



## Do functional groups of planktonic copepods differ in their ecological niches?

Fabio Benedetti, Meike Vogt, Damiano Righetti, Francois Guilhaumon, Sakina-Dorothee Ayata

### ► To cite this version:

Fabio Benedetti, Meike Vogt, Damiano Righetti, Francois Guilhaumon, Sakina-Dorothee Ayata. Do functional groups of planktonic copepods differ in their ecological niches?. *Journal of Biogeography*, 2018, 45 (3), pp.604–616. 10.1111/jbi.13166 . hal-01826722

**HAL Id: hal-01826722**

**<https://hal.science/hal-01826722>**

Submitted on 3 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Do functional groups of planktonic copepods differ in their ecological niches?

*Authors:* Fabio Benedetti<sup>1</sup>, Meike Vogt<sup>2</sup>, Damiano Righetti<sup>2</sup>, François Guilhaumon<sup>3</sup>, Sakina-Dorothee Ayata<sup>1</sup>

## *Affiliations:*

1. Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d’Océanographie de Villefranche sur mer (LOV), Observatoire Océanologique, 06230 Villefranche-sur-Mer, France.

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

3. IRD UMR 9190 MARBEC, IRD-CNRS-IFREMER-UM, Université de Montpellier, Montpellier 34095, France.

*Running title:* Copepod functional groups and their niches.

*Keywords:* Functional groups - Zooplankton - Environmental niche - Trait biogeography - Copepods.

## Acknowledgements

This work received the financial support of the MerMex (Marine Ecosystems Response in the Mediterranean Experiment)/ MISTRALS French National Program, through the PlankMed action of Work Package 5. F.B. received support from the Climate-KIC initiative of the European Institute of Innovation & Technology (EIT) through a PhD grant and an international mobility grant to work at ETH Zürich. The authors were also supported by the EuroMarine consortium through the PlankDiv and the PlankBioS workshops. They are very grateful to Philipp Brun and Pr. Nicklaus E. Zimmermann for their advice on the methods and Jean-Olivier Irisson for his help in implementing them. We thank Benjamin Randall and Richard Copley for revising the language. We thank Christine Meynard and two anonymous referees for their constructive comments.

## **Abstract**

### **Aim**

To assess the degree of overlap between the environmental niches of marine planktonic copepods and test if the distribution of copepod functional groups differs across environmental gradients.

### **Location**

The Mediterranean Sea.

### **Methods**

Functional groups were defined based on clustering of functional traits in 106 marine copepod species using a multivariate ordination analysis. Functional traits included maximum body length, feeding mode, spawning strategy and trophic group. Simultaneously, the global distribution of the species was used to model their environmental niches with six environmental variables. For each of these predictors, four niche parameters were derived from the univariate response curve of each species, to summarise their environmental preferences and ordinate the species in niche space through a PCA. Finally the differences in the position in niche space of functional groups were tested with variance analysis.

### **Results**

We identified seven copepod functional groups with different distributions along the environmental gradients covered by our study. While carnivorous functional groups were affiliated with oligotrophic and tropical conditions, large and small current-feeding herbivores are associated with colder, more seasonally-varying and productive conditions. Small cruising detritivores and other small current-feeding herbivores were not affiliated with specific conditions as their constituting species were scattered in niche space.

### **Main conclusions**

Since copepod functional groups occupy distinct ecological niches, ecosystem processes related to these groups are expected to vary across environmental gradients. Conditions favouring large current-feeding herbivores should allow for enhanced fluxes of energy and nutrients through Mediterranean Sea ecosystems, while such fluxes should be weakened where large carnivores and small passive ambush feeding copepods dominate. Our study supports the development of trait-based zooplankton functional groups in marine ecosystem models.

## Introduction

Zooplankton play a key role in the functioning of marine ecosystems. They actively transport particles to the deeper ocean through vertical migration (Jónasdóttir, Visser, Richardson & Heath, 2015), and produce rapidly sinking fecal pellets after grazing on primary producers in the euphotic layer (Turner, 2002). They participate in the remineralisation of organic matter by feeding on particle aggregates (Alldredge, 1972; Nishibe et al., 2015), and are thus involved in nutrient cycling and in the biological carbon pump (Turner, 2015). Additionally, they represent a pivotal link between the basis of the food web and the upper trophic levels (Beaugrand, Brander, Lindley, Souissi & Reid, 2003; Beaugrand & Kirby, 2010). Zooplankton are subdivided into different size classes, among which the mesozooplankton (organisms ranging between 200  $\mu\text{m}$  and 2 mm; Sieburth, Smetacek & Lenz, 1978) have received most attention because of their high biomass (Verity & Smetacek, 1996). In terms of both abundance and diversity, mesozooplankton are dominated by copepods in many regions of the global ocean (Kjørboe, 2011a). Changes in copepod diversity can be indicative of climate variability impacts on ecosystem functioning (Hooff & Peterson, 2006).

To assess the vulnerability of marine ecosystems to climate change, global biogeochemical models have been developed to describe the processes controlling ocean biogeochemistry and ecosystem functioning. At first, such models summarized pelagic ecosystems using only a few compartments but they have evolved towards capturing greater ecological complexity, and therefore improved the quantification of ecosystems responses and biogeochemical fluxes (e.g. Le Quéré et al., 2016). However, to depict variations in zooplankton biomass, most ecosystem models still rely on a limited number of size classes from micro-, and meso- to macrozooplankton (Kishi et al., 2007; Le Quéré et al., 2016) so they still only poorly capture the variety of zooplankton life-histories and traits (Kjørboe, 2011a; Litchman, Ohman & Kjørboe, 2013).

An alternative trait-based approach for describing pelagic ecosystem dynamics is to define plankton functional types based on groups of species with similar functional traits (Pomerleau, Sastri & Beisner, 2015; Benedetti, Gasparini & Ayata 2016). Functional traits are phenotypic attributes that impact the fitness of a species and its contribution to ecosystem functioning (Violle et al., 2007). The appearance of functional traits is governed by the expression of genes within organisms, and trait expression, in turn, governs the organisms' fitness under different biotic and abiotic conditions (Barton et al., 2013). A functional trait can also emerge from the interactions between other traits (trade-offs; Kjørboe, 2011b) and environmental conditions, leading to certain trait combinations being favoured under certain conditions. For zooplankton, Litchman et al. (2013) proposed a comprehensive set of traits, encompassing multiple types (i.e. life-history, morphological, physiological and behavioural) and life missions (i.e. feeding, survival, growth and reproduction). However, zooplankton traits have been described exhaustively for a few taxa only, and are mostly the results of laboratory experiments (Hébert, Beisner & Maranger, 2016). Consequently, depicting zooplankton functional diversity at the community level often requires information about many species and remains possible only for

few well-studied regions and/or a limited number of groups. For instance, Pomerleau et al. (2015) described the temporal evolution of zooplankton functional diversity in the north-eastern Pacific using time series data, and Benedetti et al. (2016) identified functional groups of copepods in the Mediterranean Sea. Brun, Payne & Kiørboe (2016) identified strong spatial and seasonal variations of body and offspring sizes, myelination rates and feeding modes for planktonic copepods in the North Atlantic. Copepods have therefore emerged as an interesting model to study marine functional biogeography (Barton et al., 2013).

