Title

Copepod functional traits and groups show contrasting biogeographies in the global ocean

Running title

Global marine copepod traits biogeography

Authors

Fabio Benedetti[1*] (0000-0002-7554-3646)

Jonas Wydler[1]

Meike Vogt[1] (0000-0002-0608-1935)

[1]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092 Zürich, Switzerland.

*Corresponding author: fabio.benedetti@usys.ethz.ch

Acknowledgements

We thank all contributors involved in the plankton species field sampling and identification throughout the world and we acknowledge the efforts made to deposit the data on publicly available online archives. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 862923. This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein. We thank Luke Gregor for editing the language of an early version of the manuscript.

Conflict of interest

The authors declare no conflict of interests.

Abstract

Aim: To define global zooplankton functional groups (FGs) and to estimate their environmental niche and habitat distribution. We model the spatial patterns of copepod FGs habitat and identify regions sharing similar functional trait expression at the community level.

Taxon: Marine planktonic Neocopepoda.

Location: Global ocean.

Methods: Factor analysis on mixed data and hierarchical clustering were used to identify copepod FGs based on five species-level functional traits. An ensemble of species distribution models was used to estimate the environmental niches of the modelled species, project the mean annual habitat suitability of the FGs, and to estimate the community weighted mean values of the traits studied. Ocean regions were defined based on their community-level mean trait expression using a principal component analysis and hierarchical clustering.

Results: Eleven global copepod FGs were identified. They displayed contrasting latitudinal patterns in mean annual habitat suitability that could be explained by differences in environmental niche preferences: two FGs were associated with polar conditions, one followed the global temperature gradient, five were associated with tropical oligotrophic gyres, and the remaining three with boundary currents and counter currents. Four main regions of varying community weighted mean trait values emerged: the Southern Ocean, the northern and southern high latitudes, the tropical gyres, and the boundary currents and upwelling systems.

Conclusions: We build on an exhaustive species trait dataset to put forward novel FGs that will improve the representation of zooplankton in global marine ecosystem models. Our results contribute to our understanding of the spatial patterns and drivers of marine plankton trait biogeography and will serve as a basis for studying the links between zooplankton biodiversity and ecosystem functioning and how they might evolve in the context of climate change.

Keywords

Planktonic copepods, Functional groups, Trait-based approach, Species Distribution Modelling, Global Ocean, Community weighted mean trait

1. Introduction

Copepods are crustaceans that dominate the biomass of the mesozooplankton size class (0.2-2.0 mm) and rank amongst the most abundant animals in the oceans (Turner, 2004; Kiørboe, 2011a). They are morphologically and functionally diverse and are adapted to almost all marine ecosystems (Kiørboe, 2011a; Bron et al., 2011). Copepods play a pivotal role in the food web, both as microplankton grazers and prey for small pelagic fishes and other larger animals (Beaugrand, Edwards, & Legendre, 2010; Steinberg & Landry, 2017). Copepods play a critical role in the marine biological carbon pump, the fixation of inorganic carbon to organic matter by photosynthesis and consequent sequestration away from the surface into the deep ocean. Planktonic copepods export carbon from the euphotic layer to the deep ocean in several ways, such as the grazing of phytoplankton cells in the sunlit layers, followed by the excretion of relatively fast-sinking fecal pellets, or by extensive vertical migration (Turner, 2015; Steinberg & Landry, 2017).

The relative contribution of planktonic copepods to the above-mentioned processes is mediated by their diversity and trait expression (Barton et al., 2013). For instance, larger copepods perform stronger diel vertical migration (Ohman & Romagnan, 2016) and produce larger fecal pellets that sink faster and thus increase the proportion of particulate organic carbon sinking to depth (Stamieszkin et al., 2015; Brun et al., 2019). The dominance of certain copepod feeding modes (i.e., the mechanism through which prey is captured) impacts food-web dynamics as feeding modes affect copepod grazing and mortality rates (Kenitz, Visser, Mariani, & Andersen, 2017; van Someren Gréve, Almeda, & Kiørboe, 2017). Yet, mechanistic ecosystem models usually represent zooplankton only through a couple of size classes (Le Quéré et al., 2005), which greatly oversimplifies the contribution of species diversity and functional traits to the functioning of ecosystems and biogeochemical cycles (Flynn et al., 2015). Multiple modelling studies show that the number of zooplankton functional groups and their grazing characteristics exerts an important control on the biomass and diversity of other trophic groups, with consequences for global biogeochemical cycles in these models (Prowe et al., 2012; Sailley et al., 2013; Vallina et al., 2014; Le Quéré et al., 2016). Ecosystem models, however, rely strongly on parameterisations of specific traits governing biological processes and food-web interactions (e.g. grazing rates or food preferences) based on scarce information, as the empirical evidence included in current models is often sourced from limited laboratory or field data from a narrow range of culturable species (Barton et al., 2013). Consequently, investigating the potential links

between copepod trait distribution and ecosystem functioning contributes to improving current ecosystem models (Stocker, 2014).

To better understand the role of planktonic copepods in marine systems, plankton ecologists are increasingly adopting trait-based approaches (Litchman, Ohman, & Kiørboe, 2013; Hébert, Beisner, & Maranger, 2017). Functional traits are species- or organism-level characteristics (morphological, behavioral, physiological, or related to life history) that affect their fitness and contribute to ecosystem functioning (Violle et al., 2007). According to Litchman et al. (2013), functional traits can be classified by the main functions they contribute to, namely, survival, feeding, growth, and reproduction. Organisms often cannot maximize all four ecological functions at the same time since there are trade-offs between these functions. For example, copepods that rely on passive feeding (i.e., copepods that remain immobile and only move to capture prey) show 8.5 times lower mortality rates than copepods that rely on active feeding (van Someren Gréve, Almeda, & Kiørboe, 2017). Yet, adult male copepods often need to move actively to find a mate for reproduction, which undermines the benefits of their feeding mode (Kiørboe, 2011b; Litchman et al., 2013).

Copepod functional traits can be investigated by grouping species into functional groups (FGs) according to similarities in functional trait combinations (Benedetti, Gasparini, & Ayata, 2016). Categorizing species according to their similarity in functional traits rather than their taxonomic classification allows one to summarize highly diverse groups, and/or communities, into more parsimonious categories defined by distinct ecological functions (Barnett, Finlay, & Beisner, 2007; Benedetti, Vogt, Righetti, Guilhaumon, & Ayata, 2018). At the regional scale, Pomerleau et al. (2015) showed that 42 zooplankton species from the North East Pacific Ocean could be clustered into five groups based on their body length, feeding and reproduction mode, and trophic group. Based on a similar set of traits, Benedetti, Vogt, et al. (2018) grouped 106 marine copepod species that dominate Mediterranean communities into seven FGs. They found that carnivorous species were associated with tropical oligotrophic conditions, whereas current-feeding herbivores were associated with more productive and seasonally varying conditions. Such groupings can improve the representation of zooplankton in an ecosystem model since FGs increase the representation of ecological function without adding taxonomic diversity and complexity. However, studies based on a regional pool of species might underestimate the real range of functions ensured by zooplankton in the global ocean, warranting the need for a study based on a global species pool.

The functional composition of observed plankton communities can be investigated through community weighted mean (CWM) values of traits (Ricotta, 2005; Pomerleau, Sastri, & Beisner, 2015; Brun et al., 2016). This is done by calculating the proportion of species in a community that exhibits a particular functional trait value by accounting for their relative contribution to community composition or abundance. Thereby, CWM values of functional traits help to identify the relationships between the emerging expression of functional traits on a community-level, the environmental conditions associated to the spatial patterns of trait expression (i.e., their biogeographical patterns), and the potential trade-offs underlying trait expression. Like for FGs, empirical CWM trait values of marine zooplankton have been explored on local or regional levels for the northern hemisphere mostly (Pomerleau et al., 2015; Pecuchet et al., 2018). They evidenced a negative impact of warming and seasonality on community-level body length and offspring size. The first truly global study by Brun et al. (2016) also found larger body sizes, higher proportions of myelinated species, and relatively smaller offspring size in latitudes >50°, because of gradients in temperature and seasonality in phytoplankton production. To our knowledge, other qualitative traits, such as trophic groups or spawning mode, have not been investigated through the CWM trait approach (McGinty et al., 2018; Benedetti, Vogt et al., 2018). Therefore, their global spatial distribution remains poorly documented and a more complete understanding of trade-offs between the expression of these diverse traits is still needed to better characterize and map ecosystem functions worldwide. If geo-referenced trait information, such as patterns of CWM values, are clustered in space and time, they can be used to define ocean regions sharing common ecological characteristics (Longhurst, 2010; Reygondeau et al., 2017; Hofmann Elizondo et al., 2021). Such regionalization schemes have never been applied on plankton communities' functional trait expressions on a global scale, although this could be more informative of large-scale marine food web and ecosystem dynamics than previous regionalization schemes based on environmental parameters alone.

In this study, we address the following questions: (i) What are the global zooplankton FGs emerging from the/a clustering of the most frequently sampled copepod species according to their functional traits (Benedetti et al., 2016)? (ii) Which parameters structure FGs most strongly in functional trait space and environmental niche space (Benedetti, Vogt, et al., 2018)? (iii) What are the habitat distribution patterns of these FGs and their community-level trait expression (estimated through CWM trait values), and which regions of the global ocean share similar trait expression? To do so, a new functional trait synthesis was carried out for a global pool of >380 frequently observed copepod species, and five traits (body size, trophic

group, feeding mode, myelination, and spawning mode) were used to identify novel FGs based on species' traits combinations. Occurrence-based species distribution models (SDMs) were used to investigate the differences in the environmental niches and monthly habitat distribution modelled for the new copepod FGs. Mean annual CWM trait values were derived from the monthly SDMs-based habitat distribution maps, and hierarchical clustering was used to identify regions that share similar community-level trait expression.

2. Materials and Methods

2.1. Species occurrence data

We used the zooplankton occurrences dataset (geo-localised and dated presences) of Benedetti et al. (2021), which has been gathered to model global zooplankton species biogeography, diversity and community composition. The dataset combines 165'716 occurrences binned into the 1° x 1° grid of the World Ocean Atlas (WOA; Boyer et al., 2013) for 385 copepod species. The occurrences were initially retrieved from the Ocean Biodiversity Information System (OBIS; https://www.obis.org), the Global Biodiversity Information Facility (GBIF; https://www.gbif.org), as well as complementary datasets (see Benedetti et al., 2021 for an exhaustive description and for the curation of species names). The occurrences corresponding to benthic and parasitic taxa and those that were not fit to model the habitat of copepod species in the surface open ocean were discarded as follows: occurrences with missing spatial coordinates, sampling dates, sampling depth, or taxonomic identification at species-level were removed. Furthermore, occurrences from drilling holes, freshwater (sea surface salinity < 20 according to the WOA, grid cells within 25 km of the nearest shoreline, or with a sampling depth >500 m were also discarded.

The copepod occurrence data displayed sampling effort biases (Appendix S1) that could inflate metrics of empirical models' performance or over-represent portions of the environmental space and thus hinder model predictability and interpretation (Veloz, 2009; Hijmans, 2012). We addressed these potential sampling biases by further thinning the species-level occurrence data (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015). For each month and each species separately, the occurrences were thinned by applying a randomization algorithm (30 randomizations per species dataset) that returned a 'thinned' dataset where monthly species occurrences were at least 500 km apart (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015). Only those species displaying at least 50 occurrences (n = 385) were retained and constituted the final list of species for which functional traits data were searched for in the literature.

2.2. Species functional trait data

The five following species-level functional traits were included in our analyses, as detailed below. These traits were chosen based on data availability from previous traits compilations and extended literature survey (see Appendices S2 and S3):

Body size (quantitative continuous): mean maximum adult female body size (i.e. length of the cephalothorax) in mm. Body size is considered a master functional trait as it impacts all life functions and scales most physiological rates (Kiørboe & Hirst, 2014; Hébert, Beisner, & Maranger, 2017) and influences predator-prey interactions (Hansen, Bjornsen, & Hansen, 1994).

Trophic group (categorical): most marine planktonic copepods are omnivorous, yet they can be grouped according to their preference in food sources (Kiørboe, 2011a; Pomerleau et al., 2015; Benedetti et al., 2016). Here, the following five groups were defined: Omnivore-Herbivore, Omnivore-Carnivore, Omnivore-Detritivore, strict Carnivore, and Omnivore. The trophic group describes the primary food source of a species and therefore its role in food-web dynamics.

Feeding mode (categorical): copepods have developed various strategies to detect and capture food items. The present feeding modes followed the definitions of Kiørboe (2011a): ambush-feeding, current-feeding, cruise-feeding, particle-feeding, current-cruise feeding, and current-ambush feeding (the last two referred to those mixed-feeding species that perform both strategies; Kiørboe, 2011a; Litchman et al., 2013). Ambush-feeding copepods lurk stationarily in the water column and detect the vibration generated by motile preys thanks to specialized appendages (i.e., hydromechanical perception) and capture prey through quick jumps. Copepods perform active ambush-feeding according to Kiørboe (2011a), which is considered a more passive feeding mode than cruise-feeding and current-feeding (Brun et al., 2017). Current-feeding describes the use of a scanning current that can detect and capture food items. This strategy is believed to favor a more efficient capture of many immobile preys like phytoplankton cells compared to ambush-feeding (Kiørboe, 2011a). Lastly, cruise-feeders correspond to copepods that swim actively through the water column in search of their prey.

Myelination (binary): myelinated copepod species have a lipid-rich myelin sheath around their nerves that increases the speed of nervous signals transmission and therefore enables faster attack or evasive reactions (Lenz, 2012). Myelin sheaths play a key role in modulating mortality and feeding rates and improving energy savings under low food conditions.

Spawning mode (binary): eggs are either released into the open water after fertilization (free-spawning) or remain attached to the female in specialized egg-sacs until hatching (sac-spawning). Sac spawning copepods display lower fecundity rates and longer hatching times compared to free-spawners (Kiørboe & Sabatini, 1994).

2.3. Definition of functional groups

We used multivariate analysis and hierarchical clustering to identify functional groups of copepod species that show similar functional trait combinations and therefore constitute different functional units in marine ecosystems. Only those copepod species with a body size value and no missing data for three out of five functional traits (343 out of the 385 species) were retained in the analysis (Appendix S2). The species' trophic groups and feeding modes were re-coded binarily to accurately represent species that belong to several feeding modes/ trophic groups (Appendix S2). This means that the analyses described below were performed on 343 species based on ten trait dimensions instead of the initial five dimensions.

