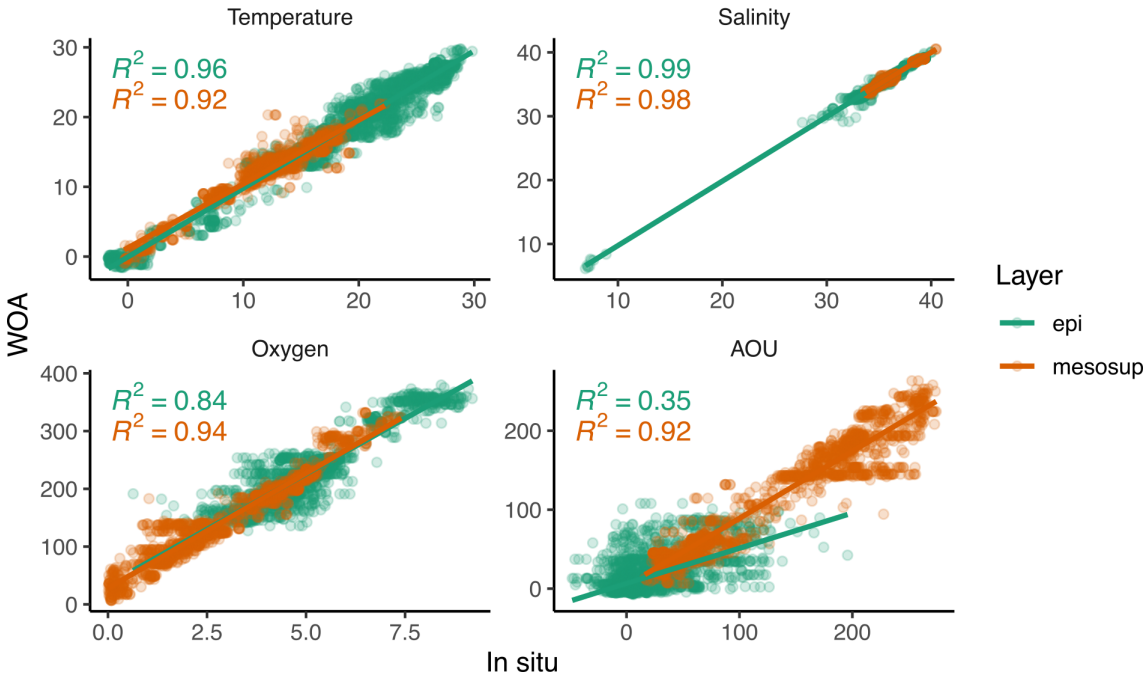

Figure S8: Correlation between *in situ* and annual WOA data at UVP5 profiles locations in the epipelagic and mesopelagic layers.