Understanding and quantifying the distribution of zooplankton functional groups along environmental gradients is critical to (i) develop more realistic marine ecosystem models, (ii) better understand the underlying environmental drivers of community structure, (iii) unravel how climate impacts the rate and amplitude of processes mediated by functional types, and (iv) predict how the three above-mentioned aspects may evolve in a global change context.

Modelling the environmental niche of each species, and then assessing differences between functional groups, on the basis of their constituent species, enables to explore the biogeography of functional groups (Brun et al., 2015). Field observations of species occurrences can be combined with environmental predictors to build environmental niche models (ENMs) that enable the characterization of the environmental preferences of a species (Colwell & Rangel, 2009). In spite of the recent development of global open access plankton occurrence databases (O'Brien, 2005; Buitenhuis et al., 2013), relatively few studies have used ENMs to investigate plankton biogeography (Robinson et al., 2011). For mesozooplankton, most studies have mined the Continuous Plankton Recorder (CPR) data to model past and future range variations of copepod species, and changes in ocean surface layer community composition (Reygondeau & Beaugrand, 2011; Chust et al., 2014; Villarino et al., 2015; Brun, Kiørboe, Licandro & Payne, 2016; Benedetti, Guilhaumon, Adloff & Ayata, 2017) but they ignored functional aspects of biodiversity. Hence, the link between trait and species biogeography is currently underexplored.

In particular, it remains to be tested how different combinations of morphological traits (e.g. body length) and life history traits (e.g. feeding mode, spawning strategy) are distributed along environmental gradients. Are scanning-current feeders restricted to more productive regions where larger phytoplankton cells emerge (Kiørboe, 2011b; Benedetti et al., 2016)? Are passive ambush-feeding species more tightly affiliated to stable and oligotrophic environments where food availability is limited and mortality rates high (Paffenhöfer, 1993; Kiørboe, 2011b)? Do large herbivorous species thrive in the same conditions as large carnivorous predators, or are they affiliated to very different niches? As ecosystem processes are mediated by such functional traits (*sensu* Hébert, Beisner et al., 2016), assessing how the latter are expressed under varying conditions is a prerequisite for understanding ecosystem functioning under climatic changes.

In this study, we attempt to bridge this gap through an investigation of the level of congruence between the environmental niches of species belonging to different copepod functional groups (FGs). We aim to (1) model the environmental niches of copepod species, (2) define relevant FGs based on species functional traits, and (3) test whether species with similar combinations of functional traits inhabit similar environmental niches.

## Material and Methods

### *Species selection and data.*

Copepod species most commonly found in the Mediterranean Sea and their trait values were taken from Benedetti et al. (2016). Multiple regional datasets were combined to obtain the geographical distribution of the copepod species. As none of the species is endemic to the Mediterranean Sea (Razouls, de Bovée, Kouwenberg & Desreumaux, 2005-2017), species niches must be calibrated at the global scale in order to avoid truncated response curves, and therefore biased niche estimates (Thuiller, Brotons, Araùjo & Lavorel, 2004). Among the 193 species initially described by Benedetti et al. (2016), only those with more than 50 observations at the regional scale, and with at least 15 additional occurrences at the global scale (OBIS; <http://www.iobis.org/>; accessed on December 11, 2014) were retained. Our final list encompasses 106 copepod species. Our list of 106 species gathers the most commonly observed ones in the epipelagic layer, including those representing most of the mesozooplankton biomass in the Mediterranean Sea (Siokou-Frangou et al., 2010; Razouls et al., 2005-2017). The average sampling depth of the presence data (for the 99.94% of the records for which sampling depth was available) was  $58 \pm 119$  m. The presence data was resampled at a resolution of  $1 \times 1^\circ$  grid cells, to match the resolution of the environmental layers (see below). The final data set contained 76 366 gridded presence points.

### *Functional groups from species functional traits.*

Four functional traits were extracted from Benedetti (2015): maximum body length (mm), trophic group (carnivore, omnivore, omnivore-carnivore, omnivore-herbivore, omnivore-detritivore), feeding mode (ambush, cruise, scanning current, or mixed), and egg-spawning strategy (broadcast-spawner, sac-spawner).

These traits were selected because of their ready availability in the literature (Brun, Payne & Kiørboe, 2017). Other important traits such as resting stages and physiological rates are still lacking for the majority of the taxa studied. However, the present traits cover different important ecological functions and can influence ecosystem processes (Hébert, Beisner et al., 2016). Body size is a “master trait” that transcends and scales with a variety of traits related to ecosystem processes such as carbon cycling, secondary productivity or nutrient transfer (Hébert, Beisner et al., 2016). Trophic group describes the primary food source of a species and therefore its role in food-web dynamics (Pomerleau et al., 2015). Though nearly all copepods are omnivorous, many species are known to have preferential food sources (Benedetti, 2016). While omnivores target phytoplankton, detritus or smaller zooplankton, carnivores prey upon earlier life stages, smaller adult copepods or microzooplankton and thus contribute to the top-down control on zooplankton. Omnivore-herbivorous copepods are known to primarily graze upon phytoplankton when available and are therefore involved in the energy transfer from the autotrophs to fishes. Omnivore-detritivores are associated with marine snow or carcasses, so they contribute to the recycling of organic matter. Feeding mode has strong implications for prey selection, energy

allocation or nutrient cycling (Litchman et al., 2013; Hébert, Beisner et al., 2016). For instance, ambush feeders that target motile preys are characterized by lower energy expenditure than active feeders, resulting in lower mortality but also a lower feeding efficiency. Similarly, spawning strategy shapes energy allocations. Species developing sacs to place their eggs in invest less energy in growth and survival (Litchman et al., 2013).

FGs for the 106 selected species were defined by performing a Multiple Correspondence Analysis (MCA; Husson, Lê & Pagès, 2010; Fig. 2) on the four functional traits to represent underlying structures in the species characteristics in a reduced functional space (Benedetti et al., 2016). MCA is an ordination method for the multivariate analysis of categorical variables. Maximum body length was thus transformed into four size classes (SC1: 0.5-1.2 mm; SC2: 1.3-1.8 mm; SC3: 1.9-3.0 mm; SC4: 3.4-8.2 mm) using hierarchical agglomerative clustering based on the Euclidean distance and a synoptic aggregation link (Husson et al., 2010; Legendre & Legendre, 2012). Trophic groups were also transformed into binary data because this allows the optimization of their description in functional space. For example, omnivore-herbivores were considered as both “omnivores” and “herbivores” and not just “omnivore-herbivore” as if it was an independent trophic group. Kaiser-Guttman’s criterion (Guttman, 1954) was used to select the significant MCA axes that constitute the functional space. The species are positioned along each MCA axis according to their combination of trait values. The coordinates of the species along the retained MCA axes were used to compute the inter-species Euclidean distance matrix. Hierarchical agglomerative clustering using Ward’s aggregation method (Legendre & Legendre, 2012) was performed on this matrix to produce a functional dendrogram. Several cutting-levels along this dendrogram were examined, and the one leading to FGs that are ecologically relevant was kept. The final cutting-level was chosen to ensure groups were functionally homogeneous (i.e. not too large) while also attempting to avoid functional redundancies between the groups (i.e. not too small and numerous).

The sensitivity of our FGs to the choice of trait definition and clustering method was tested in additional analyses. Our analysis was repeated using Gower’s distance with the UGPMA linkage method to draw the functional dendrogram (Mouchet et al., 2008). The absolute values of maximum body size were kept, since Gower’s distance can mix quantitative and qualitative data. The FGs defined using this method were very similar to those found with the initial methodology (cophenetic correlation coefficient between the original and the alternative dendrogram was equal to 0.80; Rohlf & Fisher, 1968).