Since the functional traits used represented both quantitative and qualitative variables, a factor analysis on mixed data (FAMD; Pagès, 2004) was used to investigate inter-species functional traits variance and estimate the functional distances between species. The use of a FAMD represented an improvement to the Multiple Correspondence Analysis (MCA) used in Benedetti et al. (2016) as it allowed to retain body size as a quantitative continuous variable instead of having to convert this trait into size classes. The FAMD reduced the dimensionality of the ten traits dataset by synthetizing them through a set of principal components (PCs) that described the main modes of functional trait variance. The number of retained FAMD components was chosen based on a leave-one-out cross-validation method where each object (i.e., species) is alternatively removed and predicted with a new FAMD model and a mean square error of prediction is calculated for each n components ranging from one to ten (Audigier et al., 2016). The mean square error was lowest for n = four components (error = 0.028, whereas error > 0.029 for all other values of n) and those first four FAMD components explained 80.15% of the variability in inter-species trait variance. An Euclidean distance matrix was calculated from these four components to represent the inter-species functional distance resulting from the FAMD. To assess the quality of this dimension reduction analysis, we compared this Euclidean distance to a Gower distance matrix calculated directly from the trait data (10 dimensions) following the guidelines of Mouillot et al. (2021), We calculated the co-ranking matrix between the two distance matrices and determined how the dimension reduction distorted the initial inter-species functional distance using the area under the curve

(AUC) criterion, which should be > 0.7 according to Mouillot et al. (2021). We found an AUC criterion of 0.81 which supports the “excellent” quality of our dimension reduction analysis (Mouillot et al., 2021). We used hierarchical clustering and Ward’s agglomeration link to generate the functional dendrogram. The number of clusters (i.e., FGs) was then chosen by the height at which the dendrogram is "cut".

We cared to investigate the sensitivity of the resulting FGs to the main parameters of our clustering approach (see Appendix S4 for results not reported below): (i) the effect of including species with missing values in the FAMD compared to restricting the pool to those species without any missing trait value (283 instead of 343; Appendix S2), (ii) the choice of the distance metric (Euclidean distance matrix based on FAMD components *versus* Gower’s distance matrix based on the trait values), and the agglomeration link (Ward’s aggregation link versus average link; Mouchet et al., 2008). Therefore, four combinations of distance matrices and agglomeration links were considered for two subsets of the species traits data. Including 60 copepod species with some missing trait values affected the structure FAMD space only marginally (Fig. S4.1) and did not change the number of components to be retained (n = 4), nor did it change their amount of explained variance (80.15%). Therefore, the FAMD including 343 species was kept as the standard since it allowed for a larger pool of species. To quantify the quality of the clustering procedure, four internal cluster stability indices (Calinski-Harabasz index, Connectivity, Dunn’s index, and average silhouette width) were calculated for each dataset, distance matrix choice, agglomeration link choice and for a number of clusters ranging from three to 13 (Figs. S4.2 and S4.3). Plus, Baker’s Gamma correlation coefficients (Baker, 1974) were computed between each functional dendrogram stemming from the four different combinations of distance metrics and agglomeration links to examine their pairwise similarity (Table S4.1). Lastly, we performed a careful expert inspection of each dendrogram and resulting FGs to further evaluate their similarity and ensure the ecological relevance of the final FGs (i.e., avoid few large groups that are functionally heterogenous or numerous small groups that are functionally redundant). Ultimately, eleven FGs were identified (Table S4.2).

2.4. Species distribution modelling

2.4.1. Environmental predictor selection

We considered 14 environmental predictors that drive the spatial ranges of copepod species in the open ocean and that are commonly used to model their abiotic habitats (see Benedetti et al., 2021 and references therein). A more exhaustive description of the data coverage and

initial resolution is given in Appendix S5. First, five predictors were retrieved from the World Ocean Atlas (WOA, version 2013v2): sea surface temperature (SST, °C), dissolved oxygen at 175 m depth (dO_2, μml l^{-1}), nitrate concentration (NO_3^-, μM), phosphate concentration (PO_4^{3-}, μM), silicate concentration (SiO_2, μM). The products N^* and Si^* were derived from the monthly climatologies of the above-mentioned nutrient fields: N^* refers to the excess of NO_3^- to PO_4^{3-} relative to the Redfield ratio ($N^* = [NO_3] - 16[PO_4^{3-}]$) and can be used as a tracer of denitrification and N_2 fixation. Si^* is the excess of SiO_2 to NO_3 ($Si^* = [SiO_2] - [NO_3^-]$) and is relevant to diatom growth, as healthy diatoms take up silicate and nitrate in a one-to-one ratio (Sarmiento & Gruber, 2006). Several additional variables were considered as they represent complementary niche axes: surface wind speed (Wind, m s^{-1}); eddy kinetic energy (EKE, m^2 s^{-2}) computed according to the algorithm of Qiu and Chen (2004) as a proxy for mesoscale activity; chlorophyll-a concentration (Chl, mg m^{-3}) as a proxy of surface phytoplankton biomass; surface carbon dioxide partial pressure (pCO_2, μatm); photosynthetically active radiation (PAR, μmol m^{-2} s^{-1}); and mixed-layer-depth (MLD, m). Also, photosynthetically available radiation (PAR) over the MLD (MLPAR, μmol m^{-2} s^{-1}) was added as an indicator of available light within the mixed layer. For NO_3^-, PO_4^{3-}, SiO_2, Chl and EKE, we added their log-transformed fields to achieve distributions that are closer to a normal distribution. The monthly climatologies of all predictor variables were all projected onto the 1°x1° cell grid of the WOA and their values were matched with the monthly species occurrence data.

We used a two-stage procedure to select the environmental predictors of the SDMs (section 2.4.3 for the models set up). First, we removed collinear predictors to avoid increasing the uncertainty in regression models projections through coefficients inflation (Dormann et al., 2013). The collinearity of predictors at the global scale and at the FGs-level was calculated for each species through pairwise Spearman's rank correlation coefficients. When a pair of predictors displayed an average correlation coefficient above the widely used |0.7| threshold (Dormann et al., 2013), the predictor closest to a normal distribution was kept. This first step narrowed down our initial 14 predictors to the following ten: SST, PAR, $logNO_3$, MLD, logChl, logEKE, Si^*, N^*, Wind, and pCO_2. Second, to avoid model overfitting the number of environmental predictors included in the SDMs was restricted to five to achieve a 10:1 ratio of occurrences to predictors, following Guisan et al. (2017). To choose the five most influential predictors, univariate random permutation tests were performed (n = 30 repetitions; Appendix S6). For each species and SDMs, one of the ten potential predictors was randomly reshuffled while the other nine were kept as is and SDMs were trained based on this reshuffled dataset. Then, a correlation coefficient was calculated between the original vector

of unshuffled model prediction and the vector of predictions resulting from the reshuffled dataset. The higher the correlation coefficient, the lower the importance of the reshuffled predictor on the final SDM prediction. We chose the top five predictors for each FG and SDM separately (Appendix S6). These were then used to train the SDMs used for global habitat projections. For those species that could not be assigned to a FG due to missing functional traits data, the five predictors displaying the highest ranks of relative importance across all species combined were used in the SDMs (Appendix S6).

2.4.2. Background data

Correlative SDMs such as the ones used here require both presence and absence data. Yet, the occurrence data used here typically lack absence data, therefore pseudo-absence data were generated (Benedetti et al., 2021). We followed the target-group approach of Phillips et al. (2009) which is appropriate to model plankton species distributions based on their relatively sparse data (Righetti et al., 2019; Benedetti et al., 2021). First, we investigated the spatial distribution of the occurrence data of each FG separately. All FGs displayed similar patterns of sampling effort except for FG2, FG4 and FG11 (Appendix S4), which had very few occurrences in the Southern Ocean and the North Atlantic Ocean (Appendix S1). Therefore, we chose to draw pseudo-absences from the total pool of sites (i.e., monthly 1° x 1° grid cells) that displayed at least ten occurrences of all species included, except for those species belonging to the three groups above. For species of FG2, FG4 and FG11, pseudo-absences were only drawn from those sites where at least 10 different species of their FG were found (i.e., their "target group"). The threshold of ten occurrences per site was chosen to avoid relying on those sites where only a small fraction (e.g., only one or two species) of the overall copepod community was detected. Once the sites were defined, the pseudo-absences of each species were randomly drawn based on their occurrences in both their corresponding target groups. This way, a species' background is located at the sites where its lack of presence is most likely to reflect an actual absence. For each species, ten times more background data than presences were generated following the guidelines of Barbet-Massin et al. (2012), and presences were weighted ten times more than pseudo-absences.

2.4.3. Model configuration and habitat suitability projections

Following Benedetti et al. (2021), we developed an ensemble modelling approach that combined three SDMs types which were configured to avoid model overfitting (Merow et al., 2014): Generalized Additive Models (GAMs), Generalized Linear Models (GLMs), and

Artificial Neural Networks (ANNs). The GLM were run using a quadratic formula and a logit link function. The GAMs were also set up to use a logit link function and a maximum of five smoothing terms. ANNs were run using five cross validations with 200 iterations to optimize number of unities in the hidden layer and the parameter for weight decay (Thuiller et al., 2016). The SDMs were trained on 80% of the presence/pseudo-absence data chosen at random and tested on the remaining 20%. For each SDM, ten random cross evaluations runs were carried out. For each cross-validation run, the true skill statistic (TSS) was calculated to evaluate model performance. Only species displaying a mean TSS score >0.30 were used for the final species and functional trait projections.

The SDMs were then projected in the conditions of the global ocean as a function of all monthly climatologies of the environmental predictors included in the model to obtain species-level maps of habitat suitability indices (HSI). Since we used FGs-specific and SDM-specific sets of predictors, the SDMs vary in number of successfully modelled species (254 for GAM, 244 for ANN, and 238 species for GLM). For each FG, annual mean HSI values were obtained by averaging the monthly HSI maps across each species constituting the FG and SDMs. These mean annual HSI estimates indicate which regions are suitable for each species/FG.

2.5. Positioning FGs in environmental niche space

Following the approach of Benedetti, Vogt, et al. (2018), the mean univariate response curves emerging from the GAMs were used to estimate the relative optimal conditions (i.e., niche center) and tolerance (i.e., niche width) for each species and predictor (see Appendix S7). The niche center was calculated as the weighted median and the niche width as the difference between the weighted 10th and 90th percentiles. The HSI values were used as weights. For this analysis, a set of eight predictors common to all species (i.e., unlike the top five FG-specific predictors used for mapping, as described in section 2.4.1) had to be defined based on the ranking of predictors (Appendix S6): SST, logChl, logNO_3, MLD, Si^*, PAR, N^* and logEKE. Eight predictors were chosen to cover all top ranking predictors based on the overall rankings. To investigate the inter-species and inter-FGs similarities in the above-mentioned niche characteristics and test if the FGs present distinct environmental niches, a principal component analysis (PCA; Legendre & Legendre, 2012) was performed on the niche centers and niche widths for all 8 predictors to ordinate the species according to their niche center and niche width. This space composed by the principal components (PCs) of the PCA is hereafter referred to as the niche space. To explore whether FGs differ significantly in their position in

niche space, we averaged the PCs scores of each FG based on the scores of their constituting species. Non-parametric variance analysis (Kruskal-Wallis tests) was carried out to test if FGs differ in their position in niche space). Post-hoc variance analyses (Dunn's test with a p-values adjustment following Bonferroni's method) were then performed to identify the pairs of FGs that displayed significant variations in niche space positions.

2.6. Community weighted median (CWM) proportions of traits

On top of projecting the FGs-specific mean annual patterns of HSI, the CWM trait values for body size, trophic group, feeding mode, myelination, and spawning mode were computed to explore the biogeography of copepod traits. All qualitative traits were mapped by calculating a CWM trait value based on the monthly species-specific HSI values as weights. More precisely, for each grid cell and month, the CWM proportion was calculated as the sum of all HSI values belonging to species exhibiting that trait over the sum of HSI values across all species in that community (i.e. all the species in that grid cell). For body size (i.e., the only continuous trait), we calculated the weighted median body size for each community, again using monthly HSI values as weights. The CWM values were calculated according to Cormen et al. (2009) using the matrixStats R package. Like for FGs, monthly CWM projections were performed for every SDM type and mean annual CWM trait values were derived based on the ensemble of projections (Appendices S8 and S9).

2.7. Ocean regionalization based on functional trait biogeography

To explore sub-global patterns of traits and FGs expression rather than reporting results for broad latitudinal bands (e.g., going beyond the tropics *versus* the poles), clustering was used to divide the global ocean into regions that displayed similar trait expressions (e.g., mean annual projections of CWM body size, and CWM values of each trophic groups, feeding modes, myelination, and spawning mode; Appendix S10). First, the spatial patterns of the twelve CWM traits values were summarized through a PCA again (Appendix S10) and the scores of each grid cell along the first four PCs (97.9% of total variance explained) were used to derive a global Euclidean distance matrix. Second, following the same approach as in section 2.3 and in Benedetti et al. (2021), various clustering approaches were explored based on this distance matrix (three hierarchical clustering approaches and two widely used partitioning approaches) and three internal cluster stability metrics (Dunn's index, Connectivity, and the average silhouette width) were used to help us find an optimal number (k) of regions (see Appendix S11). The three hierarchical approaches consisted in using three

alternative agglomeration linkages (average, complete and Ward's; Legendre & Legendre, 2012). The stability metrics enabled us to discard the two partitioning methods (kmeans and partitioning around medoids; Fig. S11.1) and to narrow k down to 3 to 8 regions (Fig. S11.2). Ultimately, we chose to present the CWM trait patterns obtained for k = 4 under Ward's linkage based on the profiles of cluster stability metrics and because this linkage methods minimizes intra-cluster variance. Values of k > 4 simply lead to smaller regions nested in those obtained for k = 4.

3. Results

3.1. Copepod FGs

The 343 copepod species retained were clustered based on their combinations of body size, feeding mode, trophic group, myelination, and spawning mode. Eleven FGs could be derived from the functional dendrogram, which reflects inter-species functional similarity (Fig. 1; Appendix S4). The first dichotomy in the dendrogram occured between the small to medium-sized cruise feeders (FG1, 2, 3) and other feeding modes. FG1 was composed of 14 species that are small (median ± IQR = 1.10 ± 0.52 mm) cruise-feeding herbivores. Almost all species (13 out of 14) were sac-spawning and myelinated species. All but one species belonged to the *Clausocalanus* genera. FG2 was defined by small (0.80 ± 0.30 mm) cruise-feeding detritivores. The 22 species of this group were all amyelinated and sac-spawning. All species in group 2 belonged to the Oncaeidae family. FG3 was composed of 46 medium-sized (2.61 ± 1.70 mm) detritivores. Most species in FG3 were myelinated (74%) and free-spawning (80%). Cruise- and current-feeding were the two predominant feeding modes, with *Spinocalanus*, *Metridia* and *Scaphocalanus* being the dominant genera in this FG.

The second dichotomy in the dendrogram separated the small and large sac-spawning carnivores (FG4 and 5) from the rest. FG4 consisted of 29 small (1.22 ± 0.70 mm) carnivorous ambush-feeders. They were all sac-spawning and amyelinated. The 29 species were part of the Corycaeidae family and thus belonged to either the *Corycaeus*, *Farranula* or *Vettoria* genera. FG5 was made up of 29 species of large (4.58 ± 2.75 mm) sac-spawning current- and cruise-feeding carnivores from the Sapphirinidae and Euchaetidae families.

