



**HAL**  
open science

## 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-01826722v1>**

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.

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

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

5 *Affiliations:*

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

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

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

12  
13  
14 *Running title:* Copepod functional groups and their niches.

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

## 17 18 19 **Acknowledgements**

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

30 **Abstract**

31 **Aim**

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

35

36 **Location**

37 The Mediterranean Sea.

38

39 **Methods**

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

48

49 **Results**

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

56

57 **Main conclusions**

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

## 64 **Introduction**

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

79 To assess the vulnerability of marine ecosystems to climate change, global biogeochemical  
80 models have been developed to describe the processes controlling ocean biogeochemistry and  
81 ecosystem functioning. At first, such models summarized pelagic ecosystems using only a few  
82 compartments but they have evolved towards capturing greater ecological complexity, and  
83 therefore improved the quantification of ecosystems responses and biogeochemical fluxes (e.g.  
84 Le Quéré et al., 2016). However, to depict variations in zooplankton biomass, most ecosystem  
85 models still rely on a limited number of size classes from micro-, and meso- to  
86 macrozooplankton (Kishi et al., 2007; Le Quéré et al., 2016) so they still only poorly capture the  
87 variety of zooplankton life-histories and traits (Kiørboe, 2011a; Litchman, Ohman & Kiørboe,  
88 2013).

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

104 few well-studied regions and/or a limited number of groups. For instance, Pomerleau et al.  
105 (2015) described the temporal evolution of zooplankton functional diversity in the north-eastern  
106 Pacific using time series data, and Benedetti et al. (2016) identified functional groups of  
107 copepods in the Mediterranean Sea. Brun, Payne & Kiørboe (2016) identified strong spatial and  
108 seasonal variations of body and offspring sizes, myelination rates and feeding modes for  
109 planktonic copepods in the North Atlantic. Copepods have therefore emerged as an interesting  
110 model to study marine functional biogeography (Barton et al., 2013).

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

116 Modelling the environmental niche of each species, and then assessing differences between  
117 functional groups, on the basis of their constituent species, enables to explore the biogeography  
118 of functional groups (Brun et al., 2015). Field observations of species occurrences can be  
119 combined with environmental predictors to build environmental niche models (ENMs) that  
120 enable the characterization of the environmental preferences of a species (Colwell & Rangel,  
121 2009). In spite of the recent development of global open access plankton occurrence databases  
122 (O'Brien, 2005; Buitenhuis et al., 2013), relatively few studies have used ENMs to investigate  
123 plankton biogeography (Robinson et al., 2011). For mesozooplankton, most studies have mined  
124 the Continuous Plankton Recorder (CPR) data to model past and future range variations of  
125 copepod species, and changes in ocean surface layer community composition (Reygondeau &  
126 Beaugrand, 2011; Chust et al., 2014; Villarino et al., 2015; Brun, Kiørboe, Licandro & Payne,  
127 2016; Benedetti, Guilhaumon, Adloff & Ayata, 2017) but they ignored functional aspects of  
128 biodiversity. Hence, the link between trait and species biogeography is currently underexplored.

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

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

144

## 145 **Material and Methods**

### 146 *Species selection and data.*

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

163

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

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

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

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

189 FGs for the 106 selected species were defined by performing a Multiple Correspondence  
190 Analysis (MCA; Husson, Lê & Pagès, 2010; Fig. 2) on the four functional traits to represent  
191 underlying structures in the species characteristics in a reduced functional space (Benedetti et al.,  
192 2016). MCA is an ordination method for the multivariate analysis of categorical variables.  
193 Maximum body length was thus transformed into four size classes (SC1: 0.5-1.2 mm; SC2: 1.3-  
194 1.8 mm; SC3: 1.9-3.0 mm; SC4: 3.4-8.2 mm) using hierarchical agglomerative clustering based  
195 on the Euclidean distance and a synoptic aggregation link (Husson et al., 2010; Legendre &  
196 Legendre, 2012). Trophic groups were also transformed into binary data because this allows the  
197 optimization of their description in functional space. For example, omnivore-herbivores were  
198 considered as both “omnivores” and “herbivores” and not just “omnivore-herbivore” as if it was  
199 an independent trophic group. Kaiser-Guttman’s criterion (Guttman, 1954) was used to select the  
200 significant MCA axes that constitute the functional space. The species are positioned along each  
201 MCA axis according to their combination of trait values. The coordinates of the species along the  
202 retained MCA axes were used to compute the inter-species Euclidean distance matrix.  
203 Hierarchical agglomerative clustering using Ward’s aggregation method (Legendre & Legendre,  
204 2012) was performed on this matrix to produce a functional dendrogram. Several cutting-levels  
205 along this dendrogram were examined, and the one leading to FGs that are ecologically relevant  
206 was kept. The final cutting-level was chosen to ensure groups were functionally homogeneous  
207 (i.e. not too large) while also attempting to avoid functional redundancies between the groups  
208 (i.e. not too small and numerous).