### ***Choice of the environmental predictors for niche modelling.***

Monthly data on 27 environmental predictors thought to be ecologically and physiologically relevant for copepods were considered (Helaouët & Beaugrand, 2007; Beaugrand, Mackas & Goberville, 2013). These variables comprised sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD, defined according to three different criteria), surface chlorophyll-a concentration (Chla), bathymetric depth, surface photosynthetically active radiation (PAR) and PAR integrated over the MLD (MLPAR; Brun et al., 2015), surface

dissolved oxygen concentration, sea surface wind speed and sea level anomalies (Appendix S1). To avoid the use of variables that are too highly correlated, when two predictors showed an absolute Spearman's rank correlation coefficient ( $|\rho| \geq 0.7$ , one of them was removed (Dormann et al., 2012). The final set of variables consisted of SST, the seasonal range of SST values ( $\Delta$ SST), SSS, MLD based the temperature criterion (de Boyer Montégut, Madec, Fischer, Lazar & Iudicone 2004), MLPAR (based on the same MLD product), and the logarithm of surface chlorophyll-a concentration (logChla). Subsequently, multivariate and univariate ENMs were fitted for each species using the six selected predictors (Fig. 1).

### *Species niche characteristics.*

Various types of algorithm exist to model environmental niches from species observations matched with environmental predictors (Merow et al., 2014). Here, General Additive Models (GAMs) were chosen because they can be tuned to produce unimodal response curves, while allowing skewed response curves (Fig. 1). This type of response to environmental gradients is expected for copepods (Bonnet et al., 2005; Helaouët & Beaugrand, 2007).

Since GAMs require absence data that are not available at the global scale, pseudo-absences were simulated for each species. The method employed for drawing pseudo-absences impacts ENMs quality, and should be chosen in light of the ecological characteristics of the species studied (Chefaoui & Lobo, 2008; Barbet-Massin, Jiguet, Albert & Thuiller, 2012). Mesozooplankton are ectotherms whose population dynamics are tightly coupled to climate (Hays, Richardson & Robinson, 2005), and whose individuals are passively dispersed over very large spatial scales during relatively short time periods (Jönsson & Watson, 2016). Therefore, pseudo-absences were generated using an environmental and geographical weighting method (Hengl, Sierdsema, Radović & Dilo, 2009; Hattab et al., 2013; Fig. 1). This method locates the - absences in regions of low habitat suitability for the species (i.e. in environments that are the most dissimilar to the ones where presences are located) and that are relatively far from the observed presences. The weighting procedure is based on a presence-only envelope niche model, the Ecological Niche Factor Analysis (Hirzel, Hausser, Chessel & Perrin, 2002), which provides a Habitat Suitability Index (HSI) from 0 to 1 for each species as a function of the six chosen environmental predictors. Together with the presences, the HSI is then used to produce a probability distribution  $\tau$  for locating the pseudo-absence following Hengl et al. (2009):

$$\tau(x) = \left[ \frac{d(x) + (100 - HSI(x))}{2} \right]^2$$

with  $d$  being the distance between the geographic cell and the presence points normalized by the maximum distance, so that  $d$  varies between 0 and 100. Thus,  $\tau$  is used as a probability density function to select pseudo-absences at random away from known presences and in regions of unsuitable conditions (according to the actual presences). For each species, pseudo-absences were drawn from the monthly subsets of occurrence data to ensure they follow the same temporal variability as the presences. Repeating the pseudo-absence generation procedure



(leading to a different set of randomly chosen pseudo-absences), or increasing the number of pseudo-absences, had very little impact on model quality and predictions (Benedetti et al., 2017), so only one set of monthly pseudo-absences was generated per species.

Multivariate and univariate GAMs were fitted using the logit link function (binomial response), and the parameter controlling the degree of smoothness was reduced ( $k = 5$ ). ENMs were evaluated by cross-validating 80% of the presence and pseudo-absence data against the remaining 20%. A threshold-dependent metric of the overall performance of each model was computed: the true skill statistic (TSS; Allouche, Tsoar & Kadmon, 2006). TSS ranges from -1 to +1, with 0 indicating that an ENM does not perform better than random. For each species and predictor variable, five cross-validation runs were performed and the mean TSS was calculated to evaluate the predictive power of the selected variables. The five univariate response curves derived from each cross-evaluation run were averaged to provide the univariate response curve of each species.

Following Brun et al. (2015), niche characteristics were summarized by four statistics derived from the mean univariate response curve (Fig. 1). The median of the mean curve was used to estimate the univariate niche centers, and quantile ranges between the 10<sup>th</sup> and the 90<sup>th</sup> percentiles were computed to estimate niche breadths (i.e. relative species tolerance ranges). The predictor values at the 10<sup>th</sup> and 90<sup>th</sup> percentiles served as estimates of lower and upper niche boundaries. These four niche parameters were used to test the differences in the environmental preferences of FGs (next section). Confidence intervals for TSS values and each univariate niche parameter were obtained by applying a bootstrapping procedure on the species occurrences (presences and pseudo-absences) which generated 200 replicates of every univariate model and niche characteristic.

The multivariate models could have been used to derive response curves but this possibility was discarded because multivariate models generate response curves that are much harder to interpret (Irwin, Nelles & Finkel, 2012; Brun et al., 2015). The response curve of a multivariate model is generated through the variation of one predictor, while setting the value of the other predictors to their average. Even though this corrects for the effect of the other variables in the response, it can flatten parts of the response due to variable collinearity, even if patterns exist (Brun et al., 2015). Furthermore, the identity of misrepresented variables varies between species, making it even harder to compare species niche characteristics. This is why the univariate GAMs were selected over multivariate ones to depict the species niches.

TSS values for the 106 multivariate GAMs show that the six selected variables lead to a fair modelling of the species distribution (Appendix S2). The average TSS for all the multivariate models is 0.82 ( $\pm 0.11$ ), and species mean TSS ranges between 0.46 (*Oithona decipiens*) and 0.97 (*Labidocera wollastoni*). The univariate models differ greatly in their performance (Appendix S2): the best performing variables are SST (mean TSS = 0.57), SSS (mean TSS = 0.64) and  $\Delta$ SST (mean TSS = 0.63), while the other three variables obtained significantly lower average TSS values (pairwise Wilcoxon tests,  $p$ -value < 0.05). MLD (mean TSS = 0.39) performs slightly better than logChla (mean TSS = 0.34) and MLPAR which displays the lowest

average performance score (mean TSS = 0.291).

### ***Functional groups and functional traits in niche space.***

The niche characteristics of each species were used in a scaled Principal Component Analysis (PCA; Legendre & Legendre, 2012) to ordinate them according to their environmental preferences in a reduced space (hereafter “niche space”; Fig. 1). We assess the similarity between the niche characteristics of species by comparing their coordinates along the retained Principal Components (i.e. their position in niche space). To create a parsimonious niche space, Spearman’s rank correlation coefficients were used to examine the correlation between the univariate niche characteristics. When two niche parameters displayed  $|\rho| \geq 0.7$ , one of them was removed. From the correlation matrix (Appendix S3), the following eleven niche characteristics were kept to create the niche space: SST center and breadth (as center is correlated to both lower and upper boundaries), SSS niche center (negatively correlated to breadth and lower boundary), upper and lower boundaries of  $\Delta$ SST (as SST breadth is highly correlated to  $\Delta$ SST breadth), MLD center (which summarizes all other 3 MLD characteristics), all MLPAR parameters but the lower boundary (highly correlated to SST and MLPAR centers), and both logChla center and breadth (logChla niche center summarises the two boundaries). Kaiser-Guttman’s criterion was used to select the number of significant PCs that will generate the niche space (Fig. 1).