The third dichotomy of the dendrogram separated the large and medium-sized current feeders (FG6 and 7) from the rest. FG6 was a group of 13 very large (6.80 ± 2.15 mm) species that were either current-feeding herbivores (85%) or fully omnivorous (15%). They were all myelinated free-spawners. The two most represented genera were *Calanus* and *Eucalanus*. FG7 was the largest FG with 55 medium-sized (1.80 ± 2.26 mm) and current-feeding

herbivores. Most were myelinated (95%) and free-spawning (89%). The main genera contributing to the composition of FG7 were *Calanus*, *Calocalanus* and *Paracalanus*.

The fourth dichotomy occured between the small and medium-sized omnivores and the medium free spawning carnivores (FG8). FG8 consisted of 49 medium-sized (3.40 ± 1.7 mm) omnivorous-carnivorous species which were predominantly amyelinated (86%) and free-spawning (80%). This diverse group mixed current- (60%) and ambush-feeders (40%) and gathered 12 genera with *Candacia*, *Haloptilus* and *Heterorhabdus* being the most dominant. FG9 was composed of 40 medium-sized (2.76 ± 1.98 mm), current-feeding omnivorous species. All were free-spawning and mostly amyelinated (72%). *Pleuromamma*, *Gaetanus*, and *Labidocera* were the main genera.

Finally, the fifth and last main dichotomy separated the small and medium-sized mixed-feeding omnivores (FG11) from the small ambush feeders (FG10). FG10 was also rather homogeneous and contained 14 small (1.10 ± 0.52 mm), ambush-feeding, amyelinated and sac-spawning omnivores. All species belonged to either the *Oithona* or the *Dioithona* genera from the Oithonidae family. FG11 was a group of 32 small (1.50 ± 0.67 mm) species predominantly belonging to the *Acartia* and *Centropages* genera. The species belonging to *Acartia* were omnivorous-herbivorous, whereas the others were omnivorous. All species in FG11 were amyelinated free-spawning current-ambush feeders.

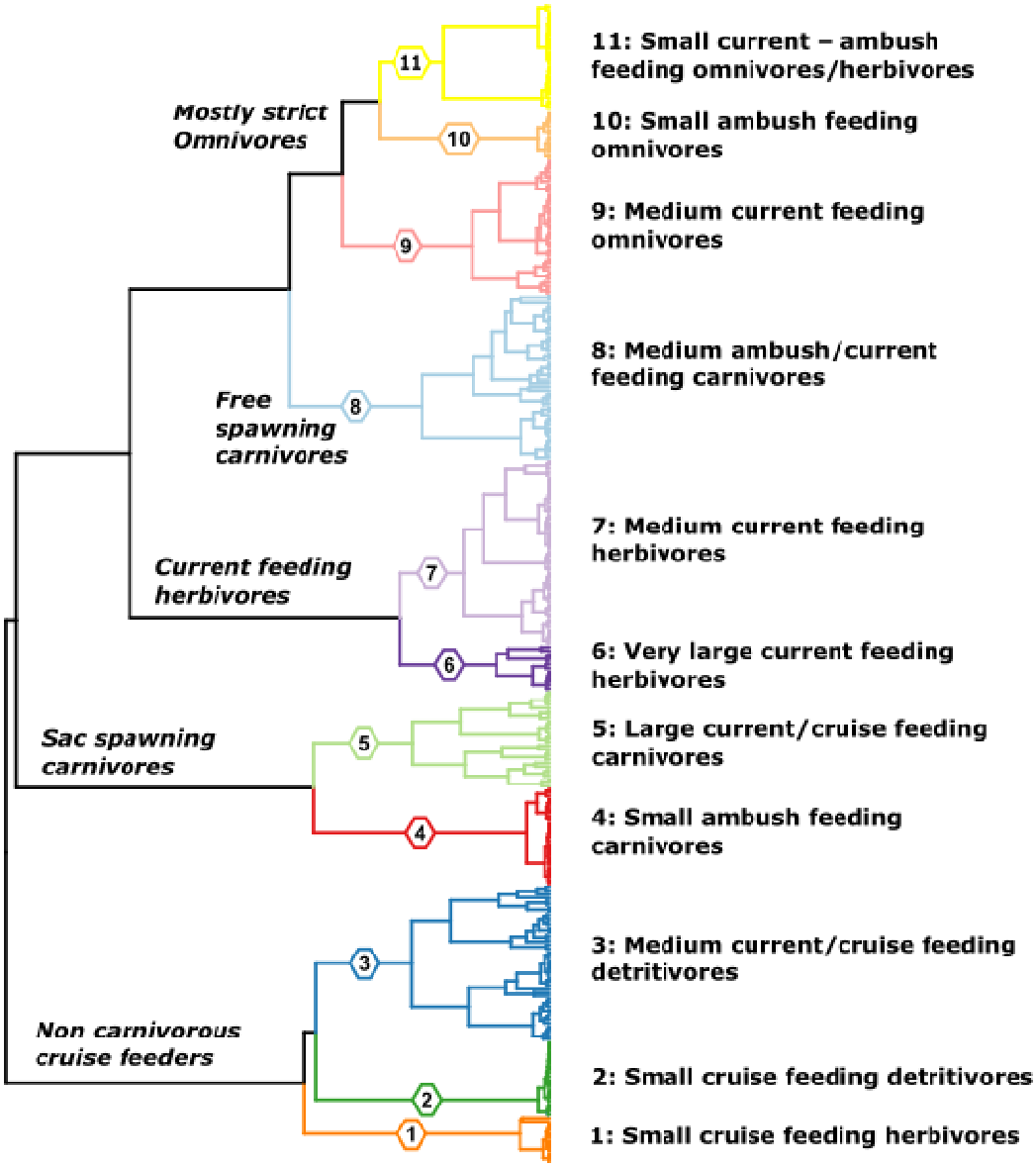


Figure 1: Functional dendrogram representing the inter-species traits dissimilarity for 343 copepod species based on their functional trait combinations. The hierarchical clustering was performed on a Euclidean distance matrix issued from the species coordinates ensuing from a Factor Analysis of Mixed Data (FAMD). Each leaf of the functional dendrogram represents a copepod species and the eleven functional groups (FGs) identified through the hierarchical clustering approach are numbered and highlighted in color. The main taxonomic groups representing the FGs are as follows: FG1: *Clausocalanus* spp.; FG2: Oncaeidae; FG3: *Spinocalanus*, spp., *Scaphocalanus* spp. and *Metridia* spp.; FG4: Corycaeidae; FG5: Sapphirinidae and Euchaetidae; FG6: Calanidae; FG7: *Paracalanus* spp., *Calocalanus* spp., and *Calanus* spp.; FG8: *Haloptilus* spp., *Heterorhabdus* spp. and *Candacia* spp.; FG9: *Pleuromamma* spp., *Gaetanus* spp., and *Labidocera* spp.; FG10: Oithonidae; FG11: *Acartia* spp. and *Centropages* spp.

The functional space defined by the PCs of the FAMD (Appendix S12) served as an alternative description of the reduced functional space. The first four PC of the FAMD explained 79.21% of the total variance in functional traits. The largest functional distance along the first component was found between FG4 and FG6. The Euclidean distance matrix represented on the functional dendrogram (Fig. 1) was compared to a Gower distance matrix computed from the functional traits (i.e., no dimension reduction; Appendix S4) based on the AUC criterion. We found an AUC of 0.808, which indicated that the reduced FAMD space results in a fairly good representation of the inter-species functional dissimilarity according to Mouillot et al. (2021). Furthermore, the importance of a functional trait in scoring the functional space was quantified by examining the drop in AUC score if said trait was omitted. Body size was found to be the most important (-0.322 in AUC), followed by trophic group (-0.265), feeding mode (-0.163), spawning mode (-0.091), and myelination (-0.063).

3.2. FGs in environmental niche space

The environmental niches of the copepod species were described by their univariate niche centers and widths for the top eight predictors we defined as part of our core set or predictors (section 2.5, Appendix S6), and these were used to examine FGs position in niche space based on a PCA (Fig. 2; see Appendix S13 for PCs 3 and 4). The niche characteristics that contributed the most positively to PC1 (relative contribution to PC1 given in brackets when >5%) were: logChl center (12.16%), logChl width (12.51%), EKE width (11.78%), MLD center (9.96%), MLD width (10.78%), N^* width (7.48%), Si^* width (5.64%) and Si^* center (5.19%). The species with negative scores on PC1 were those that are characterized by higher SST width (5.65%) and N^* center (4.45%). The three most important niche characteristics scoring PC2 were: Si^* center (30.88%), Si^* width (27.90%) and SST center (11.17%). Species with positive PC1 scores were those affiliated with wider niches and conditions of higher concentrations of nutrients and chlorophyll-a, stronger seasonal variations, and overall higher water column turbulence and mixing.

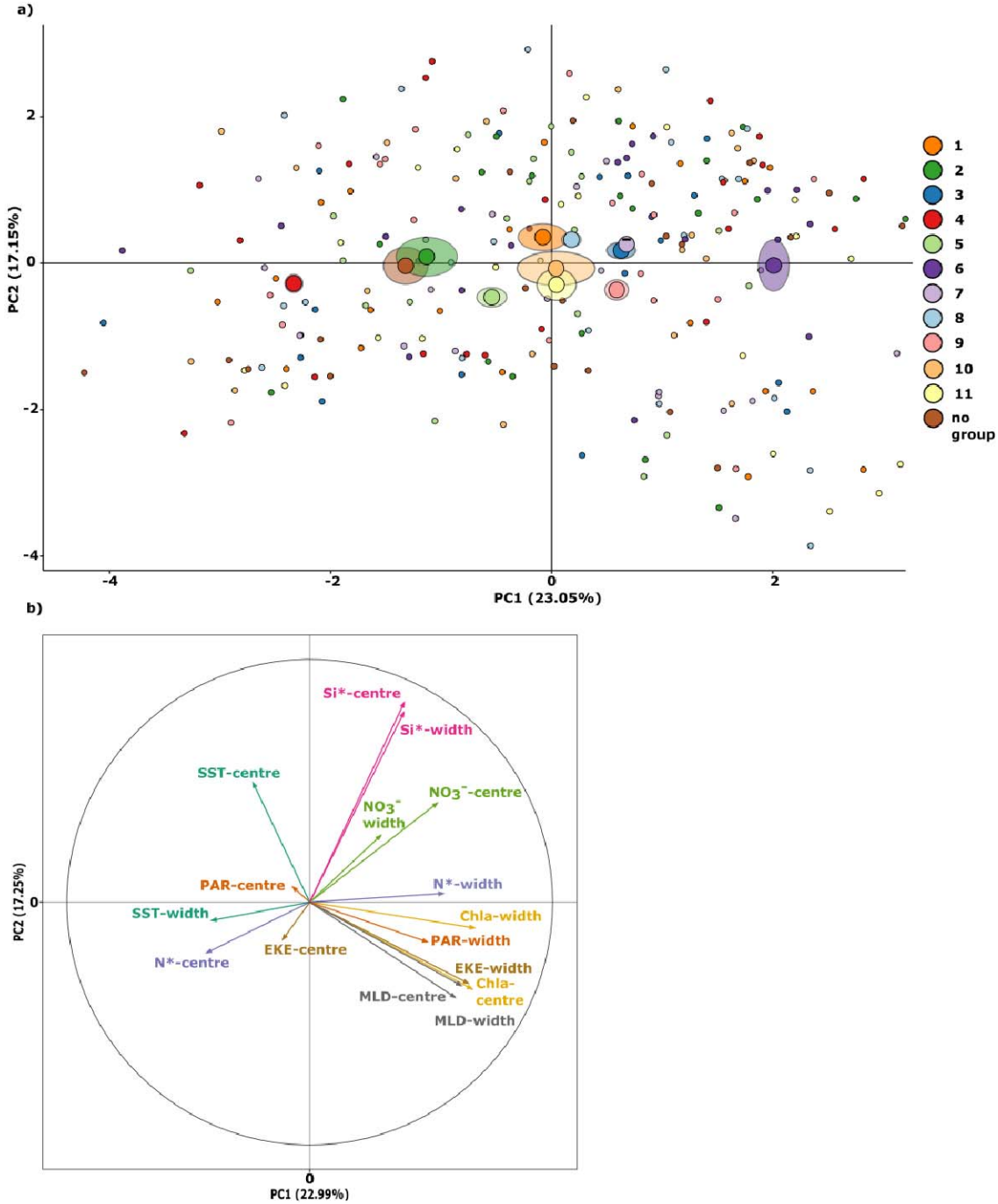


Figure 2: Position of a) functional groups (FGs, larger circles) and species (smaller circles) in environmental niche space according to the first two principal components (PCs) of a principal component analysis (PCA) performed on b) the species-level niche characteristics derived from GAM-based univariate response curves of the chosen environmental predictors. Niche centers were calculated as the weighted median value of the corresponding predictor as an estimate of the species' optimum for each predictor. Niche widths were calculated as the range between the weighted 10th and 90th quantiles and estimate the relative tolerance range of the species to the corresponding predictor. The semi-transparent ellipses indicate two times the value of the standard errors associated with the mean PC scores of the FGs.

Post-hoc variance analyses showed significant (Dunn's tests; $p < 0.05$) inter-FGs variations in niche characteristics and PC scores (see Appendix S14). The largest distance in niche space (PC1) was found between FG4 and FG6 (Fig. 2a; p = 6.5e-10). Along PC1, FG4 also differed significantly from FG1 (p = 0.042), FG3 (p = 1.6e-06), FG7 (p = 1.3e-08), FG8 (p = 1.4e-4), FG9 (p = 2.3e-06), FG10 (p = 0.011) and FG11 (p = 0.005). Conversely, FG6 differed significantly from FG2 (p = 4.5e-4), FG5 (p = 0.001) and FG8 (p = 0.049) along PC1. FG6 also differed significantly from FG2 (p = 0.001) and FG5 (p = 0.002). None of the FGs did showed significant variations along PC2. Only FG3 and FG7 showed significant variations along PC3 (p = 4.2e-3; Appendices S13 and S14). To summarize, when FGs 4 and 6 were not accounted for, none of the remaining nine FGs showed significant variations in PC1 and PC2 scores (all $p > 0.05$). This was due to the high level of inter-FG environmental niche characteristics overlap, which implies that distinct FGs display broad environmental niches and can share similar abiotic habitats.

3.3. Mean annual habitat suitability indices (HSI) patterns

For all three SDMs and 11 FGs combined, we found SST to be the most important predictor for constraining the models, followed by PAR, $logNO_3$, MLD, logChl, logEKE, Si^*, N^*, Wind, and pCO_2 (Appendix S6). SST remained the most important predictor for every FG when looking at predictor importance per FG. Therefore, most FGs displayed mean annual HSI patterns that were driven by the latitudinal temperature gradients at the first-order. The impact of second-order predictors became clearer on a sub-global scale. Aggregating the average monthly HSI projected across SDMs per FG allowed us to estimate their mean annual habitat suitability patterns (Fig. 3).