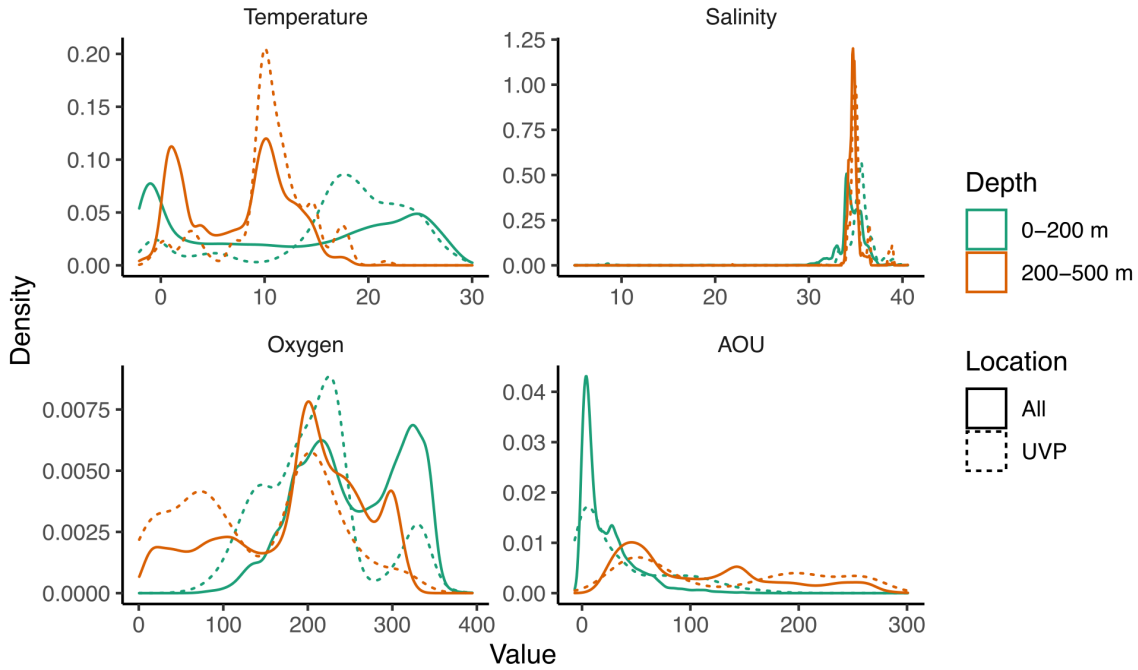

Figure S9: Distribution of annual WOA data all over the globe and at UVP5 profiles locations.

Tables

Table S1: List of oceanographic campaigns included in the study.

Campaign	Year	Nb profiles	UVP5
BOUM	2008	177	sd
CASSIOPEE	2015	13	sd
CCELTER 2008	2008	73	sd
CCELTER 2011	2011	56	zd
CCELTER 2012	2012	59	sd
CCELTER 2014	2014	60	sd
CCELTER 2017	2017	68	hd
DEWEX	2013	1	sd
MSM22	2012	101	sd
MSM23	2012	64	sd
M105	2014	8	sd
M106	2014	114	sd
M107	2014	71	sd
PS88b	2014	36	sd
M116	2015	74	sd
M121	2015	84	sd
M135	2017	138	sd
GreenEdge 2016	2016	121	hd
MSM060	2017	126	hd
IPS Amundsen 2018	2018	6	sd
JERICO 2017	2017	24	sd
KEOPS	2011	13	zd
LOHAFEX	2009	55	sd
MALINA	2009	16	sd
MooseGE[1]	2015	3	sd
NAAMES02	2016	21	hd
OUTPACE	2015	193	sd
P16N	2015	14	sd
Sargasso	2014	84	sd
SOMBA	2014	6	sd
Tara Oceans	2009-2013	643	sd

[1]https://doi.org/10.18142/235

Table S2: Definition of productive (1) and non-productive (0) seasons based on latitude and month.

Latitude band	Month											
	J	F	M	A	M	J	J	A	S	O	N	D
90°N - 66.5°N	0	0	0	0	0	1	1	1	0	0	0	0
66.5°N - 23.5°N	0	0	1	1	1	0	0	0	1	1	0	0
23.5°N - 23.5°S	0	0	0	0	0	0	0	0	0	0	0	0
23.5°S - 66.5°S	0	0	1	1	0	0	0	0	1	1	1	0
66.5°S - 90°S	1	1	0	0	0	0	0	0	0	0	0	1

This model is based on light intensity and nutrients availability. In polar regions, light availability is often limited (namely in winter) but becomes sufficient after the summer ice breakup, allowing productivity. In mid-latitudes, both light and nutrients become available in spring and autumn, generating phytoplankton blooms. In tropical regions, productivity is limited all year by nutrients and remains low.

Table S3: Variance in plankton community composition explained by diel and seasonal cycles computed from RDA. n = number of profiles included in each modality (day / night profiles for diel cycle; non-productive / productive season for seasonal cycle.) All p-value < 0.001.

Regionalisation	Epipelagic		Mesopelagic	
	n	R^2	n	R^2
Diel	1595 / 922	1.1%	1088 / 659	0.9%
Seasonal	1925 / 592	1.3%	-	-