In order to test for the differences between the niches of species pertaining to different FG, the central positions of the groups in niche space were computed as the average coordinates of the constituting species. Kruskal-Wallis variance analysis (Kruskal & Wallis, 1952) was performed to test whether FGs differed significantly in their positions in niche space. Similarly, variations between the positions in niche space of each functional trait were tested (for instance, between the different trophic groups) in order to explore whether some are more closely linked to certain environmental conditions. All statistical analyses were conducted with R v.3.4.0 (R Core Team 2017). The *biomod2* package (Thuiller, Georges & Engler, 2013) was used to develop the niche models and extract the response curves. The *FactoMineR* package (Husson, Josse, Lê & Mazet, 2017) was used for the multivariate analyses.

## **Results**

### ***Copepod functional groups.***

The first separation in the functional dendrogram (Fig. 2) distinguishes carnivorous species (FG 1 and 2) from the non-carnivorous ones (FG 3 to 7). Among the latter, the next cutting level distinguishes egg-broadcasters (FG 3 and 4) from sac-spawners (FG 5, 6 and 7). Further down on the dendrogram, trophic groups and size classes drive the composition of the FGs. Hence, all traits contribute to the structure of the functional space, with carnivorous diet being the primary discriminative trait.

Given that the 106 species were classed into different FGs on the basis of a dendrogram resulting from a cluster analysis of four different traits, some species were assigned to a FG when they differ in one particular trait with respect to other species because they present similar values of the other traits considered. FG 1 is composed of large sac-spawning carnivores that either feed by cruising or through a scanning current, but five of the ten species constituting this FG lack information on their feeding strategy. FG 2 is defined by 15 smaller and strictly carnivorous species that mainly feed through ambush tactics. The spawning strategy is equally divided between broadcasters and sac-spawners. Most of them inhabit the epi- and mesopelagos, and are thought to exhibit weak to no diel vertical migration (DVM) behaviour. FG 3 is defined by eight scanning current-feeding species belonging to the largest size class. All but two species (*Eucalanus hyalinus* and *Pleuromamma abdominalis*) are usually considered herbivores. The group's main spawning strategy is broadcasting. Some of them are established as strong migrants in the Mediterranean Sea (*Calanus helgolandicus*, *P. abdominalis* and *Neocalanus gracilis*), while the remainder have rarely exhibited any DVM behaviour. FG 4 is the largest group (n = 27) and gathers smaller (compared to FG 3) scanning current-feeding herbivores and omnivores displaying mixed feeding strategies. Like FG 3, the copepod species are all broadcasters, thus size class was the factor dividing these two FG. Most of the species in FG 4 preferentially inhabit the epipelagos and are known to exhibit weak to no DVM behaviour. FG 5 is the second largest group (n = 21) and clusters the small sac-spawning detritivorous species together. The dominant feeding strategy is cruising, as only 4 species rely on a scanning current to capture their food. Additionally, 4 species of this group are small cruise-feeders. Most of these species are known to occur within a broad depth range. Small ambush-feeding omnivores belonging to the *Oithona* genus constitute FG 6, together with three species (*Haloptilus longicornis*, *Isias clavipes* and *Lubbockia squillimana*) for which information about feeding strategy was lacking, but that are characterized by similar sizes and trophic groups. FG 7 gathers small herbivores that feed either by cruising or a scanning current. They are separated from the herbivores of FG 4 because they are sac-spawners and not broadcasters. The corresponding species are known to mainly occur in the epipelagos.

#### ***Species position in niche space based on the niche characteristics.***

A PCA based on individual species niche parameters was used to summarize the species' relative position in environmental space. The first three PCs were retained, accounting for 72.27% of the total variance. The first PC (PC1, 35.68%; Fig. 3) separates species affiliated with oligotrophic conditions (warm SST, higher SSS and low seasonality) from those affiliated with more productive conditions (colder and less saline waters, often characterized by higher seasonality). The species with higher values of SST centers, SSS centers, MLPAR centers and upper MLPAR boundaries have positive coordinates on PC1. Meanwhile, species displaying higher logChla centers and upper  $\Delta$ SST boundaries are located on the negative side of PC1. The second PC (PC2, 21.74%; Fig. 3) distinguishes species with broader tolerance to variations in SST and in chlorophyll-a concentrations

(positive coordinates) from species affiliated to habitats of lesser seasonal variability (i.e. characterized by higher values of lower  $\Delta$ SST boundaries). So, copepod species with positive PC2 coordinates can endure colder temperatures and seasonally varying environments. Meanwhile, species that present positive coordinates on PC1 and negative coordinates along PC2 are specialized in more oligotrophic conditions. The PCA reveals that species show distinct co-variations between their niche characteristics, thus demonstrating that they have different environmental preferences. The values of the species niche parameters are given in the Appendix S2.

### ***Do functional groups occupy distinct environmental niches?***

The FGs were represented in niche space after the mean position of each group was computed to assess if different FGs occupy distinct environmental niches (Fig. 4). The PC1 coordinates of the species differ significantly between FGs (Kruskal-Wallis test, p-value = 0.007, H-value = 17.52), which indicates that species between FGs present distinct environmental niches. No significant differences were found between the FG coordinates along PC2: the niche characteristics structuring this PC do not differ between FGs. The groups gathering carnivorous species (FG 1 and 2) were on average located on the positive side of PC1 while FG 3 (large scanning current-feeding herbivores), FG 6 (small ambush-feeding omnivores), and FG 7 (small sac-spawning herbivores) were on the negative side. This supports the hypothesis that there are differences in the environmental niches between groups: FG 1 and 2 are affiliated to warmer, saltier, more stable and less productive conditions in the open ocean; meanwhile, FG 3, 6 and 7 are associated with conditions of stronger seasonal variations, colder temperatures, lower salinities and higher chlorophyll-a concentrations. The mean positions of FG 4 (small broadcasting scanning current-feeding herbivores) and FG 5 (detritivores) were predominantly central in the environmental niche space. Therefore, these two FGs are composed of species that are spread out in niche space. With regard to the distribution of functional traits in niche space, only the mean position of each trophic group (Carnivores vs. Omnivores vs. Herbivores vs. Detritivores) showed significant differences along PC1 (Kruskal-Wallis test, p-value = 0.023, H-value = 11.38).

## **Discussion**

### ***Mediterranean copepod functional groups differ in their niches.***

The main finding of our study is that, on average, the environmental niches of the species within a FG differ from those of species constituting another FG (Fig. 4). Under the more tropical conditions where FGs 1 and 2 are favoured, we expect the Mediterranean zooplankton food-web dynamics to be dominated by top-down processes. The small passive ambush-feeding taxa (Corycaids and some *Oithona* species) benefit from a metabolism that is well adapted to food-depleted conditions where competition and predation are high (Landry, 1985; Kiørboe, 2011a).