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

216

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

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

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

232

### 233 *Species niche characteristics.*

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

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

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

255

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



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

265 Multivariate and univariate GAMs were fitted using the logit link function (binomial response),  
266 and the parameter controlling the degree of smoothness was reduced ( $k = 5$ ). ENMs were  
267 evaluated by cross-validating 80% of the presence and pseudo-absence data against the  
268 remaining 20%. A threshold-dependent metric of the overall performance of each model was  
269 computed: the true skill statistic (TSS; Allouche, Tsoar & Kadmon, 2006). TSS ranges from -1 to  
270 +1, with 0 indicating that an ENM does not perform better than random. For each species and  
271 predictor variable, five cross-validation runs were performed and the mean TSS was calculated  
272 to evaluate the predictive power of the selected variables. The five univariate response curves  
273 derived from each cross-evaluation run were averaged to provide the univariate response curve  
274 of each species.

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

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

294 TSS values for the 106 multivariate GAMs show that the six selected variables lead to a fair  
295 modelling of the species distribution (Appendix S2). The average TSS for all the multivariate  
296 models is 0.82 ( $\pm 0.11$ ), and species mean TSS ranges between 0.46 (*Oithona decipiens*) and  
297 0.97 (*Labidocera wollastoni*). The univariate models differ greatly in their performance  
298 (Appendix S2): the best performing variables are SST (mean TSS = 0.57), SSS (mean TSS =  
299 0.64) and  $\Delta$ SST (mean TSS = 0.63), while the other three variables obtained significantly lower  
300 average TSS values (pairwise Wilcoxon tests,  $p$ -value < 0.05). MLD (mean TSS = 0.39)  
301 performs slightly better than logChla (mean TSS = 0.34) and MLPAR which displays the lowest

302 average performance score (mean TSS = 0.291).

303

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

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

322 In order to test for the differences between the niches of species pertaining to different FG, the  
323 central positions of the groups in niche space were computed as the average coordinates of the  
324 constituting species. Kruskal-Wallis variance analysis (Kruskal & Wallis, 1952) was performed  
325 to test whether FGs differed significantly in their positions in niche space. Similarly, variations  
326 between the positions in niche space of each functional trait were tested (for instance, between  
327 the different trophic groups) in order to explore whether some are more closely linked to certain  
328 environmental conditions. All statistical analyses were conducted with R v.3.4.0 (R Core Team  
329 2017). The *biomod2* package (Thuiller, Georges & Engler, 2013) was used to develop the niche  
330 models and extract the response curves. The *FactoMineR* package (Husson, Josse, Lê & Mazet,  
331 2017) was used for the multivariate analyses.

332

## 333 **Results**

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

335 The first separation in the functional dendrogram (Fig. 2) distinguishes carnivorous species (FG  
336 1 and 2) from the non-carnivorous ones (FG 3 to 7). Among the latter, the next cutting level  
337 distinguishes egg-broadcasters (FG 3 and 4) from sac-spawners (FG 5, 6 and 7). Further down on  
338 the dendrogram, trophic groups and size classes drive the composition of the FGs. Hence, all  
339 traits contribute to the structure of the functional space, with carnivorous diet being the primary  
340 discriminative trait.