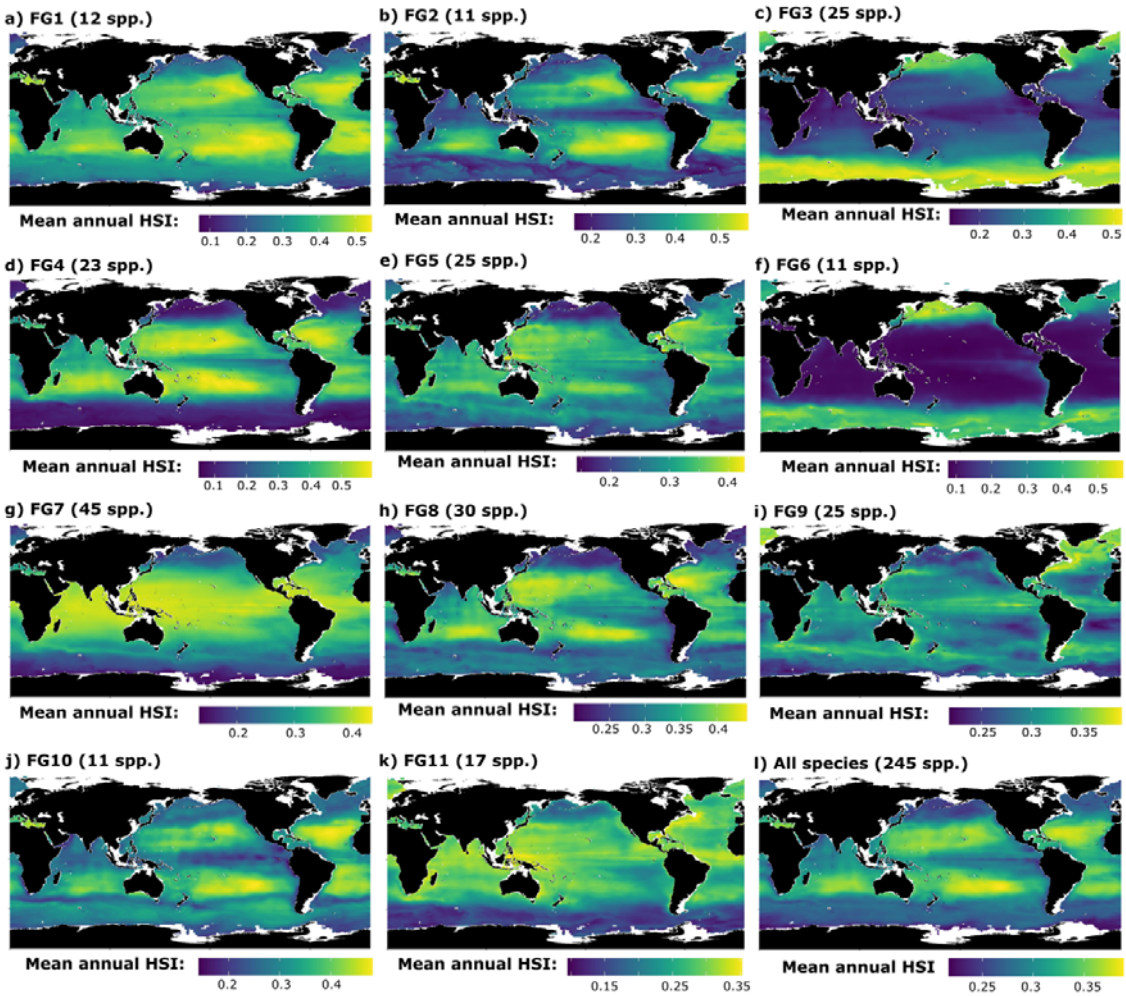


Figure 3: Maps of mean annual habitat suitability index (HSI) the eleven copepod functional groups (FG) identified in the present study: a) FG1 (small myelinated cruise-feeding omnivores-herbivores), b) FG2 (small non myelinated cruise-feeding detritivores), c) FG3 (medium-sized myelinated mixed-feeding or cuise-feeding detritivores), d) FG4 (small amyelinated ambush-feeding carnivores), e) FG5 (large current- or cruise-feeding carnivores), f) FG6 (very large myelinated current-feeding omnivores-herbivores), g) FG7 (medium-sized myelinated current-feeding omnivores-herbivores), h) FG8 (medium-sized amyelinated ambush- or current-feeding carnivores), i) FG9 (medium-sized amyelinated current-feeding omnivores), j) FG10 (small amyelinated ambush-feeding omnivores), k) FG11 (small amyelinated mixed-feeding omnivores) and for l) all species together. Mean annual estimates were derived from the 12 monthly estimates of mean HSI obtained for each of the three species distribution models (SDMs) used (generalized linear models, generalized additive models and artificial neural networks). The FG were defined based on the functional traits combinations of the 245 copepod species described by a factorial analysis on mixed data (FAMD) whose principal components were used to perform hierarchical clustering on a Euclidean distance matrix with Ward's aggregation link.

The mean annual HSI of FG3 was found to be maximal towards the poles and decrease progressively towards the equator. In contrast, the mean annual HSI of FG1, FG2, FG4, FG8, and FG10 were found to be highest in the tropics and decreased towards higher latitudes. For those five FGs, slighter decreases in HSI were modelled towards the tropical upwelling

systems (e.g., Peru, Benguela), the northern part of the Indian Ocean and the Pacific Equatorial counter current. Therefore, these five FGs reached maximal HSI in the oligotrophic conditions of the tropical gyres. Meanwhile, FG5, FG9, and FG11 show less marked latitudinal gradients in mean annual HSI values. No clear hotspot in HSI could be found for these three FGs, but they all displayed lower HSI in higher latitudes than in the tropics. Yet, the region where their mean annual HIS was lowest varied between groups: the North Pacific and parts of the Southern Ocean for FG5 and FG9, the whole Southern Ocean for FG11. The mean annual HSI pattern of FG7 followed the annual SST gradient very closely, with maximal HSI values near the equator and progressive decrease towards the poles.

3.4. Functional trait biogeography from CMW trait values and regionalization

We projected the CWM body size and the CWM values of trophic group, feeding mode, spawning mode, and myelination for the global open ocean on a mean annual scale to illustrate the biogeography of these key traits (Appendix S9). To summarize the main spatial gradients of trait biogeography and assess how these could overlap with known large scale oceanographic currents (i.e., currents and fronts), the global ocean was clustered based on similarities in CWM values after reducing their dimensionality (see Appendices S10 and S11). We identified four ocean regions that display significant contrasts in CWM trait proportions (Fig. 4; Appendix S15).

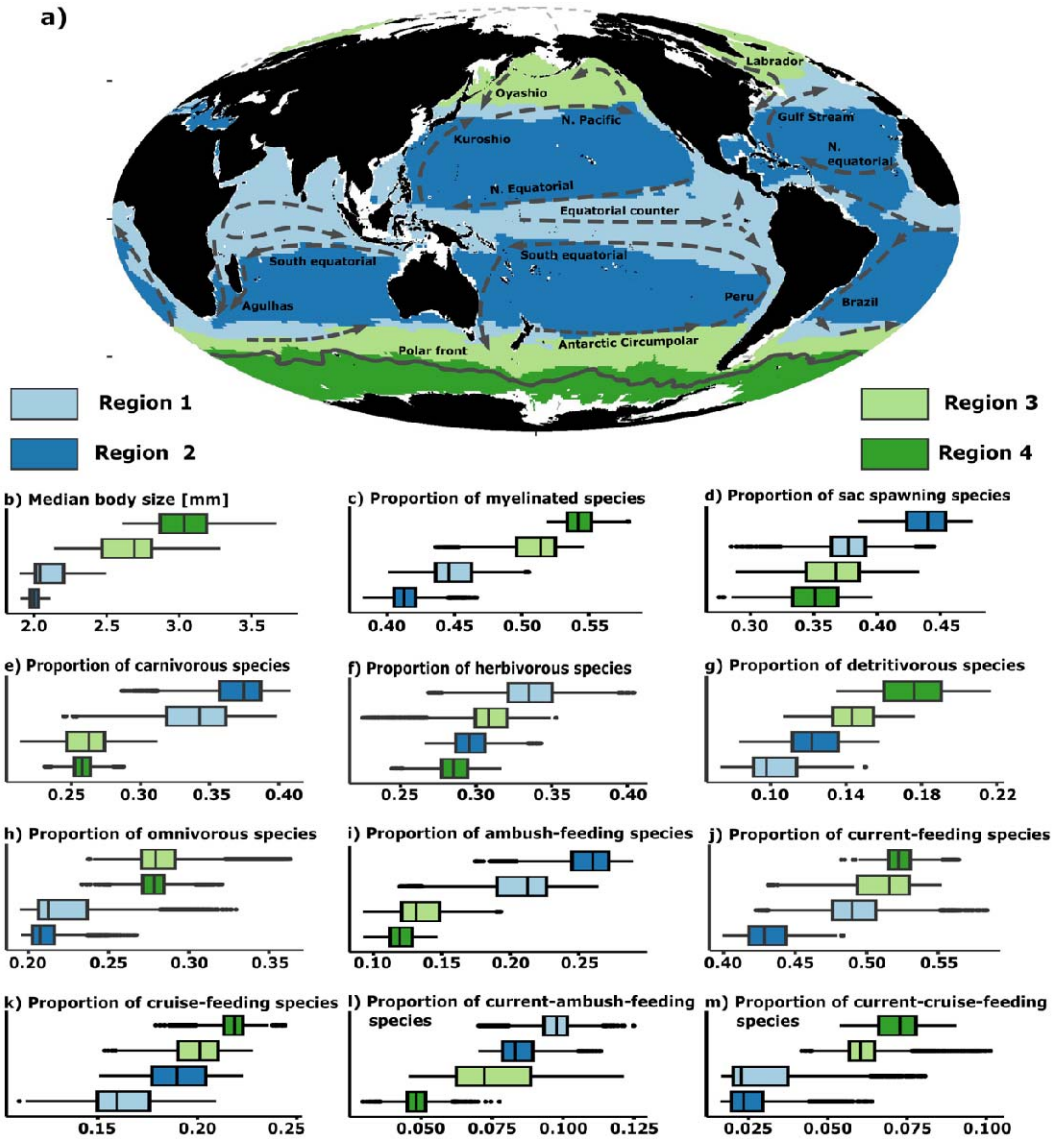


Figure 4: The global ocean divided into a) four regions according to the principal components of a Principal Component Analysis (PCA) based on the CWM values of the functional traits. Major oceanographic circulation features (i.e., surface currents and fronts) are illustrated through dashed arrows to highlight their overlap between the regions' boundaries. The distribution of CWM trait values between the four regions are shown through the boxplots: b) median body size, CWM values of c) myelinated species, d) sac spawning species, e) carnivores, f) herbivores, g) detritivores, h) omnivores, i) ambush-feeders, j) current-feeders, k) cruise-feeders, l) current-ambush-feeders, and m) current-cruise-feeders. The position of the Antarctic Polar Front was taken from Orsi and Harris (2019). The positions of the main large-scale ocean currents were drawn according to Pidwirny (2006). The lower, middle, and upper boundaries of the boxplots correspond to the 25th, 50th, and 75th percentiles respectively. The lower and upper whiskers extend no further than 1.5*IQR (interquartile range) from the lower and upper hinges.

Region 1 primarily fell within the equatorial band and comprised coastal upwelling regions and the main oxygen minimum zones. The copepod communities of Region 1 displayed lower

CWM body size (median ± IQR = 2.043 mm ± 0.194), higher proportions of carnivores (0.342 ± 0.040) and herbivores (0.335 ± 0.029) but lower proportions of omnivores (0.213 ± 0.030) and detritivores (0.098 ± 0.023). The CWM values of myelinated (0.446 ± 0.026), sac-spawning (0.377 ± 0.026), and ambush-feeding (0.212 ± 0.036) species were lower than the CWM traits values found for the communities of Region 2 and higher than those found for the high latitude communities of Regions 3 and 4 (Appendix S13).

Region 2 comprised the tropical gyres. We found that communities in Region 2 showed the lowest CW median body size (2.002 ± 0.057 mm) but the highest CWM values of ambush-feeding (0.259 ± 0.027) carnivorous (0.372 ± 0.028), and sac-spawning (0.440 ± 0.030) copepods. Conversely, the CWM values of myelinated (0.413 ± 0.016) species in Region 2 were the lowest across all four regions.

Region 3 was located poleward to the previous two regions and covered the North Atlantic Ocean, the North Pacific Ocean and the waters located between the Polar Front and the Antarctic Circumpolar Current. The CWM body size (2.691 ± 0.342 mm), the CWM values of myelinated (0.514 ± 0.028) and omnivorous (0.279 ± 0.021) species were higher there than in Regions 1 and 2. In contrast, the CWM values of sac-spawners (0.368 ± 0.041), carnivores (0.266 ± 0.026), and ambush feeders (0.132 ± 0.027) was lower.

Region 4 mainly corresponded to the Southern Ocean (i.e., grid cells south of the Antarctic Polar Front, Fig. 4a). Region 4 displayed the highest CWM body size (3.035 ± 0.321 mm), higher CWM values of myelinated (0.541 ± 0.018), omnivorous (0.278 ± 0.013), and detritivorous (0.176 ± 0.030) species, and the lowest CWM values of sac-spawners (0.351 ± 0.035). The CWM values of current-feeding (0.522 ± 0.015) and cruise-feeding (0.218 ± 0.009) species was higher than those of ambush-feeding (0.120 ± 0.016).

4. Discussion

4.1. Towards meaningful global zooplankton FGs in marine ecology

Here, we identified eleven copepod FGs based on combinations of species-level functional traits. Nine of these eleven FGs had also been found, or were nested within larger groups, in previous studies based on regional species pools (Pomerleau et al., 2015; Benedetti, Gasparini, & Ayata, 2016; Benedetti, Vogt, et al., 2018; Becker et al., 2021; summarized in Table 1), which suggests that most of the functions performed by copepods on a global scale should also be expressed at regional scales.

Table 1: Table summarizing the overlap between the copepod functional groups (FGs) defined in the present study and those found in previous studies based on species composition and functional trait composition.