When these small species dominate, we also expect size-related ecosystem fluxes (e.g. secondary productivity, nutrient recycling, excretion or carbon transfer etc.) to be weakened (Hébert, Beisner et al., 2016), which is the case in the ultra-oligotrophic eastern Mediterranean Sea (Siokou-Frangou et al., 2010; The MerMex Group, 2011). Larger predatory copepods (FG 1) may also occur, as they are able to feed on smaller copepods but also larger gelatinous zooplankton like Doliolids or Appendicularia (Ohtsuka & Onbé, 1989; Takahashi et al., 2013). Under the colder, more variable and more productive conditions where FGs 3 and 7 are favoured, we expect Mediterranean zooplankton food-web dynamics to be dominated by bottom-up processes, as these functional entities thrive on phytoplankton of different size classes. The larger current-feeding taxa (FG 3) are able to feed on the larger phytoplankton and we expect those species to mediate stronger energy fluxes from the surface to the deeper ocean (Jónasdóttir et al., 2015; Visser, Grønning & Jónasdóttir, 2017). Rates of secondary productivity, nutrient recycling and carbon transfer should be enhanced in the communities where such species dominate (Hébert, Beisner et al., 2016). These FGs are known to be more abundant in the coldest and most productive regions of the Mediterranean Sea: the northwestern regions (Gulf of Lion and Ligurian Sea), the northern Adriatic and Aegean Sea (Siokou-Frangou et al., 2010; The MerMex Group, 2011). Regarding small ambush-feeding generalists (FG 5), the positioning of *O. similis* at the extreme negative end of the niche space (Fig. 3) skews the average position of FG 5, whereas most of the species constituting this group are located at the opposite side of the niche space. *O. similis* is known to reach high abundances and biomass in colder environments (Gallienne & Robins, 2001; Castellani, Licandro, Fileman, Di Capua & Mazzocchi, 2016), but it has been frequently misidentified with *O. helgolandica* and is sometimes even considered as a conglomerate of cryptic species (Razouls et al., 2005-2017). Based on the present results, we do not expect cruise-feeding detritivorous species to be favoured under any particular conditions in the Mediterranean Sea.

The extent to which our findings apply to the global scale remains to be tested and depends on the representativeness of the Mediterranean copepod fauna relative to the global one. Yet it should be noted that the occurrence data we used here already covers a nearly global latitudinal and environmental range (Appendices S4 and S5). The geographical range of the data used to model the copepod niches is comparable across FGs, with data mainly located in the Mediterranean Sea and its surrounding basins: the North Atlantic and Indian Oceans. A notable exception is FG 6 because of the wider distribution of *Oithona similis* in polar oceans. Similarly, the distribution of the environmental predictors associated with the monthly presence of species largely overlaps between FGs (Appendix S5), suggesting that the differences observed in niche space (Fig. 4) result from the differential segregation of the FG presence data in geographical and/or environmental space.

Based on these findings, we anticipate that carnivorous FGs should dominate community composition in tropical conditions, whilst small ambush-feeding omnivores (*Oithona* spp.), large current-feeding herbivores and some smaller current-feeding herbivores should prevail in temperate, subpolar and polar oceans. Our results are in line with those of Woodd-Walker, Ward

& Clarke (2002) who found a higher proportion of carnivorous copepods at lower latitudes, relative to herbivorous and omnivorous copepods. Primary production is low in warmer seas, so there is a clear ecological advantage in specializing on food sources other than phytoplankton. Alternatively, species that primarily graze on phytoplankton and store lipids, such as many large Calanoida (Barton et al., 2013) are disadvantaged under tropical conditions. Our results are also in agreement with Atkinson (1998), who observed that the strong seasonality in food availability at high latitudes favours two types of life history strategies: (i) small generalists (typified by *Oithona* spp.) that present extended feeding periods thanks to a wide range of feeding mechanisms and preys (Castellani, Irigoien, Harris & Lampitt, 2005); and (ii) larger herbivorous lipid-storing species (typified by *Calanus* and other large Calanoida) that are able to enter diapause at greater depth for overwintering. This is in line with the association of large current-feeders with colder and more productive conditions.

### ***Implications for ecosystem functions and modelling.***

Based on our analyses we recommend the integration of selected, observation-based zooplankton functional groups in regional to global marine ecosystem models to better represent the diversity of plankton and their functional roles in ecosystems (Le Quéré et al., 2016). At present, copepods are represented by the mesozooplankton compartment in most models, a size class that was included due to its important role in the carbon cycle via the formation of particulate organic matter, and its link between phytoplankton and higher trophic levels (Le Quéré et al., 2005). In current models, the mesozooplankton group has a fixed trophic level and comprises groups with different trophic levels and distinct traits. Such a simplistic representation prevents an accounting of food-web dynamics involving multi-trophic interactions and energy transfer from the first trophic level upwards through multiple channels, with potentially different time scales and transfer rates. A recent study demonstrates the importance of zooplankton food web interactions for ecosystem characteristics in a model with higher heterotroph complexity (Le Quéré et al., 2016). The authors suggest that the inclusion of multiple zooplankton taxa at different trophic levels, and with different physiological and behavioral characteristics, may be a promising avenue in order to improve the representation of marine ecosystem dynamics in a more realistic way. Our results confirm these findings and show that even within a single size class, the available functional trait data is now ample enough to inform the observation-based definition of functional groups based on physiological and behavioural characteristics.

The delineation of FGs presented here is relevant for Mediterranean and global ecosystem models because FGs (i) play different roles in food webs, (ii) are characterized by different, though overlapping, environmental and physiological requirements, and (iii) impact the abundance of other PFTs and energy and matter transfer patterns within marine food-webs (Hood et al., 2006; Le Quéré et al., 2016). Current-feeding herbivores primarily graze upon the phytoplankton and the protozooplankton and then are preyed upon by predators (i.e. carnivorous copepods or fishes; Lynam et al., 2017). Carnivorous FGs contribute to the top-down control of

other FGs. The detritivores graze upon particulate organic matter, therefore affecting its re-mineralization and attenuation at depth (Nishibe et al., 2015).

#### ***Caveats and robustness of our study.***

In this study, all niche characteristics were derived from GAMs, one of the many existing ENMs (Merow et al., 2014). ENM choice is the primary cause of uncertainty in niche model projections (Diniz-Filho et al., 2009; Garcia, Burgess, Cabeza, Rahbek, Araújo & 2012; Benedetti et al., 2017) and is therefore likely to be the largest uncertainty factor in our niche parameters. We visually compared the response curves with those obtained from generalized linear models (GLMs), maximum entropy (MaxEnt; Phillips, Anderson & Schapire, 2006), flexible discriminant analysis (FDA), multivariate adaptive regression splines (MARS), neural network analysis (ANN), classification tree analysis (CTA), boosted regression trees (BRT) and random forest (RF). Most of these models (GLMs, FDA, MARS and MaxEnt) generated response curves with shapes similar to those obtained from GAMs, while deviations in certain models (CTA, BRT) were explained by methodological reasons, thus suggesting that our results are robust to model choice (Appendix S6) and model set-up.