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

369

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

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

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

390

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

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

410

## 411 **Discussion**

412

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

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

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

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

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

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

472

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

474 Based on our analyses we recommend the integration of selected, observation-based zooplankton  
475 functional groups in regional to global marine ecosystem models to better represent the diversity  
476 of plankton and their functional roles in ecosystems (Le Quéré et al., 2016). At present, copepods  
477 are represented by the mesozooplankton compartment in most models, a size class that was  
478 included due to its important role in the carbon cycle via the formation of particulate organic  
479 matter, and its link between phytoplankton and higher trophic levels (Le Quéré et al., 2005). In  
480 current models, the mesozooplankton group has a fixed trophic level and comprises groups with  
481 different trophic levels and distinct traits. Such a simplistic representation prevents an accounting  
482 of food-web dynamics involving multi-trophic interactions and energy transfer from the first  
483 trophic level upwards through multiple channels, with potentially different time scales and  
484 transfer rates. A recent study demonstrates the importance of zooplankton food web interactions  
485 for ecosystem characteristics in a model with higher heterotroph complexity (Le Quéré et al.,  
486 2016). The authors suggest that the inclusion of multiple zooplankton taxa at different trophic  
487 levels, and with different physiological and behavioral characteristics, may be a promising  
488 avenue in order to improve the representation of marine ecosystem dynamics in a more realistic  
489 way. Our results confirm these findings and show that even within a single size class, the  
490 available functional trait data is now ample enough to inform the observation-based definition of  
491 functional groups based on physiological and behavioural characteristics.

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

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

501

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

503 In this study, all niche characteristics were derived from GAMs, one of the many existing ENMs  
504 (Merow et al., 2014). ENM choice is the primary cause of uncertainty in niche model projections  
505 (Diniz-Filho et al., 2009; Garcia, Burgess, Cabeza, Rahbek, Araújo & 2012; Benedetti et al.,  
506 2017) and is therefore likely to be the largest uncertainty factor in our niche parameters. We  
507 visually compared the response curves with those obtained from generalized linear models  
508 (GLMs), maximum entropy (MaxEnt; Phillips, Anderson & Schapire, 2006), flexible  
509 discriminant analysis (FDA), multivariate adaptive regression splines (MARS), neural network  
510 analysis (ANN), classification tree analysis (CTA), boosted regression trees (BRT) and random  
511 forest (RF). Most of these models (GLMs, FDA, MARS and MaxEnt) generated response curves  
512 with shapes similar to those obtained from GAMs, while deviations in certain models (CTA,  
513 BRT) were explained by methodological reasons, thus suggesting that our results are robust to  
514 model choice (Appendix S6) and model set-up.

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

537 Although copepods represent the largest fraction of biomass and diversity in the Mediterranean  
538 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.



543 **References**

- 544 Alldredge, A. L. (1972). Abandoned larvacean houses: a unique food source in the pelagic environment.  
545 *Science*, 177(4052), 885-887.
- 546
- 547 Allouche, O., Tsoar A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models:  
548 prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.
- 549
- 550 Atkinson, A. (1998). Life cycle strategies of epipelagic copepods in the Southern Ocean. *Journal of*  
551 *Marine Systems*, 15(1), 289-311.
- 552
- 553 Barbet-Massin, M., Jiguet F., Albert C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species  
554 distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.
- 555
- 556 Barton, A. D., Pershing A. J., Litchman E., Record N. R., Edwards K. F., Finkel Z. V., Kiørboe T., &  
557 Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, 16(4), 522-534.
- 558
- 559 Beaugrand, G., Brander K. M., Lindley J. A., Souissi S., & Reid, P. C. (2003). Plankton effect on cod  
560 recruitment in the North Sea. *Nature*, 426(6967), 661-664.
- 561
- 562 Beaugrand, G., & Kirby, R. R. (2010). Climate, plankton and cod. *Global Change Biology*, 16(4), 1268-  
563 1280.
- 564
- 565 Beaugrand, G., Mackas D., & Goberville, E. (2013). Applying the concept of the ecological niche and a  
566 macroecological approach to understand how climate influences zooplankton: advantages, assumptions,  
567 limitations and requirements. *Progress in Oceanography*, 111, 75-90.
- 568
- 569 Benedetti, F. (2015). Mediterranean copepods' functional traits. In: *Supplement to: Benedetti,*  
570 *Fabio; Gasparini, Stéphane; Ayata, Sakina-Dorothee (2016): Identifying copepod functional groups from species*  
571 *functional traits. Journal of Plankton Research*, 38(1), 159-166, <https://doi.org/10.1093/plankt/fbv096>.
- 572 PANGAEA
- 573
- 574 Benedetti, F., Gasparini S., & Ayata, S-D. (2016). Identifying copepod functional groups from species  
575 functional traits. *Journal of Plankton Research*, 38(1), 159-166.
- 576
- 577 Benedetti, F., Guilhaumon F., Adloff F., & Ayata, S-D. (2017). Investigating uncertainties in zooplankton  
578 composition shifts under climate change scenarios in the Mediterranean Sea. *Ecography*, doi:  
579 10.1111/ecog.02434.
- 580
- 581 Bonnet, D., Richardson A., Harris R., Hirst A., Beaugrand G., Edwards M., Ceballos S., Diekman R.,  
582 Lopez-Urrutia A., & Valdes, L. (2005). An overview of *Calanus helgolandicus* ecology in European  
583 waters. *Progress in Oceanography*, 65(1), 1-53.
- 584