This study	Main clade	Main traits	Pomerleau et al. (2015) - North Pacific Ocean	Benedetti et al. (2016) - Mediterranean Sea	Benedetti et al. (2018) - Mediterranean Sea	Becker et al. (2021) - South Atlantic Ocean
FG1	Clausocalanus	Small myelinated cruise-feeding omnivores-herbivores	Small current- and cruise-feeding omnivore-herbivores (Group 6)	Small cruise-feeding omnivores-herbivores (subset of Group 6)	Small cruise-feeding omnivores-herbivores (Group 7)	Small and large myelinated cruise- and current-feeding omnivores-herbivores (subset of Group B)
FG2	Oncaeidae	Small non myelinated cruise-feeding detritivores	No equivalent	Small cruise-feeding omnivores-detritivores (subset of Group 6)	Small sac-spawning detritivores (subset of Group 5)	Small amyelinated ambush- or cruise-feeding carnivores or detritivores (small subset of Group D)
FG3	Spinocalanus, Scaphocalanus, Metridia	Medium-sized myelinated mixed-feeding or cuise-feeding detritivores	No equivalent	Small cruise-feeding omnivores-detritivores (subset of Group 6)	Small sac-spawning detritivores (small subset of Group 5)	No equivalent
FG4	Corycaeidae	Small amyelinated ambush-feeding carnivores	No equivalent	Small ambush-feeding carnivores (Group 2)	Small ambush-feeding carnivores (Group 2)	Small amyelinated ambush- or cruise-feeding carnivores or detritivores (small subset of Group D)
FG5	Sapphirinidae, Euchaetidae	Large current- or cruise-feeding carnivores	No equivalent	Large cruise-feeding carnivores (Group 1)	Large cruise- or current-feeding feeding carnivores (Group 1)	Small and large myelinated cruise- and current-feeding omnivores-herbivores (subset of Group B)
FG6	Calanus, Eucalanus	Very large myelinated current-feeding omnivores-herbivores	No equivalent	No equivalent	No equivalent	No equivalent
FG7	Calanus, Paracalanus, Calocalanus	Medium-sized myelinated current-feeding omnivores-herbivores	Small and large current-feeding omnivores-herbivores (Group 5c)	Small and large current-feeding omnivores-herbivores (Group 4)	Small and large current-feeding omnivores-herbivores (Groups 3 and 4)	Large myelinated current-feeding omnivores-herbivores (subset of Group C)
FG8	Candacia, Haloptilus, Heterorhabdus	Medium-sized amyelinated ambush- or current-feeding carnivores	No equivalent	Large cruise-feeding carnivores (Group 1)	Large cruise- or current-feeding feeding carnivores (Group 1)	No equivalent
FG9	Pleuromamma, Gaetanus,	Medium-sized amyelinated	No equivalent	Small and large current-feeding	Small and large current-feeding	No equivalent

	Labidocera	current-feeding omnivores		omnivores-herbivores (subset of Group 4)	omnivores-herbivores (small subset of Group 4)	
FG10	Oithonidae	Small amyelinated ambush-feeding omnivores	No equivalent	Small ambush-feeding omnivores (Group 5)	Small ambush-feeding omnivores (Group 6)	Small amyelinated mixed- or ambush-feeding omnivores (Group A)
FG11	Acartia, Centropages	Small amyelinated mixed-feeding omnivores	Small ambush-feeding omnivores (subset of Group 4)	Small mixed-feeding omnivores (Group 3)	Small mixed-feeding omnivores (small subset of Group 4)	Small amyelinated mixed- or ambush-feeding omnivores (Group A)

The four FGs defined by Becker et al., (2021) based on a series of cruises in the South Atlantic Ocean were quite broad and often functionally heterogeneous (i.e., large carnivorous copepods mixed with smaller current-feeding omnivorous-herbivorous ones, or small cruise-feeding particles feeders mixed with small ambush-feeding carnivores). As a result, our FGs are often nested within, or scattered across, the groups defined by Becker et al. (2021).

Only the present FG3 and FG6 have no true counterparts in previous studies. These two groups find their most suitable habitats towards the poles (Fig. 3) whose zooplankton communities were not covered by the regional studies mentioned. The copepods of FG3 (*Spinocalanus* spp., *Metridia* spp., or *Scaphocalanus* spp.) are known to mainly inhabit deeper ocean layers where they feed on falling particulate organic matter and zooplankton carcasses (Yamaguchi et al., 2002; Sano et al., 2013). Therefore, this group contributes to the remineralization of organic matter and marine snow at higher latitudes and/or in colder conditions. FG6 comprises the largest current-feeding omnivorous-herbivorous copepods from the Calanidae family (*Eucalanus* spp. and *Calanus* spp.). Such large copepods are not present in the warm and oligotrophic conditions of the Mediterranean Sea hence their absence in Benedetti, Gasparini, & Ayata (2016) and Benedetti, Vogt, et al. (2018). FG6 is a key group for the biological carbon pump and lipid pump (Jónasdóttir et al., 2015; Visser et al., 2017; Steinberg & Landry, 2017; Brun et al., 2019) as it represents large-bodied grazers that can actively feed on microphytoplankton, perform relatively strong vertical migrations and generate large and fast sinking pellets (Stamieszkin et al., 2015; Ohman & Romagnan 2016; Brun et al., 2019). This is confirmed by the position of FG6 in niche space (Fig. 2) as it is affiliated to turbulent and seasonally varying conditions with higher nutrient and chlorophyll-

a concentrations. The ecological roles and functions ensured by the other nine FGs have already been detailed and discussed in those previous studies. We support the statements of these authors and how they described the ecological roles of the copepod FGs. Here, we found a larger range of FGs compared to previous studies due to three main reasons: (i) we investigated a global and larger (hundreds of species instead of tens) pool of species, (ii) we accounted for myelination as an additional trait, and (iii) we retained body size as a continuous trait, contrary to Benedetti, Gasparini & Ayata (2016) and Benedetti, Vogt et al. (2018) whom relied on size classes which likely smoothed out important trait variations.

Overall, the largest differences in niche space (Fig. 2a, Appendix S14) occurred between two main sets of FGs: (i) myelinated free-spawning large-bodied, or medium-bodied, current-feeding herbivores (FG3, 6, 7 and 9), and (ii) amyelinated sac-spawning small-bodied, or large-bodied, cruise- and ambush-feeding detritivores and carnivores (FG2, 4 and 5). The first set is associated with conditions of stronger mixing and higher nutrients and chlorophyll-a concentrations (Fig. 2b). The species constituting these FGs also show larger niche widths, suggesting they display broader tolerances to monthly variations in environmental conditions. Therefore, these FGs are found more frequently in high latitude environments, or boundary current systems, where either seasonality or horizontal and vertical mixing lead to higher mean annual productivity (Sarmiento & Gruber, 2006; Roy 2018). In contrast, the FGs of the second set display narrower niches and are associated with conditions typical of the warmer tropical oligotrophic gyres (weaker water mixing and lower nutrients and chlorophyll-a concentrations; Fig. 2b). Meanwhile, the groups that are in the center of the niche space (FG1, 8, 10 and 11) are those whose species are very scattered across said space, meaning these FGs could not be associated to any particular environment at the scale of our study.

This continuum of FGs in niche space was also found in functional trait space. Indeed, the FGs of sets (i) and (ii) were often found on opposite sides of the FAMD dimensions (see Appendix S12). The global spatial distribution of the FGs and CWM trait values (Figs. 3 and 4) further support the statements above: larger myelinated free-spawning and active feeding copepods occur more frequently near the poles, whereas smaller amyelinated sac-spawning and passive feeding copepods tend to occur in tropical oligotrophic gyres. Similar continuums of copepod functional traits and environmental niches were found at smaller scales (references in Table 1) or globally (Brun et al., 2016).

Towards meaningful zooplankton FGs in marine ecosystem models

Our combination of analyses supports the view that planktonic copepods display a continuum of functional traits with a strong latitudinal gradient (Fig. 3 and 4; Appendix S9), driven by global gradients in abiotic conditions. Larger and myelinated active feeders present higher metabolic rates as well as higher feeding, excretion and mortality rates (Kiørboe 2011a; Lenz 2012; Litchman et al., 2013; Kiørboe & Hirst, 2014; Stamieszkin et al., 2015; van Someren Gréve, Almeda, & Kiørboe, 2017; Brun et al., 2019). A similar continuum can occur within a trophic group: larger current-feeding phytoplankton grazers (e.g., FG6 and 7) can feed on larger cells than their smaller congeners (FG1 and 11; Hansen et al., 1994), or larger active-feeding predators (FG5 and 8) can exert a top-down control on smaller ambush-feeding predators (FG4) or smaller zooplankton in general. Consequently, such a continuum should be explicitly represented in regional and global marine ecosystem models (e.g. similar to Serra-Pompei et al., 2020), which too often rely on a few size classes only (Le Quéré et al., 2005; Sailley et al., 2013). However, with a few exceptions (Henson et al., 2021), global marine ecosystem models do not yet routinely include >10 zooplankton FGs due to computational constraints. Since these models aim to represent biomass dynamics (in units of carbon) between food web components and their interactions with climate, their improvement should first focus on including copepod FGs that contribute the most to community biomass, and biogeochemical function (Le Quéré et al. 2005).

To try to rank the present FGs based on their contribution to total copepod community abundance in the upper layers of the global ocean, we implemented a preliminary synthesis of copepod abundance observations from various large-scale data sources (see Appendix S16). When examining the relative contribution of the eleven FGs to mean annual community abundance (Fig. S16b), three FGs emerge as the most abundant regardless of latitude and sampling gears: FG10 (Oithonids; small ambush-feeding omnivores), FG2 (Oncaeids; small cruise-feeding detritivores) and FG1 (*Clausocalanus* spp.; small cruise-feeding omnivores-herbivores). FG11 (*Acartia* spp. and *Centropages* spp.; small omnivorous mixed-feeders) also contributes substantially to total copepod abundance in the northern Atlantic and Pacific Oceans. Consequently, we encourage modelers to represent those three FGs more explicitly in future global ecosystem models. Some additional groups show a higher relative abundance on a regional scale, like FG4 in the tropics, or FG7 and FG9 in the Southern Ocean, which may make these groups important for regional ecosystem modeling efforts. The remaining FGs (FG3, 5, 6 and 8) display very low contributions to mean annual abundance, regardless of the latitude and the sampling gear. Such groups might be of a lower priority for improving the representation of zooplankton in marine ecosystem models. Nonetheless, these four FGs are

characterized by larger body sizes than FG1, 2 or 10 (Table S4.2) so their actual contribution to biomass production could be substantially higher, especially FG6 which gathers the largest filter-feeding Calanidae that are several millimeters longer than the species in FG1, FG2 or FG10. Future work will help model the contribution of the present copepod FGs to zooplankton biomass, and to clarify their relative priorities for inclusion in marine ecosystem models.

4.2. Why do copepod functional traits show contrasting biogeographic patterns in the ocean?

We defined four main ocean regions according to their similarity in community-level functional trait expression (Fig. 4a, Appendices S9 and S15). The first-order pattern corresponded to the separation between those regions poleward to ~45° latitude (regions 3 and 4) and those equatorward to ~45° latitude (regions 1 and 2; Appendix S10). Then, two key second-order patterns emerged: the regions south of the Antarctic Polar Front (region 4) showed a community trait composition distinct than those of the other high latitudes (e.g., the Arctic Ocean or the North Pacific Ocean), and the oligotrophic tropical gyres (region 2) were separated from the boundary upwelling and the equatorial current systems (region 1). The largest dissimilarity in community trait expression was found between regions 2 and 4 (Fig. 4b-m) and was driven by differences in CWM body size, myelination, carnivory and ambush-feeding vs. current-feeding (Appendices S9, S10 and S15). These trait patterns emerge from environment-based distribution models, so they reflect combinations of environmental conditions, but they also reflect biological processes that lead to the selection of trade-offs between traits through abiotic and/or biotic filtering. In other words, certain trait combinations are more competitive than others under varying conditions of temperature or food availability because of physiological constraints or because they lead to lower mortality rates (Litchman et al., 2013; Barton et al., 2013; van Someren Gréve et al., 2017; McGinty et al., 2021).

Median copepod body size decreased from the poles to the equator with a slight increase in upwelling systems. Such a pattern has been documented by other studies and is primarily driven by the strong negative relationship between the body size of marine ectotherms and temperature, according to Bergmann's rule (Brun et al., 2016; McGinty et al., 2018; Evans et al., 2020; Brandão et al., 2021; Campbell et al., 2021). The precise processes underlying Bergmann's rule remain debated, but it is likely that warmer temperatures (or a factor confounded with temperature) decrease growth efficiency and/or promote the maturation of adults at smaller body sizes (Atkinson 1994; Isla et al., 2008). Dissolved oxygen concentration could also limit maximal body size in marine ectotherms under the "oxygen

hypothesis" (Audzijonyte et al., 2019), but so far temperature outperformed oxygen in explaining variations in body size structure (Campbell et al., 2021).

Bergmann's rule occurs at the intraspecific level and at the interspecific level, meaning that warming-induced decreases in median body size can emerge if smaller clades replace larger ones along a latitudinal gradient. Such a turnover in size classes has been observed as well (Evans et al., 2020; Brandão et al., 2021) and is supported by our SDM projections. Indeed, functional traits that represent shifts in clade composition (i.e., those that show strong taxonomic clustering) such as myelination, spawning strategy, carnivory or feeding modes displayed latitudinal gradients that are positively, or negatively, collinear with the body size gradient (Appendices S9 and S10). The latitudinal patterns of these functional traits likely result from changes in food availability, quality, and predation pressure that are known to vary greatly from polar regions to tropical gyres (Brun et al., 2016; Horne et al., 2016; van Someren Gréve et al., 2017; Roy, 2018).

Lipid-rich myelin sheaths enable a faster conduction of nerve responses. This promotes faster reaction times and thus more efficient feeding or escape behaviors (Lenz, 2012). Myelin sheaths are cholesterol-rich so they require larger metabolic investments of dietary lipids (Lenz, 2012), which helps explain why copepod communities show larger proportion of small amyelinated taxa in tropical gyres where smaller lipid-poor phytoplankton dominate (Roy, 2018). Therefore, it is not surprising that the proportions of myelinated species follow the same spatial pattern as body size and current-feeding (Fig. 4b,c,j,m) and peak in productive environments characterized by larger and lipid-rich plankton (Roy 2018).

Similarly, sac-spawning is a more energy-conservative spawning strategy as it reduces egg-mortality at the cost of fecundity and hatching speed. The increased proportions of sac-spawners in tropical gyres (Fig. 4d) could reflect an adaptation to limited food availability (Kiørboe & Sabatini, 1994; Barton et al., 2013) and higher rates of carnivory (Fig. 4e; Woodd-Walker et al., 2002) and egg cannibalism among copepods (Ohman & Hirche, 2001; Segers & Taborsky, 2011). Furthermore, ambush-feeding is a passive feeding mode that lowers predation risk and energy costs compared to current- and cruise-feeding, but at the expense of feeding efficiency (Kiørboe, 2011a; van Someren Gréve et al., 2017). Consequently, the increased proportion of ambush-feeders in region 2 (Fig. 4i) should also result from trade-offs in functional traits expression driven by abiotic and biotic filtering (Litchman et al., 2013), as oligotrophic gyres seem to promote food-webs with increased carnivorous predation (Fig. 4e) and resource competition (Wood-Walker et al., 2002; Prowe, Visser, Andersen, Chiba, & Kiørboe, 2019).

However, our approach is based on presence data and habitat suitability indices rather than abundances, which likely underestimates the contribution of very abundant ambush-feeding species like *Oithona similis* at high latitudes (Gallienne & Robins, 2001; Pinkerton et al., 2010; Prowe, Visser, Andersen, Chiba, & Kiørboe, 2019). As a result, we probably underestimate the proportion of ambush-feeders in regions such as the Southern Ocean compared to Prowe et al. (2018), although these authors discarded other ambush-feeding copepods such as the Corycaeidae (Benedetti et al., 2016; Brun et al., 2017). Together, these elements support the fact that our modelled zooplankton functional traits patterns emerge from interactions between environmental conditions and the relative fitness (i.e., trade-offs) resulting from different trait combinations.