Although our traits span diverse types and functions, many other quantitative and qualitative functional traits exist for zooplankton (Litchman et al., 2013). Our choice of traits was limited by the amount of available information for more than 100 species that are representative of Mediterranean planktonic ecosystems. In spite of the growing literature publishing trait tables for zooplankton, physiological rates are still only measured for a small pool of calanoid copepods (Kiørboe & Hirst, 2014; Hébert et al., 2016; Brun et al., 2017). Ultimately, the identified FGs and the covered range of trait values depend on the species pool studied. In the present framework, missing functional trait values have relatively minor impact on the FGs definition because species with more than two missing traits are treated as supplementary objects in the MCA, meaning they are placed in the functional space a posteriori and according to their similarity with the fully-informed species (Benedetti et al., 2016). A species with only one informed trait will deviate from the center of the MCA space only along the dimensions that are scored by this very trait. Ignoring rare species was necessary in our study since those largely lack both trait information and distribution data. The species with several missing trait values studied here are also those less frequently sampled in the Mediterranean Sea (Razouls et al., 2005-2017; Mazzocchi et al., 2014). Apart from the species belonging to the Siphostomatoida order, which are mostly semi-parasitic, it is unlikely that including rare species would add trait combinations that are completely novel compared to the ones covered here. As rarer copepod species are much less abundant in the plankton, it is also unlikely that they play an equally significant role in Mediterranean ecosystem functioning than the species studied here. However, we acknowledge that it remains to be tested how accounting for additional species, or clades, would change the distribution of FGs in niche space.

Although copepods represent the largest fraction of biomass and diversity in the Mediterranean zooplankton (Siokou-Frangou et al., 2010), not accounting for other taxa, such as pteropods,

539 chaetognaths, salps and appendicularians, may potentially ignore important ecological functions  
540 performed by these other groups. The next step will be to enlarge the species pool to the  
541 dominant taxa of other basins to test whether our findings apply to zooplankton communities  
542 worldwide.



## References

- Allredge, A. L. (1972). Abandoned larvacean houses: a unique food source in the pelagic environment. *Science*, 177(4052), 885-887.
- Allouche, O., Tsoar A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.
- Atkinson, A. (1998). Life cycle strategies of epipelagic copepods in the Southern Ocean. *Journal of Marine Systems*, 15(1), 289-311.
- Barbet-Massin, M., Jiguet F., Albert C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.
- Barton, A. D., Pershing A. J., Litchman E., Record N. R., Edwards K. F., Finkel Z. V., Kiørboe T., & Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, 16(4), 522-534.
- Beaugrand, G., Brander K. M., Lindley J. A., Souissi S., & Reid, P. C. (2003). Plankton effect on cod recruitment in the North Sea. *Nature*, 426(6967), 661-664.
- Beaugrand, G., & Kirby, R. R. (2010). Climate, plankton and cod. *Global Change Biology*, 16(4), 1268-1280.
- Beaugrand, G., Mackas D., & Goberville, E. (2013). Applying the concept of the ecological niche and a macroecological approach to understand how climate influences zooplankton: advantages, assumptions, limitations and requirements. *Progress in Oceanography*, 111, 75-90.
- Benedetti, F. (2015). Mediterranean copepods' functional traits. In: *Supplement to: Benedetti, Fabio; Gasparini, Stéphane; Ayata, Sakina-Dorothee (2016): Identifying copepod functional groups from species functional traits. Journal of Plankton Research*, 38(1), 159-166, <https://doi.org/10.1093/plankt/fbv096>.
- PANGAEA
- Benedetti, F., Gasparini S., & Ayata, S-D. (2016). Identifying copepod functional groups from species functional traits. *Journal of Plankton Research*, 38(1), 159-166.
- Benedetti, F., Guilhaumon F., Adloff F., & Ayata, S-D. (2017). Investigating uncertainties in zooplankton composition shifts under climate change scenarios in the Mediterranean Sea. *Ecography*, doi: 10.1111/ecog.02434.
- Bonnet, D., Richardson A., Harris R., Hirst A., Beaugrand G., Edwards M., Ceballos S., Diekman R., Lopez-Urrutia A., & Valdes, L. (2005). An overview of *Calanus helgolandicus* ecology in European waters. *Progress in Oceanography*, 65(1), 1-53.

- Brun, P., Vogt M., Payne M. R., Gruber N., O'Brien C. J., Buitenhuis E. T., Le Quéré C., Leblanc K., & Luo, Y. W. (2015). Ecological niches of open ocean phytoplankton taxa. *Limnology and Oceanography*, 60(3), 1020-1038.
- Brun, P., Kiørboe T., Licandro P., & Payne, M. R. (2016). The predictive skill of species distribution models for plankton in a changing climate. *Global Change Biology*, 22(9), 3170-3181.
- Brun, P., Payne M. R., & Kiørboe, T. (2016). Trait biogeography of marine copepods—an analysis across scales. *Ecology Letters*, 19(12), 1403-1413.
- Brun, P., Payne M. R., & Kiørboe, T. (2017). A trait database for marine copepods. *Earth System Science Data*, 9, 99-113. doi: 10.5194/essd-2016-30.
- Buitenhuis, E., Vogt M., Moriarty R., Bednarsek N., Doney S., Leblanc K., Le Quéré C., Luo Y.-W., O'Brien C., & O'Brien, T. (2013). MAREDAT: towards a world atlas of MARine Ecosystem DATA, *Earth System Science Data*, 5, 227-239.
- Castellani, C., Irigoien X., Harris R. P., & Lampitt, R. S. (2005). Feeding and egg production of *Oithona similis* in the North Atlantic. *Marine Ecology Progress Series*, 288, 173-182.
- Castellani, C., Licandro P., Fileman E., Di Capua I., & Mazzocchi, M. G. (2016). *Oithona similis* likes it cool: evidence from two long-term time series. *Journal of Plankton Research*, 38(3), 703-717.
- Chefaoui, R. M., & Lobo, J. M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210(4), 478-486.
- Chust, G., Castellani C., Licandro P., Ibaibarriaga L., Sagarminaga Y., & Irigoien, X. (2014). Are *Calanus* spp. shifting poleward in the North Atlantic? A habitat modelling approach, *ICES Journal of Marine Science: Journal du Conseil*, 71(2), 241-253.
- Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences*, 106(Supplement 2), 19651-19658.
- de Boyer Montégut, C., Madec G., Fischer A. S., Lazar A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, 109(C12).
- 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, 897-906.
- Dormann, C. F., Schymanski S. J., Cabral J., Chuine I., Graham C., Hartig F., Kearney M., Morin X., Römermann C., & Schröder, B. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, 39(12), 2119-2131.