585 Brun, P., Vogt M., Payne M. R., Gruber N., O'Brien C. J., Buitenhuis E. T., Le Quéré C., Leblanc K., &  
586 Luo, Y. W. (2015). Ecological niches of open ocean phytoplankton taxa. *Limnology and Oceanography*,  
587 60(3), 1020-1038.

588

589 Brun, P., Kiørboe T., Licandro P., & Payne, M. R. (2016). The predictive skill of species distribution  
590 models for plankton in a changing climate. *Global Change Biology*, 22(9), 3170-3181.

591

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

594

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

597

598 Buitenhuis, E., Vogt M., Moriarty R., Bednarsek N., Doney S., Leblanc K., Le Quéré C., Luo Y.-W.,  
599 O'Brien C., & O'Brien, T. (2013). MAREDAT: towards a world atlas of MARine Ecosystem DATA,  
600 *Earth System Science Data*, 5, 227-239.

601

602 Castellani, C., Irigoien X., Harris R. P., & Lampitt, R. S. (2005). Feeding and egg production of *Oithona*  
603 *similis* in the North Atlantic. *Marine Ecology Progress Series*, 288, 173-182.

604

605 Castellani, C., Licandro P., Fileman E., Di Capua I., & Mazzocchi, M. G. (2016). *Oithona similis* likes it  
606 cool: evidence from two long-term time series. *Journal of Plankton Research*, 38(3), 703-717.

607

608 Chefaoui, R. M., & Lobo, J. M. (2008). Assessing the effects of pseudo-absences on predictive  
609 distribution model performance. *Ecological Modelling*, 210(4), 478-486.

610

611 Chust, G., Castellani C., Licandro P., Ibaibarriaga L., Sagarminaga Y., & Irigoien, X. (2014). Are  
612 *Calanus* spp. shifting poleward in the North Atlantic? A habitat modelling approach, *ICES Journal of*  
613 *Marine Science: Journal du Conseil*, 71(2), 241-253.

614

615 Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. *Proceedings of*  
616 *the National Academy of Sciences*, 106(Supplement 2), 19651-19658.

617

618 de Boyer Montégut, C., Madec G., Fischer A. S., Lazar A., & Iudicone, D. (2004). Mixed layer depth  
619 over the global ocean: An examination of profile data and a profile-based climatology. *Journal of*  
620 *Geophysical Research: Oceans*, 109(C12).

621

622 Diniz-Filho, J. A. F., Mauricio Bini, L., Fernando Rangel, T., Loyola, R. D., Hof, C., Nogués-Bravo, D. & Araújo,  
623 M. B. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate  
624 change. *Ecography*, 32, 897-906.

625

626 Dormann, C. F., Schymanski S. J., Cabral J., Chuine I., Graham C., Hartig F., Kearney M., Morin X.,  
627 Römermann C., & Schröder, B. (2012). Correlation and process in species distribution models: bridging a  
628 dichotomy. *Journal of Biogeography*, 39(12), 2119-2131.

629

630 Gallienne, C., & Robins, D. (2001). Is *Oithona* the most important copepod in the world's oceans?  
631 *Journal of Plankton Research*, 23(12), 1421-1432.

632

633 Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149-  
634 161.

635

636 Hattab, T., Lasram F. B. R., Albouy C., Sammari C., Romdhane C. M., Cury P., Leprieur F., & Le Loc'h,  
637 F. (2013). The use of a predictive habitat model and a fuzzy logic approach for marine management and  
638 planning. *PLoS ONE*, 8(10), e76430.

639

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

642

643 Hébert, M-P., Beisner B. E., & Maranger, R. (2016). A compilation of quantitative functional traits for  
644 marine and freshwater crustacean zooplankton. *Ecology*, 97(4), 1069-1081, doi: 10.1890/15-1275

645

646 Hébert, M-P., Beisner B. E., & Maranger, R. (2016). Linking zooplankton communities to ecosystem  
647 functioning: toward an effect-trait framework. *Journal of Plankton Research*, 39, 3-12.