The climatological conditions used to train the SDMs are influenced by large scale oceanic currents as highlighted by our regionalization analysis (Fig. 4). We found the boundaries of the four main regions to overlap with the trajectories of well-known boundary currents as well as equatorial counter currents and fronts. Unfortunately, we cannot disentangle the effects of dispersal by currents on the observed community trait expression (and their modelled spatial patterns) from the effects of gradients in temperature and/or productivity, which select species based on their physiological requirements. Although ocean basin connectivity is rather high at the scale of our study (Jönsson & Watson, 2016), dispersal by marine currents and mesoscale processes are known to impact plankton community structure (Richter et al., 2020; Sommeria-Klein et al., 2021). For instance, the Antarctic Polar Front seems to be the main delimiter between regions 3 and 4, which coincides with the view that it imposes a strong physical barrier on passively drifting zooplankton (Murphy et al., 2021). Yet, recent evidence showed that many epipelagic plankton can cross Southern Ocean fronts thanks to the very dynamic meandering eddies (Murphy et al., 2021), which could explain the deviations of the region 4 boundaries from the Antarctic Polar Front. Therefore, the gradient in community trait expression separating regions 3 and 4 may also stem from the latitudinal gradient in temperature and biogeochemical conditions that select species and their traits based on physiological and metabolic constraints. Similarly, the gradients in CWM traits (i.e., the poleward succession of regions 2, 1 and 3) modelled in the North Pacific and North Atlantic Oceans could be driven by the temperature gradient or the presence of western boundary currents (e.g., the Gulf Stream and the North pacific Current) which carry southern warm-water communities eastward while they are getting mixed with northern cold-water communities. Such processes could explain why the CWM trait distribution of region 1 lies in between those of regions 2 and 3 (Fig. 4). This uncertainty calls for more studies combining

distribution models with dispersal rates and limitations (D'Amen et al., 2018; Shipley et al., 2021) to study the link between zooplankton functional traits and the "seascape" (Sommeria-Klein et al., 2021; Richter et al., 2020).

4.3. Caveats and future directions

Our results are sensitive to the three main steps of our framework: (i) the way the species were positioned in a functional space and how the FGs were defined from the latter, (ii) the quantity and quality of the functional trait data considered, and (iii) the SDMs chosen to generate the CWM traits values. We carefully investigated the sensitivity of our FGs definition to alternative choices made in our clustering approach (Appendix S4). Most of the comparable studies working with functional trait spaces rely on a Gower distance matrix and a principal coordinate analysis (PCoA; Legendre & Legendre, 2012) to ordinate taxa as a function of their trait combinations (see the synthesis by Mouillot et al., 2021). Here, we relied on an alternative dimensionality reduction analysis (i.e., the FAMD) to avoid negative eigenvalues (i.e., imaginary dimensions) and the relatively low levels of explained variance that are often inherent to the use of a PCoA (Legendre & Legendre, 2012; Mouillot et al., 2021). We assessed how using this alternative dimension reduction analysis affected the quality of the functional trait space by computing the AUC criterion recommended by Mouillot et al. (2021). We found an AUC value of 0.81, which is substantially higher than the recommended 0.7 threshold and indicates a "high quality trait space" according to these authors. In addition, the first four PCs of our FAMD explained nearly 80% of the variance in species traits, which reinforces our confidence in the quality of our functional trait space.

Beyond the dimensionality reduction step, the choice of the distance matrix and aggregation link could have been determining factors in drawing the functional dendrogram and thus defining the FGs (Fig. 1). We evaluated the similarity between the functional dendrograms emerging from alternative clustering approaches and found that they were all highly positively correlated (mean Baker's Gamma correlation coefficient was 0.75 ± 0.12; Appendix S4). This indicates that all dendrograms displayed very similar structure and explains how these dendrograms would have led to similar FGs composition. Ultimately, the approach that combined the FAMD, an Euclidean distance matrix, and Ward's aggregation linkage was chosen as our standard approach, since it provided ecologically meaningful FGs (i.e., groups that are neither too large and functionally heterogeneous nor too numerous and functionally redundant).

The quantity and the quality of the species-level functional trait information considered for our study largely determined the FGs and their ecological meaning. The traits chosen for this study only cover a fraction of the traits mentioned in the literature (Litchman et al., 2013; Brun et al., 2017; Appendix S3). Therefore, our study is likely underestimating the true diversity of copepod functions and FGs present in the ocean. However, it could also be that the present functional groups are quite representative of natural copepod communities and that adding further traits would only subdivide those FGs into smaller and functionally homogeneous ones. The trait compilation of Brun et al. (2017) included the following additional functional traits: growth rate, clearance rate, ingestion rate, egg diameter and production, or the production of resting eggs and diapausing stages throughout the life cycle. Yet, the availability and the coverage of those traits remained too poor for the wide pool of species studied here (Appendix S3), reflecting the historic measurement biases that largely focused on larger Calanoida species such as *Calanus* spp. Since a broad range of metabolic and physiological traits scale allometrically with body size (Kiørboe & Hirst, 2014), additional traits could be inferred based on body size measurements and phylogenetic distances (Molina-Velegas et al., 2018). It also implies that the present FGs characterized by larger body sizes (e.g., FG5 and 6) should display larger metabolic rates. Further lab-based and observation-based are needed to uncover the functional trait expression for a larger range of copepod species and families.

Differences between model projections are the main source of uncertainty in species community composition and diversity under a SDMs ensemble approach (Benedetti, Guilhaumon, Adloff, & Ayata, 2018; Diniz-Filho et al., 2009). We found some regional differences between the mean annual HSI projections of the SDMs. When considering all copepod species together, the GLM estimated higher mean HSI than the two other model types for the Southern Ocean (Appendix S8). Such variability could arise from how the various SDMs cope with limited predictors and species occurrence availability in winter conditions at the very high latitudes. Indeed, the copepod occurrence data (Appendix S1) and the climatological satellite observations (e.g., PAR and logChl) have poor coverage at latitudes >60° in winter. The sampling effort in the Southern Ocean is slightly lower than the global average between April and October and is extremely low between May and September (Appendix S1), and satellite-based PAR and chlorophyll-a measurements are not available for winter months due to the low light levels at high latitudes. Therefore, our mean annual projections for the polar oceans are likely more representative of summer conditions. Most GLMs and GAMs included one or both of these satellite-based climatologies as predictors

(Appendix S6). Yet only GLMs showed these relatively high HSI levels in the Southern Ocean. Plus, SST was found to be the most influential predictor in both SDMs types (Appendix S6). This implies that this discrepancy between GLMs and GAMs was not driven by predictor selection but rather by how the responses of the species to SST and logChla/PAR gradients were captured by those two types of models (Merow et al., 2014). Here, the less complex response curves of the GLMs lead to higher average HSI at very cold temperatures compared to GAMs. The HSI patterns obtained from the GAM and ANN were closer to the latitudinal diversity gradients that were previously observed and modelled for marine ectotherms, which seldom show increases in richness or habitat suitability towards the poles (Tittensor et al., 2010; Benedetti et al., 2021). Consequently, we have more confidence in the GAM-based and ANN-based annual HSI estimates that in the GLM-based ones.

Interestingly, inter-SDMs variability was much lower when looking at CWM traits projections rather than HSI patterns (Appendix S8). All three SDMs lead to very similar spatial patterns in CWM traits and the amplitude of these CWM values differed only slightly. The regions displaying higher CWM traits variability depended on the trait considered, and variability remained quite low overall (but see Appendix S8). Therefore, the present CWM traits patterns display relatively low uncertainty compared to the HSI patterns. The fact that CWM body size and CWM myelination show very similar spatial patterns and ranges to the respective estimates of Brun et al. (2016) gives us further confidence in our approach and spatial projections.

To conclude, we recommend the inclusion of multiple functional groups of copepods in marine ecosystem models, based on our functional dendrogram (Fig. 1), with a level of complexity of zooplankton representation that should be modulated by the scientific question, the FGs dominating community biomass at the scale of the study region(s), the traits of interest, and computing efficiency constraints. Ideally, future global marine ecosystems, that cannot efficiently include 10 zooplankton groups, will include a few (3-4) FGs that cover the main gradients observed in trait space (Fig. 1; Appendix S12), niche space (Fig. 2), geographical space (Figs. 3, 4; Appendices S9, S10) and that represent important fractions of biomass (Appendix S16). This would divide copepods into a few but ecologically meaningful groups with clearly distinct ecological niches and also functional trait characteristics. For other ecological applications or to represent food-web interactions across trophic levels (Ward et al., 2012; Serra-Pompei et al., 2020), these groups could be split into further FGs with specific characteristics (e.g., size classes, high lipid content, grazing and mortality rates etc.). To achieve this, future studies are required to improve the coverage of functional traits data

towards more taxonomic groups and more quantitative traits (Appendix S3). Ongoing compilations of zooplankton groups biomass data will also help rank the importance of the present FGs worldwide and across regions. Ultimately, the present results and future data compilations will help us study global patterns of zooplankton functional diversity (Mouillot et al., 2013), which may link more closely to ecosystem function and service provision than taxonomic diversity. Marine ecosystems will experience increasingly stressful conditions in the next century due to anthropogenic climate change (IPCC, 2021). This will strongly restructure zooplankton community richness and composition as species migrate poleward to track suitable habitats (Benedetti et al., 2021). How these changes will impact functional trait expression and functional diversity at the community-level remains too poorly understood. The present study is thus a key step towards our understanding of how future climate change will reshape the expression of marine functional diversity worldwide.

Data Availability Statement

The copepod species occurrences data used to train the species distribution models are publicly available on Zenodo (10.5281/zenodo.5101349). The newly implemented species-level functional trait table is available as Table S2 and will be made available through an open access repository upon publication of our study. Any computer code used to generate the results of the study are freely available upon request to the authors and all R codes are currently stored on the GitHub account of J.W. (https://github.com/jonas-wydler).

References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spthin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541–545.

Atkinson, D. (1994). Temperature and Organism Size—A Biological Law for Ectotherms? In M. Begon & A. H. Fitter (Eds.), *Advances in Ecological Research* (Vol. 25, pp. 1-58): Academic Press.

Audigier, V., Husson, F., & Josse, J. (2016). A principal component method to impute missing values for mixed data. *Advances in Data Analysis and Classification*, 10(1), 5–26.

Audzijonyte, A., Barneche, D. R., Baudron, A. R., Belmaker, J., Clark, T. D., Marshall, C. T., . . . van Rijn, I. (2019). Is oxygen limitation in warming waters a valid mechanism to explain decreased body sizes in aquatic ectotherms? *Global Ecology and Biogeography*, 28(2), 64-77. doi:10.1111/geb.12847

Baker, F. B. (1974). Stability of two hierarchical grouping techniques case i: sensitivity to data errors. *Journal of the American Statistical Association*, 69(346), 440– 445.

Barnett, A. J., Finlay, K., & Beisner, B. E. (2007). Functional diversity of crustacean zooplankton communities: towards a trait-based classification. *Freshwater Biology*, 52(5), 796–813.

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., . . . Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, 16(4), 522–534. doi: 10.1111/ele.12063

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. doi:10.1073/pnas.0913855107

Becker, É. C., Mazzocchi, M. G., de Macedo-Soares, L. C. P., Costa Brandão, M., & Santarosa Freire, A. (2021). Latitudinal gradient of copepod functional diversity in the South Atlantic Ocean. *Progress in Oceanography*, 199, 102710. doi:https://doi.org/10.1016/j.pocean.2021.102710

Benedetti, F., Gasparini, S., & Ayata, S.-D. (2016). Identifying copepod functional groups from species functional traits. *Journal of Plankton Research*, 38(1), 159-166. doi:10.1093/plankt/fbv096

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. doi: 10.1111/ecog.02434

Benedetti, F., Vogt, M., Righetti, D., Guilhaumon, F., & Ayata, S.-D. (2018). Do functional groups of planktonic copepods differ in their ecological niches? *Journal of Biogeography*, 45(3), 604–616. doi: 10.1111/jbi.13166

Benedetti, F., Vogt, M., Elizondo, U. H., Righetti, D., Zimmermann, N. E., & Gruber, N. (2021). Major restructuring of marine plankton assemblages under global warming. *Nature Communications*, 12(1), 5226. doi:10.1038/s41467-021-25385-x

Boyer, T.P., J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T.D. O'Brien, C.R. Paver, J.R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, 2013: World Ocean Database 2013, NOAA Atlas NESDIS 72, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp., http://doi.org/10.7289/V5NZ85MT

Bron, J. E., Frisch, D., Goetze, E., Johnson, S. C., Lee, C. E., & Wyngaard, G. A. (2011). Observing copepods through a genomic lens. *Frontiers in Zoology*, 8(1), 1–15. https://doi.org/10.1186/1742-9994-8-22

Brun, P., Payne, M. R., & Kiørboe, T. (2016). Trait biogeography of marine copepods–an analysis across scales. *Ecology Letters*, 19(12), 1403–1413. doi: 10.1111/ele.12688

Brun, P., Payne, M. R., & Kiørboe, T. (2017). A trait database for marine copepods. *Earth System Science Data*, 9(1), 99–113. doi:10.5194/essd-2016-30

Brun, P., Stamieszkin, K., Visser, A. W., Licandro, P., Payne, M. R., & Kiørboe, T. (2019). Climate change has altered zooplankton-fuelled carbon export in the North Atlantic. *Nature Ecology & Evolution*, 3(3), 416–423. https://doi.org/10.1038/s41559-018-0780-3

Campbell, M. D., Schoeman, D. S., Venables, W., Abu-Alhaija, R., Batten, S. D., Chiba, S., . . . Richardson, A. J. (2021). Testing Bergmann's rule in marine copepods. *Ecography*, 44(9), 1283-1295. doi:https://doi.org/10.1111/ecog.05545

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to algorithms. *MIT and McGraw*.

D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2018). Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography*, 41(8), 1233-1244. doi:https://doi.org/10.1111/ecog.03148

Diniz-Filho, J. A. F., Mauricio Bini, L., Fernando Rangel, T.,Loyola, R. D., Hof, C., Nogués-Bravo, D., & Araújo, M. B. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, 32(6), 897–906. doi: 10.1111/j.1600-0587.2009.06196.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., . . . others (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Flynn, K. J., St John, M., Raven, J. A., Skibinski, D. O. F., Allen, J. I., Mitra, A., & Hofmann, E. E. (2015). Acclimation, adaptation, traits and trade-offs in plankton functional type models: reconciling terminology for biology and modelling. Journal of Plankton Research, 37(4), 683-691. doi:10.1093/plankt/fbv036

Gallienne, C., & Robins, D. (2001). Is *Oithona* the most important copepod in the world's oceans? *Journal of Plankton Research*, 23(12), 1421–1432. https://doi.org/10.1093/plankt/23.12.1421

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). Habitat suitability and distribution models: with applications in R. Cambridge University Press.