- Gallienne, C., & Robins, D. (2001). Is *Oithona* the most important copepod in the world's oceans? *Journal of Plankton Research*, 23(12), 1421-1432.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149-161.
- Hattab, T., Lasram F. B. R., Albouy C., Sammari C., Romdhane C. M., Cury P., Leprieur F., & Le Loc'h, F. (2013). The use of a predictive habitat model and a fuzzy logic approach for marine management and planning. *PLoS ONE*, 8(10), e76430.
- Hays, G. C., Richardson A. J., & Robinson, C. (2005). Climate change and marine plankton. *Trends in Ecology & Evolution*, 20(6), 337-344.
- Hébert, M-P., Beisner B. E., & Maranger, R. (2016). A compilation of quantitative functional traits for marine and freshwater crustacean zooplankton. *Ecology*, 97(4), 1069-1081, doi: 10.1890/15-1275
- Hébert, M-P., Beisner B. E., & Maranger, R. (2016). Linking zooplankton communities to ecosystem functioning: toward an effect-trait framework. *Journal of Plankton Research*, 39, 3-12.
- Helaouët, P., & Beaugrand, G. (2007). Macroecology of *Calanus finmarchicus* and *C. helgolandicus* in the North Atlantic Ocean and adjacent seas. *Marine Ecology Progress Series*, 345, 147-165.
- Hengl, T., Sierdsema H., Radović A., & Dilo, A. (2009). Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling*, 220(24), 3499-3511.
- Hirzel, A., Hausser J., Chessel D., & Perrin, N. (2002). Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83(7), 2027-2036.
- Hooff, R.C. & Peterson, W.T. (2006). Copepod biodiversity as an indicator of changes in ocean and climate conditions of the northern California current ecosystem. *Limnology and Oceanography*, 51, 2607-2620.
- Husson, F., Lê S., & Pagès, J. (2010). Exploratory multivariate analysis by example using R, CRC press. Taylor & Francis Group. United States of America.
- Husson, F., Josse, J., Lê, S. & Mazet, J. (2017). Package 'FactoMineR'. *Multivariate Exploratory Data Analysis and Data Mining*: <http://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>
- Irwin, A. J., Nelles, A. M., & Finkel, Z. V. (2012). Phytoplankton niches estimated from field data. *Limnology and Oceanography*, 57(3), 787-797.

- 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.
- Jönsson, B. F., & Watson, J. R. (2016). The timescales of global surface-ocean connectivity. *Nature Communications*, 7, 11239.
- Kjørboe, T. (2011a). What makes pelagic copepods so successful? *Journal of Plankton Research*, 33(5), 677-685.
- Kjørboe, T. (2011b). How zooplankton feed: mechanisms, traits and trade-offs. *Biological Reviews*, 86(2), 311-339.
- Kjørboe, 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.
- Kishi, M. J., Kashiwai M., Ware D. M., Megrey B. A., Eslinger D. L., Werner F. E., Noguchi-Aita M., Azumaya T., Fujii M., & Hashimoto, S. (2007). NEMURO—a lower trophic level model for the North Pacific marine ecosystem. *Ecological Modelling*, 202(1), 12-25.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.
- Legendre, P., & Legendre, L. (2012). *Numerical ecology*. 3<sup>rd</sup> English Edition. Elsevier Science BV, Amsterdam.
- Le Quéré, C., Harrison S. P., Colin Prentice I., Buitenhuis E. T., Aumont O., Bopp L., Claustre H., Cotrim Da Cunha L., Geider R., & Giraud, X. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11), 2016-2040.
- Le Quéré, C., Buitenhuis E. T., Moriarty R., Alvain S., Aumont O., Bopp L., Chollet S., Enright C., Franklin D. J., Geider R. J., Harrison S. P., Hirst A. G., Larsen S., Legendre L., Platt T., Prentice I. C., Rivkin R. B., Sailley S., Sathyendranath S., Stephens N., Vogt M., & Vallina, S. M. (2016). Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles. *Biogeosciences*, 13(14), 4111-4133.
- Litchman, E., Ohman M. D., & Kjørboe, T. (2013). Trait-based approaches to zooplankton communities. *Journal of Plankton Research*, 35(3), 473-484.
- Lynam, C. P., Llope M., Möllmann C., Helaouët P., Bayliss-Brown G. A., & Stenseth, N. C. (2017). Interaction between top-down and bottom-up control in marine food webs. *Proceedings of the National Academy of Sciences*, 114(8), 1952-1957.
- Mauchline, J. (1998). *The biology of Calanoid copepods*. Academic Press. San Diego, California, USA.

- Mazzocchi, M., Siokou-Frangou I., Tirelli V., Bandelj V., de Puellès M. F., Örek Y. A., de Olazabal A., Gubanov A., Kress N., and Protopapa, M. (2014). Regional and seasonal characteristics of epipelagic mesozooplankton in the Mediterranean Sea based on an artificial neural network analysis. *Journal of Marine Systems*, 135, 64-80.
- Merow, C., Smith M. J., Edwards T. C., Guisan A., McMahon S. M., Normand S., Thuiller W., Wüest R. O., Zimmermann N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37(12), 1267-1281.
- 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.
- Nishibe, Y., Takahashi K., Ichikawa T., Hidaka K., Kurogi H., Segawa K., & Saito, H. (2015). Degradation of discarded appendicularian houses by oncaeid copepods. *Limnology and Oceanography*, 60(3), 967-976.
- O'Brien, T. (2005). COPEPOD: A global plankton database. *NOAA Technical Memorandum NMFS-F/SPO-73*, 19.
- Paffenhöfer, G-A. (1993). On the ecology of marine cyclopoid copepods (Crustacea, Copepoda). *Journal of Plankton Research*, 15(1), 37-55.
- 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.
- Phillips, S. J., Anderson R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190(3), 231-259.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Razouls C., de Bovée F., Kouwenberg J. & Desreumaux, N. (2005-2017). Diversity and Geographic Distribution of Marine Planktonic Copepods. Available at <http://copepodes.obs-banyuls.fr/en>.
- Reygondeau, G., & Beaugrand, G. (2011). Future climate-driven shifts in distribution of *Calanus finmarchicus*. *Global Change Biology*, 17(2), 756-766.
- Robinson, L., Elith J., Hobday A., Pearson R., Kendall B., Possingham H., & Richardson, A. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, 20(6), 789-802.

- Sieburth, J. M., Smetacek V., & Lenz, J. (1978). Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology and Oceanography*, 23(6), 1256-1263.
- Siokou-Frangou, I., Christaki U., Mazzocchi M. G., Montresor, M., Ribera d'Alcala M., Vaqué D., & Zingone, A. (2010). Plankton in the open Mediterranean Sea: a review. *Biogeosciences*, 7(5), 1543-1586.
- The MerMex Group (2011). Marine ecosystems' responses to climatic and anthropogenic forcings in the Mediterranean. *Progress in Oceanography*, 91(2), 97-166.
- Thuiller, W., Brotons L., Araújo M. B., & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, 27(2), 165-172.
- Turner, J. T. (2002). Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquatic Microbial Ecology*, 27(1), 57-102.
- Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, 130, 205-248.
- Verity, P., & Smetacek, V. (1996). Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Marine Ecology Progress Series*, 130, 277-293.
- Villarino, E., Chust G., Licandro P., Butenschön M., Ibaibarriaga L., Larrañaga A., & Irigoien, X. (2015). Modelling the future biogeography of North Atlantic zooplankton communities in response to climate change. *Marine Ecology Progress Series*, 531, 121-142.
- 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.
- Visser, A. W., Grønning J., & Jónasdóttir, S. H. (2017). *Calanus hyperboreus* and the lipid pump. *Limnology and Oceanography*, 62(3), 1155-1165.
- 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 currently a post-doc at the Observatoire Océanologique de Villefranche-sur-Mer where he is studying zooplankton trait biogeography at multiple scales. His main research interests focus on the co-variation of biodiversity components with environmental gradients and how these affect ecosystem processes simultaneously.

Author contributions: F.B. gathered the biological data and designed the study with M.V. and S-D.A. D.R. provided the environmental layers, and helped in optimizing the methods with F.G. F.B. led the writing of the manuscript that all co-authors contributed to.

## Figure Legends

**Figure 1:** Flowchart describing the main steps of the methodology for the present study. From the top to the bottom, the left-hand side shows how the species functional traits table was used to define functional groups from a multivariate functional space. The right-hand side shows how environmental niche modelling was performed to quantify the species' environmental preferences (i.e. niche parameters). The niche parameters were selected according to their correlation in order to perform a parsimonious Principal Component Analysis (PCA) to generate the niche space. Finally (bottom), the position of the functional group (FG) in niche space was assessed and variance analysis was performed to test if different FGs present different niches.

**Figure 2:** Functional dendrogram showing the ordination of the 106 copepod species according to the combination of functional traits. A Multivariate Correspondence Analysis (MCA) based on the functional traits was performed to ordinate the species in a reduced functional space. The coordinates of the species along the four retained axes of the functional space were used to compute a Euclidean distance matrix. Agglomerative hierarchical clustering was used to draw the functional dendrogram. The chosen cutting-level is shown and the species are coloured according to the seven derived functional groups.

**Figure 3:** Principal Component Analysis (PCA) based on the selected eleven niche characteristics of the 106 copepod species. The contribution of the eleven niche parameters to the niche space, as well as their correlation (i.e. the angles between the arrows) are evidenced. The abbreviations of the environmental predictors are as follows: sea surface temperature (SST), sea surface salinity (SSS), seasonal range of SST values ( $\Delta$ SST), mixed layer depth (MLD), active radiation (PAR) integrated over the MLD (MLPAR), and logarithm of surface chlorophyll-a concentration (logChla).

**Figure 4:** Position of the seven functional groups in niche space. Smaller symbols correspond to the 106 copepod species with the colors and shapes varying according to the functional group they belong to. Larger symbols correspond to the average position of the functional groups, based on the coordinates of the species they comprise. The first principal component (PC1) represents a gradient characterizing oligotrophic conditions on the positive side (higher centers of SST, SSS, MLPAR) and more productive conditions on the negative side (broader SST and SSS niches, and higher MLD and logChla centers). The second principal component (PC2) distinguishes species with broader tolerance to SST, logChla variations (positive side) from species affiliated to habitats of lesser seasonal variability (negative side). The standard error of each FG' coordinates along PC1 and PC2 are illustrated with the error bars. The abbreviations of the environmental predictors are as follows: sea surface temperature (SST), sea surface salinity (SSS), seasonal range of SST values ( $\Delta$ SST), mixed layer depth (MLD), active radiation (PAR) integrated over the MLD (MLPAR), and the logarithm of surface chlorophyll-a concentration (logChla).



## Supporting Information

**Supporting Information S1:** Table describing the set of environmental predictors tested in the present study. The time period covered by the monthly climatologies is given, as well as the corresponding references.

**Supporting Information S2:** Table summarising for every 106 copepod species: the number of observations (presences at the global scale, after re-sampling on the predictors cell grid); the average TSS value of the multivariate niche models and its standard deviation; the average niche characteristics derived from the univariate models and their associated standard deviations computed thanks to the bootstrapping procedure.

**Supporting Information S3:** Pair-wise Spearman's correlation coefficient ( $\rho$ ) values of all the univariate niche characteristics ( $n = 24$ ).

**Supporting Information S4:** Maps of the sampling effort (i.e. number of species presences) for each of the seven copepod functional group defined in the study. Presences are at a  $1^\circ \times 1^\circ$  cell grid resolution after re-sampling the observed occurrences on the cell grid of the environmental predictors.

**Supporting Information S5:** Distribution of the latitude and the six environmental predictors' values for each of the seven copepod functional groups defined in the study. Values correspond to the environmental monthly data that were fitted on each species monthly occurrence.

**Supporting Information S6:** Examples of response curves to Sea Surface Temperature (SST) variations for *Calanus helgolandicus*, *Acartia Acartiura clausi* and *Clausocalanus mastigophorus*, according to nine different types of environmental niche models (ENMs): Generalized Additive Models (GAM), Generalized Linear Models (GLM), Multi-Adaptive Regression Splines (MARS), Flexible Discriminant Analysis (FDA), Maximum Entropy (MAXENT), Boosted Regression Trees (BRT), Random Forest (RF), Classification Tree Analysis (CTA), and Neural Network Analysis (ANN). Additionally, the niche spaces (based on a Principal Component Analysis (PCA) performed on the chosen eleven niche characteristics) and positions of the seven copepod functional groups according to various parameters of the niche modelling framework designed for the present study: the choice of the ENM, the choice of the Generalized Additive Models (GAMs) smoothing parameter value ( $k$ ), and the percentage of the species' monthly occurrence data considered to train (and test) the GAMs.

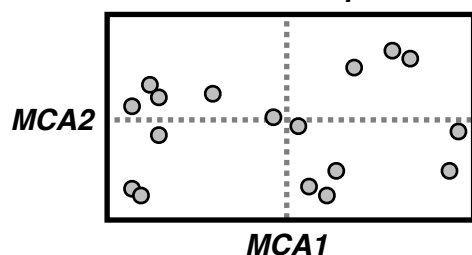
## Functional traits and groups

Functional traits table  
maximum body length,  
feeding mode, spawning  
strategy and trophic group  
106 species

Categorize  
all traits

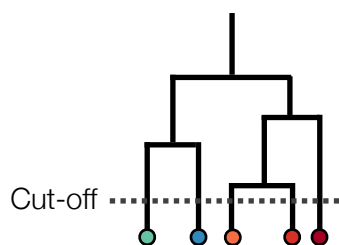
### Multiple Correspondence Analysis (MCA)

*Functional space*



Hierarchical clustering  
on MCA coordinates

*Functional dendrogram*



*Functional groups*

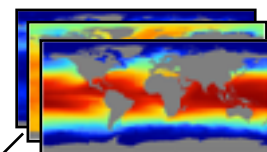
Project functional  
groups in niche space

## Environmental niche modelling

Presences  
Global, monthly,  $1^\circ \times 1^\circ$



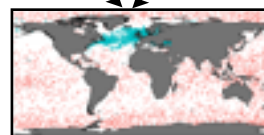
Predictors (n=6)  
Global, monthly,  $1^\circ \times 1^\circ$



Spatial buffer

Environmental  
weighing

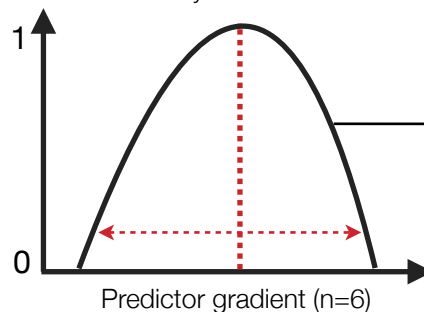
Presences +  
pseudo-absences  
(P/psA)



### Generalized Additive Models (GAMs) - cross-validation

*Average univariate response curves*

Habitat suitability



*Niche parameters*

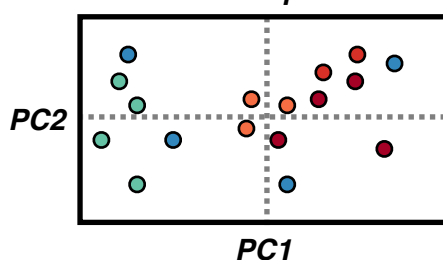
Center  
Breadth  
Lower boundary  
Upper boundary

x106 species

Parameter filtering  
(Spearman's rank correlation)

### Principal Component Analysis (PCA)

*Niche space*



Variance analysis  
on PC scores

*Are functional groups  
differently distributed  
in niche space ?*