Hansen, B., Bjornsen, P. K., & Hansen, P. J. (1994). The size ratio between planktonic predators and their prey. *Limnology & Oceanography*, 39(2), 395–403.

Hébert, M.-P., Beisner, B. E., & Maranger, R. (2017). Linking zooplankton communities to ecosystem functioning: toward an effect-trait framework. *Journal of Plankton Research*, 39(1), 3–12. doi:10.1093/plankt/fbw068

Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, 12(1), 5372. doi:10.1038/s41467-021-25699-w

Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3), 679-688. doi:https://doi.org/10.1890/11-0826.1

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. doi: 10.1111/geb.12460

Isla, J. A., Lengfellner, K., & Sommer, U. (2008). Physiological response of the copepod *Pseudocalanus* sp. in the Baltic Sea at different thermal scenarios. *Global Change Biology*, 14(4), 895-906. doi:https://doi.org/10.1111/j.1365-2486.2008.01531.x

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

Jónasdóttir, S. H., Visser, A. W., Richardson, K., & Heath, M. R. (2015). Seasonal copepod lipid pump promotes carbon sequestration in the deep North Atlantic. *Proceedings of the National Academy of Sciences, 112*(39), 12122-12126. doi:10.1073/pnas.1512110112

Jönsson, B. F., & Watson, J. R. (2016). The timescales of global surface-ocean connectivity. *Nature Communications*, 7, 11239. doi:10.1038/ncomms11239

Kenitz, K. M., Visser, A. W., Mariani, P., & Andersen, K. H. (2017). Seasonal succession in zooplankton feeding traits reveals trophic trait coupling. *Limnology & Oceanography*, 62(3), 1184–1197.

Kiørboe, T. (2011a). What makes pelagic copepods so successful? *Journal of Plankton Research*, 33(5), 677– 685. doi:10.1093/plankt/fbq159

Kiørboe, T. (2011b). How zooplankton feed: mechanisms, traits and trade-offs. *Biological Reviews*, 86(2), 311– 339. doi: 10.1111/j.1469-185X.2010.00148.x

Kiørboe, T., & Hirst, A. G. (2014). Shifts in mass scaling of respiration, feeding, and growth rates across life-form transitions in marine pelagic organisms. *The American Naturalist*, 183(4), E118–E130.

Kiørboe, T., & Sabatini, M. (1994). Reproductive and life cycle strategies in egg-carrying cyclopoid and free-spawning calanoid copepods. *Journal of Plankton Research*, 16(10), 1353–1366.

Legendre, P., & Legendre, L. (2012). Numerical Ecology. Elsevier.

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., . . . others (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11), 2016–2040. doi: 10.1111/j.1365-2486.2005.01004.x

Le Quéré, C., Buitenhuis, E. T., Moriarty, R., Alvain, S., Aumont, O., Bopp, L., . . . Vallina, S. M. (2016). Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles. *Biogeosciences*, 13(14), 4111-4133. doi:10.5194/bg-13-4111-2016

Lenz, P. H. (2012). The biogeography and ecology of myelin in marine copepods. *Journal of Plankton Research*, 34(7), 575–589. doi:10.1093/plankt/fbs037

Litchman, E., Ohman, M. D., & Kiørboe, T. (2013). Trait-based approaches to zooplankton communities. *Journal of Plankton Research*, 35(3), 473–484. doi:10.1093/plankt/fbt019

Longhurst, A. R. (2010). Ecological geography of the sea. Elsevier.

McGinty, N., Barton, A. D., Record, N. R., Finkel, Z. V., & Irwin, A. J. (2018). Traits structure copepod niches in the North Atlantic and Southern Ocean. *Marine Ecology Progress Series*, 601, 109-126. https://doi.org/10.3354/meps12660

McGinty, N., Barton, A. D., Record, N. R., Finkel, Z. V., Johns, D. G., Stock, C. A., & Irwin, A. J. (2021). Anthropogenic climate change impacts on copepod trait biogeography. *Global Change Biology*, *27*(7), 1431-1442. doi:https://doi.org/10.1111/gcb.15499

Merow, C., Smith, M. J., Edwards Jr, T. C., Guisan, A., McMahon, S. M., Normand, S., . . . Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37(12), 1267–1281. doi: 10.1111/ecog.00845

Molina-Venegas, R., Moreno-Saiz, J. C., Castro Parga, I., Davies, T. J., Peres-Neto, P. R., & Rodríguez, M. Á. (2018). Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography*, 41(10), 1740-1749. doi:https://doi.org/10.1111/ecog.03480

Mouchet, M., Guilhaumon, F., Villéger, S., Mason, N. W., Tomasini, J.-A., & Mouillot, D. (2008). Towards a consensus for calculating dendrogram-based functional diversity indices. *Oikos*, 117(5), 794–800. doi: 10.1111/j.2008.0030-1299.16594.x

Mouillot, D., Graham, N. A., Villéger, S., Mason, N. W., & Bellwood, D. R. (2013). A functional approach reveals community responses to disturbances. *Trends in Ecology & Evolution*, 28(3), 167-177. doi: 10.1016/j.tree.2012.10.004

Mouillot, D., Loiseau, N., Grenié, M., Algar, A. C., Allegra, M., Cadotte, M. W., . . . Auber, A. (2021). The dimensionality and structure of species trait spaces. *Ecology Letters*, 24(9), 1988-2009. doi:https://doi.org/10.1111/ele.13778

Murphy, E. J., Johnston, N. M., Hofmann, E. E., Phillips, R. A., Jackson, J. A., Constable, A. J., . . . Xavier, J. C. (2021). Global Connectivity of Southern Ocean Ecosystems. *Frontiers in Ecology and Evolution*, 9(454). doi:10.3389/fevo.2021.624451

Ohman, M. D., & Hirche, H. J. (2001). Density-dependent mortality in an oceanic copepod population. *Nature*, *412*(6847), 638-641. doi:10.1038/35088068

Ohman, M. D., & Romagnan, J.-B. (2016). Nonlinear effects of body size and optical attenuation on diel vertical migration by zooplankton. *Limnology & Oceanography*, 61(2), 765–770. doi: 10.1002/lno.10251

Orsi, A., & Harris, U. (2019). Fronts of the Antarctic circumpolar current–GIS data, ver. 1.

Pagès, J. (2004). Analyse factorielle de données mixtes: principe et exemple d'application. Montpellier SupAgro, http://www. agro-montpellier. fr/sfds/CD/textes/pages1.pdf

Pecuchet, L., Reygondeau, G., Cheung, W. W. L., Licandro, P., van Denderen, P. D., Payne, M. R., & Lindegren, M. (2018). Spatial distribution of life-history traits and their response to environmental gradients across multiple marine taxa. *Ecosphere*, 9(10), e02460. doi:https://doi.org/10.1002/ecs2.2460

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. *Ecological Applications*, 19(1), 181–197.

Pidwirny, M. (2006). Surface and subsurface ocean currents: Ocean current map. fundamentals of physical geography, 2nd edition. Retrieved 2021-08.20, from http://www.physicalgeography.net/ fundamentals/8q_1.html

Pinkerton, M. H., Smith, A. N. H., Raymond, B., Hosie, G. W., Sharp, B., Leathwick, J. R., & Bradford-Grieve, J. M. (2010). Spatial and seasonal distribution of adult Oithona similis in the Southern Ocean: Predictions using boosted regression trees. *Deep Sea Research Part I: Oceanographic Research Papers*, 57(4), 469-485. doi:https://doi.org/10.1016/j.dsr.2009.12.010

Pomerleau, C., Sastri, A. R., & Beisner, B. E. (2015). Evaluation of functional trait diversity for marine zooplankton communities in the northeast subarctic pacific ocean. *Journal of Plankton Research*, 37(4), 712–726. doi:10.1093/plankt/fbv045

Prowe, A. E. F., Visser, A. W., Andersen, K. H., Chiba, S., & Kiørboe, T. (2019). Biogeography of zooplankton feeding strategy. *Limnology & Oceanography*, 64(2), 661–678. doi: 10.1002/lno.11067

Qiu, B., & Chen, S. (2004). Seasonal modulations in the eddy field of the south pacific ocean. *Journal of Physical Oceanography*, 34(7), 1515–1527.

Reygondeau, G., Guieu, C., Benedetti, F., Irisson, J.-O., Ayata, S.-D., Gasparini, S., & Koubbi, P. (2017). Biogeochemical regions of the Mediterranean Sea: An objective multidimensional and multivariate environmental approach. *Progress in Oceanography*, 151, 138-148. doi:https://doi.org/10.1016/j.pocean.2016.11.001

Richter, D. J., Watteaux, R., Vannier, T., Leconte, J., Frémont, P., Reygondeau, G., . . . Da Silva, O. (2020). Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *BioRxiv*, 867739.

Ricotta, C. (2005). A note on functional diversity measures. *Basic and Applied Ecology*, 6(5), 479–486.

Sailley, S., Vogt, M., Doney, S., Aita, M., Bopp, L., Buitenhuis, E., ... Yamanaka, Y. (2013). Comparing food web structures and dynamics across a suite of global marine ecosystem models. *Ecological Modelling*, 261, 43–57.

Segers, F. H. I. D., & Taborsky, B. (2011). Egg size and food abundance interactively affect juvenile growth and behaviour. *Functional Ecology*, 25(1), 166-176. doi:https://doi.org/10.1111/j.1365-2435.2010.01790.x

Serra-Pompei, C., Soudijn, F., Visser, A. W., Kiørboe, T., & Andersen, K. H. (2020). A general size- and trait-based model of plankton communities. *Progress in Oceanography*, 189, 102473. doi:https://doi.org/10.1016/j.pocean.2020.102473

Shipley, B. R., Bach, R., Do, Y., Strathearn, H., McGuire, J. L., & Dilkina, B. (2021). megaSDM: integrating dispersal and time-step analyses into species distribution models. *Ecography*, 44, 1-12. doi:https://doi.org/10.1111/ecog.05450

Sommeria-Klein, G., Watteaux, R., Ibarbalz Federico, M., Pierella Karlusich Juan, J., Iudicone, D., Bowler, C., & Morlon, H. (2021). Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science*, *374*(6567), 594-599. doi:10.1126/science.abb3717

Stamieszkin, K., Pershing, A. J., Record, N. R., Pilskaln, C. H., Dam, H. G., & Feinberg, L. R. (2015). Size as the master trait in modeled copepod fecal pellet carbon flux. *Limnology & Oceanography*, 60(6), 2090–2107. doi: 10.1002/lno.10156

Steinberg, D. K., & Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Annual Review of Marine Science*, 9, 413–444. doi: 10.1146/annurev-marine-010814-015924

Stocker, T. (2014). Climate change 2013: the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press.

Thuiller, W., Georges, D., Engler, R., Georges, M. D., & Breiner F. (2020). biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.4.6. https://CRAN.R-project.org/package=biomod2

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. doi:10.1038/nature09329

Turner, J. T. (2004). The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zoological Studies*, 43(2), 255–266.

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, 130, 205–248. http://dx.doi.org/10.1016/j.pocean.2014.08.005

van Someren Gréve, H., Almeda, R., & Kiørboe, T. (2017). Motile behavior and predation risk in planktonic copepods. *Limnology & Oceanography*, 62(5), 1810– 1824. doi: 10.1002/lno.10535

Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12), 2290–2299.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116(5), 882–892. doi: 10.1111/j.2007.0030-1299.15559.x

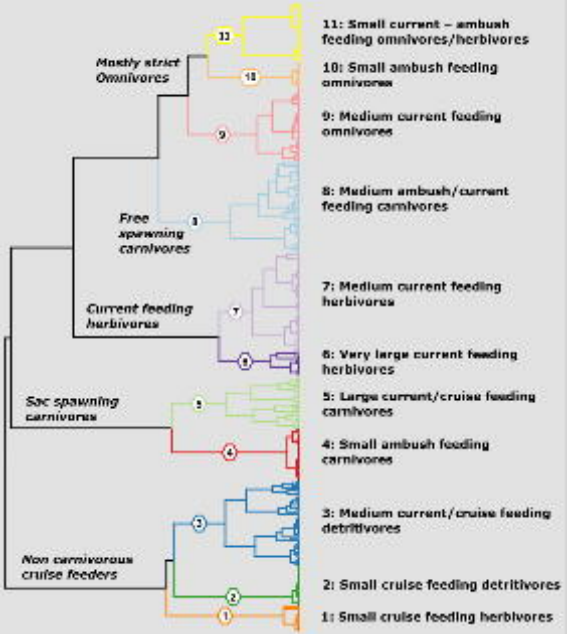
Visser, A. W., Grønning, J., & Jónasdóttir, S. H. (2017). *Calanus hyperboreus* and the lipid pump. *Limnology & Oceanography*, 62(3), 1155-1165. doi:10.1002/lno.10492

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.

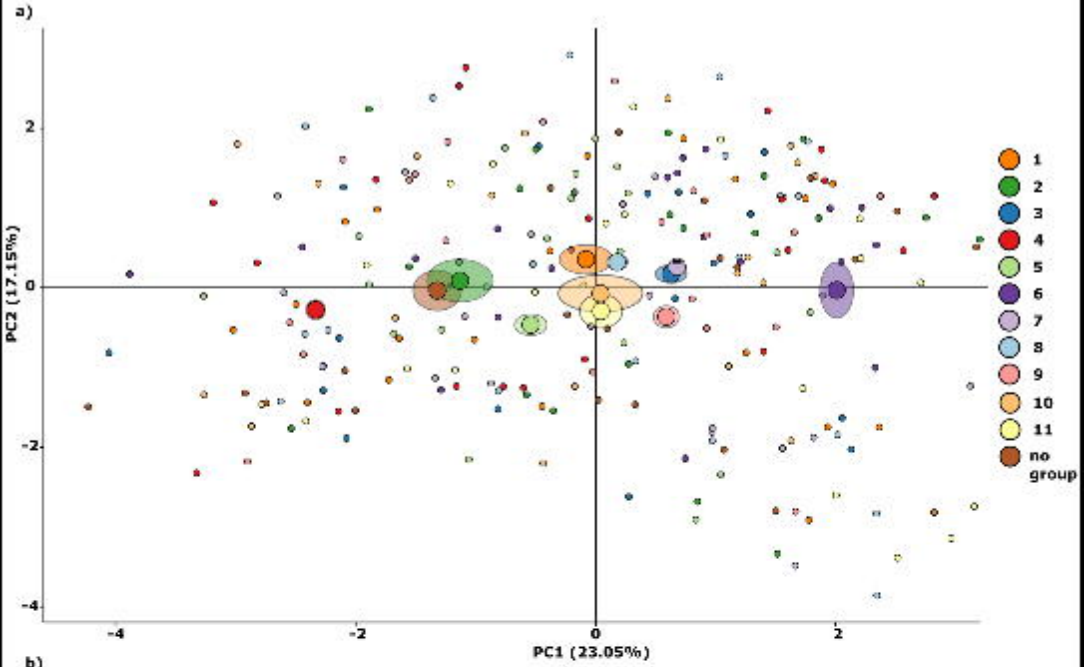
Biosketch

Fabio Benedetti is a postdoctoral researcher and **Meike Vogt** a senior research scientist in the Environmental Physics (UP) group of ETH Zürich. Both share broad interests in plankton biogeography and functional diversity and theirs links with biodiversity, ecosystem function and biogeochemical cycles in the global ocean. F.B. is a macroecologist specialized in trait-based approaches and plankton diversity modelling. M.V. is a marine ecosystem modeler specialized plankton functional types. **Jonas Wydler** has successfully completed his MSc degree in Environmental System Sciences at ETH Zürich, under the supervision of F.B. and M.V., which is the subject of this work.

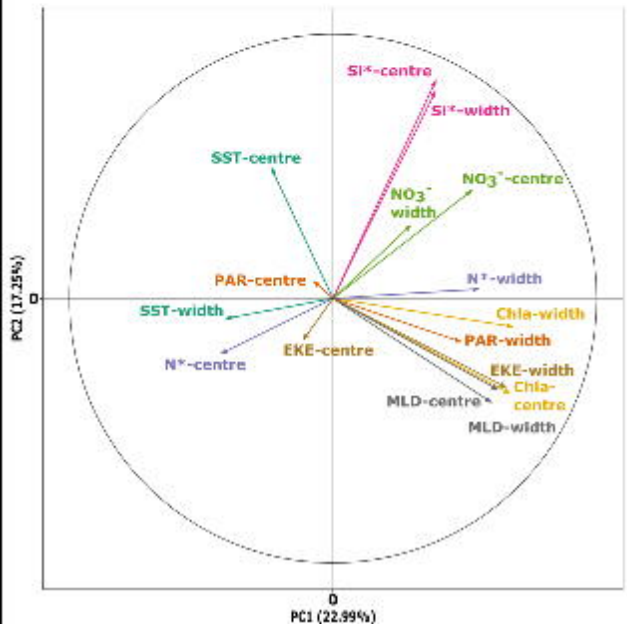
F.B. and M.V. co-designed the study and F.B. collated the data used in the analyses and provided expertise with regard to every methodology used. J.W. conducted the numerical analyses and wrote the first version of the manuscript under the supervision of F.B. and M.V. F.B. wrote the final version of the manuscript with input from both M.V. and J.W.

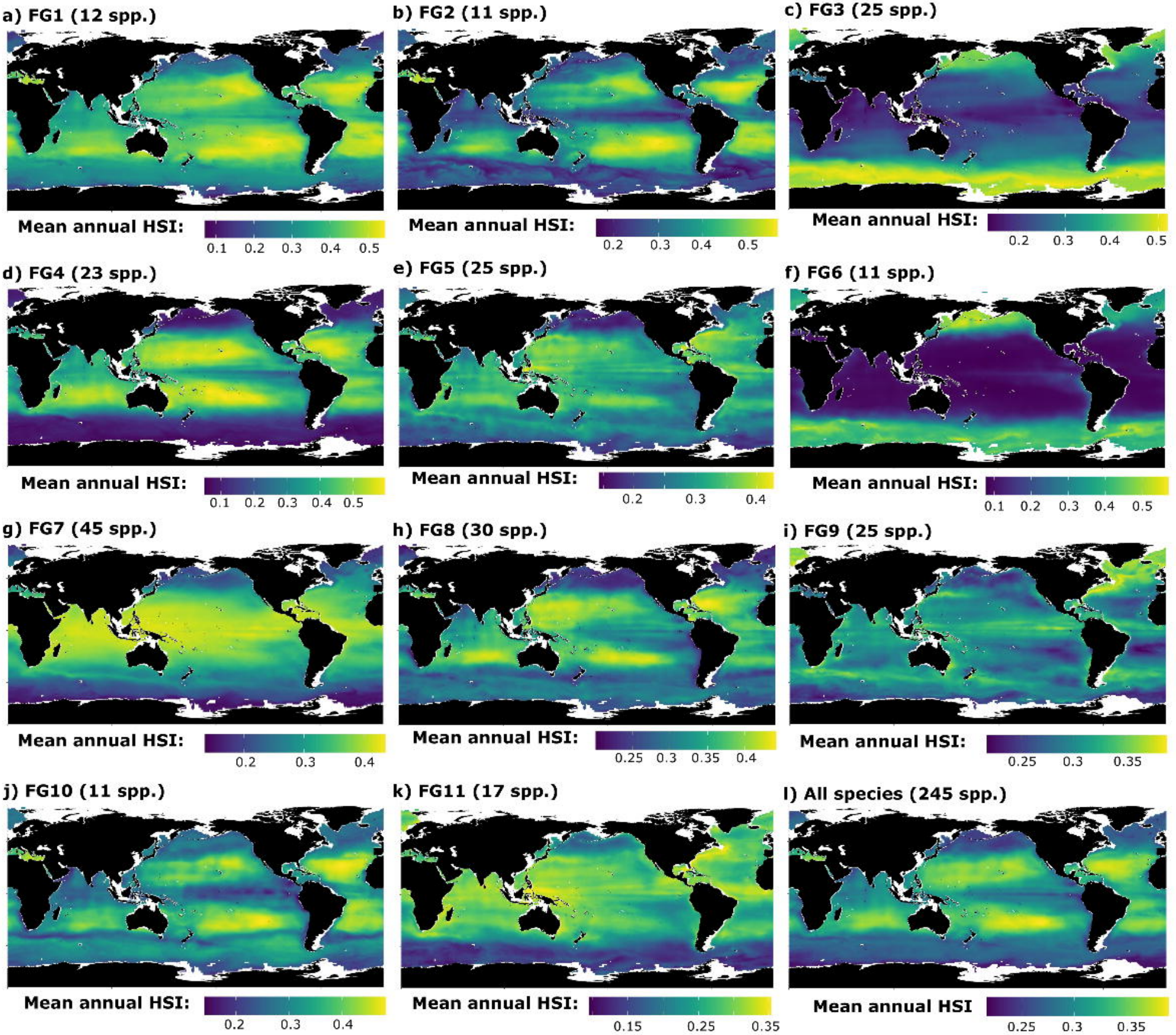

11: Small current – ambush feeding omnivores/herbivores
10: Small ambush feeding omnivores
9: Medium current feeding omnivores
8: Medium ambush/current feeding carnivores
7: Medium current feeding herbivores
6: Very large current feeding herbivores
5: Large current/cruise feeding carnivores
4: Small ambush feeding carnivores
3: Medium current/cruise feeding detritivores
2: Small cruise feeding detritivores
1: Small cruise feeding herbivores
Mostly strict Omnivores
Free spawning carnivores
Current feeding herbivores
Sac spawning carnivores
Non carnivorous cruise feeders
11
10
9
8
7
6
5
4
3
2
1

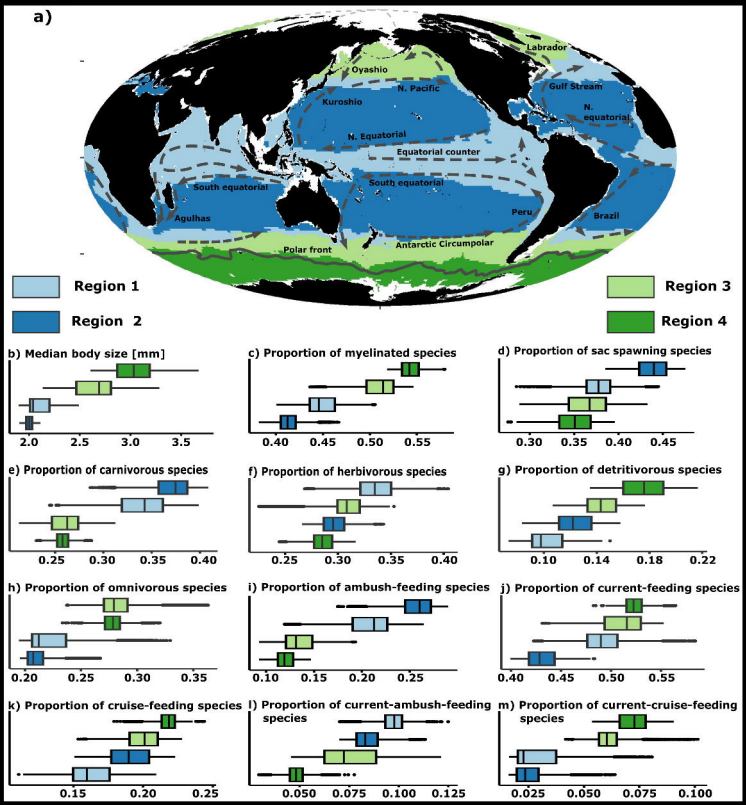
a)


PC2 (17.15%)
PC1 (23.05%)
2
0
-2
-4
-4
-2
0
2
1
2
3
4
5
6
7
8
9
10
11
no group

b)


PC2 (17.25%)
PC1 (22.99%)
0
0
Si*-centre
Si*-width
SST-centre
NO3*-centre
NO3*-width
PAR-centre
N*-width
SST-width
Chla-width
PAR-width
EKE-centre
N*-centre
EKE-width
Chla-centre
MLD-centre
MLD-width


a) FG1 (12 spp.)
Mean annual HSI:
0.1 0.2 0.3 0.4 0.5
b) FG2 (11 spp.)
Mean annual HSI:
0.2 0.3 0.4 0.5
c) FG3 (25 spp.)
Mean annual HSI:
0.2 0.3 0.4 0.5
d) FG4 (23 spp.)
Mean annual HSI:
0.1 0.2 0.3 0.4 0.5
e) FG5 (25 spp.)
Mean annual HSI:
0.2 0.3 0.4
f) FG6 (11 spp.)
Mean annual HSI:
0.1 0.2 0.3 0.4 0.5
g) FG7 (45 spp.)
Mean annual HSI:
0.2 0.3 0.4
h) FG8 (30 spp.)
Mean annual HSI:
0.25 0.3 0.35 0.4
i) FG9 (25 spp.)
Mean annual HSI:
0.25 0.3 0.35
j) FG10 (11 spp.)
Mean annual HSI:
0.2 0.3 0.4
k) FG11 (17 spp.)
Mean annual HSI:
0.15 0.25 0.35
l) All species (245 spp.)
Mean annual HSI
0.25 0.3 0.35


a)
Labrador
Oyashio
N. Pacific
Gulf Stream
Kuroshio
N. equatorial
N. Equatorial
Equatorial counter
South equatorial
South equatorial
Agulhas
Peru
Brazil
Antarctic Circumpolar
Polar front
Region 1
Region 2
Region 3
Region 4
b) Median body size [mm]
2.0 2.5 3.0 3.5
c) Proportion of myelinated species
0.40 0.45 0.50 0.55
d) Proportion of sac spawning species
0.30 0.35 0.40 0.45
e) Proportion of carnivorous species
0.25 0.30 0.35 0.40
f) Proportion of herbivorous species
0.25 0.30 0.35 0.40
g) Proportion of detritivorous species
0.10 0.14 0.18 0.22
h) Proportion of omnivorous species
0.20 0.25 0.30 0.35
i) Proportion of ambush-feeding species
0.10 0.15 0.20 0.25
j) Proportion of current-feeding species
0.40 0.45 0.50 0.55
k) Proportion of cruise-feeding species
0.15 0.20 0.25
l) Proportion of current-ambush-feeding species
0.050 0.075 0.100 0.125
m) Proportion of current-cruise-feeding species
0.025 0.050 0.075 0.100

This study	Main clade	Group definition	Pomerleau et al. (2015) - North Pacific Ocean	Benedetti et al. (2016) - Mediterranean Sea	Benedetti et al. (2018) - Mediterranean Sea	Becker et al. (2021) - South Atlantic Ocean
FG1	Clausocalanus	Small myelinated cruise-feeding omnivores-herbivores	Small current- and cruise-feeding omnivore-herbivores (Group 6)	Small cruise-feeding omnivores-herbivores (subset of Group 6)	Small cruise-feeding omnivores-herbivores (Group 7)	Small and large myelinated cruise- and current-feeding omnivores-herbivores (subset of Group B)
FG2	Oncaeidae	Small non myelinated cruise-feeding detritivores	No equivalent	Small cruise-feeding omnivores-detritivores (subset of Group 6)	Small sac-spawning detritivores (subset of Group 5)	Small amyelinated ambush- or cruise-feeding carnivores or detritivores (small subset of Group D)
FG3	Spinocalanus, Scaphocalanus, Metridia	Medium-sized myelinated mixed-feeding or cuise-feeding detritivores	No equivalent	Small cruise-feeding omnivores-detritivores (subset of Group 6)	Small sac-spawning detritivores (small subset of Group 5)	No equivalent
FG4	Corycaeidae	Small amyelinated ambush-feeding carnivores	No equivalent	Small ambush-feeding carnivores (Group 2)	Small ambush-feeding carnivores (Group 2)	Small amyelinated ambush- or cruise-feeding carnivores or detritivores (small subset of Group D)
FG5	Sapphirinidae, Euchaetidae	Large current- or cruise-feeding carnivores	No equivalent	Large cruise-feeding carnivores (Group 1)	Large cruise- or current-feeding feeding carnivores (Group 1)	Small and large myelinated cruise- and current-feeding omnivores-herbivores (subset of Group B)
FG6	Calanus, Eucalanus	Very large myelinated current-feeding omnivores-herbivores	No equivalent	No equivalent	No equivalent	No equivalent
FG7	Calanus, Paracalanus, Calocalanus	Medium-sized myelinated current-feeding omnivores-herbivores	Small and large current-feeding omnivores-herbivores (Group 5c)	Small and large current-feeding omnivores-herbivores (Group 4)	Small and large current-feeding omnivores-herbivores (Groups 3 and 4)	Large myelinated current-feeding omnivores-herbivores (subset of Group C)
FG8	Candacia, Haloptilus, Heterorhabdus	Medium-sized amyelinated ambush- or current-feeding carnivores	No equivalent	Large cruise-feeding carnivores (Group 1)	Large cruise- or current-feeding feeding carnivores (Group 1)	No equivalent
FG9	Pleuromamma, Gaetanus, Labidocera	Medium-sized amyelinated current-feeding omnivores	No equivalent	Small and large current-feeding omnivores-herbivores (subset of Group 4)	Small and large current-feeding omnivores-herbivores (small subset of Group 4)	No equivalent
FG10	Oithonidae	Small amyelinated ambush-feeding omnivores	No equivalent	Small ambush-feeding omnivores (Group 5)	Small ambush-feeding omnivores (Group 6)	Small amyelinated mixed- or ambush-feeding omnivores (Group A)
FG11	Acartia, Centropages	Small amyelinated mixed-feeding omnivores	Small ambush-feeding omnivores (subset of Group 4)	Small mixed-feeding omnivores (Group 3)	Small mixed-feeding omnivores (small subset of Group 4)	Small amyelinated mixed- or ambush-feeding omnivores (Group A)