648

649 Helaouët, P., & Beaugrand, G. (2007). Macroecology of *Calanus finmarchicus* and *C. helgolandicus* in  
650 the North Atlantic Ocean and adjacent seas. *Marine Ecology Progress Series*, 345, 147-165.

651

652 Hengl, T., Sierdsema H., Radović A., & Dilo, A. (2009). Spatial prediction of species' distributions from  
653 occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological*  
654 *Modelling*, 220(24), 3499-3511.

655

656 Hirzel, A., Hausser J., Chessel D., & Perrin, N. (2002). Ecological-niche factor analysis: how to compute  
657 habitat-suitability maps without absence data? *Ecology*, 83(7), 2027-2036.

658

659 Hooff, R.C. & Peterson, W.T. (2006). Copepod biodiversity as an indicator of changes in ocean and climate  
660 conditions of the northern California current ecosystem. *Limnology and Oceanography*, 51, 2607-2620.

661

662 Husson, F., Lê S., & Pagès, J. (2010). Exploratory multivariate analysis by example using R, CRC press.  
663 Taylor & Francis Group. United States of America.

664

665 Husson, F., Josse, J., Lê, S. & Mazet, J. (2017). Package 'FactoMineR'. *Multivariate Exploratory Data Analysis*  
666 *and Data Mining*: <http://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>

667

668 Irwin, A. J., Nelles, A. M., & Finkel, Z. V. (2012). Phytoplankton niches estimated from field data. *Limnology*  
669 *and Oceanography*, 57(3), 787-797.

670

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

674

675 Jönsson, B. F., & Watson, J. R. (2016). The timescales of global surface-ocean connectivity. *Nature*  
676 *Communications*, 7, 11239.

677

678 Kiørboe, T. (2011a). What makes pelagic copepods so successful? *Journal of Plankton Research*, 33(5),  
679 677-685.

680

681 Kiørboe, T. (2011b). How zooplankton feed: mechanisms, traits and trade-offs. *Biological Reviews*,  
682 86(2), 311-339.

683

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

686

687 Kishi, M. J., Kashiwai M., Ware D. M., Megrey B. A., Eslinger D. L., Werner F. E., Noguchi-Aita M.,  
688 Azumaya T., Fujii M., & Hashimoto, S. (2007). NEMURO—a lower trophic level model for the North  
689 Pacific marine ecosystem. *Ecological Modelling*, 202(1), 12-25.

690

691 Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the*  
692 *American statistical Association*, 47(260), 583-621.

693

694 Legendre, P., & Legendre, L. (2012). *Numerical ecology*. 3<sup>rd</sup> English Edition. Elsevier Science BV,  
695 Amsterdam.

696

697 Le Quéré, C., Harrison S. P., Colin Prentice I., Buitenhuis E. T., Aumont O., Bopp L., Claustre H.,  
698 Cotrim Da Cunha L., Geider R., & Giraud, X. (2005). Ecosystem dynamics based on plankton functional  
699 types for global ocean biogeochemistry models. *Global Change Biology*, 11(11), 2016-2040.

700

701 Le Quéré, C., Buitenhuis E. T., Moriarty R., Alvain S., Aumont O., Bopp L., Chollet S., Enright C.,  
702 Franklin D. J., Geider R. J., Harrison S. P., Hirst A. G., Larsen S., Legendre L., Platt T., Prentice I. C.,  
703 Rivkin R. B., Salliey S., Sathyendranath S., Stephens N., Vogt M., & Vallina, S. M. (2016). Role of  
704 zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles.  
705 *Biogeosciences*, 13(14), 4111-4133.

706

707 Litchman, E., Ohman M. D., & Kiørboe, T. (2013). Trait-based approaches to zooplankton communities.  
708 *Journal of Plankton Research*, 35(3), 473-484.

709

710 Lynam, C. P., Llope M., Möllmann C., Helaouët P., Bayliss-Brown G. A., & Stenseth, N. C. (2017).  
711 Interaction between top-down and bottom-up control in marine food webs. *Proceedings of the National*  
712 *Academy of Sciences*, 114(8), 1952-1957.

713

714 Mauchline, J. (1998). *The biology of Calanoid copepods*. Academic Press. San Diego, California, USA.

715  
716 Mazzocchi, M., Siokou-Frangou I., Tirelli V., Bandelj V., de Puellas M. F., Örek Y. A., de Olazabal A.,  
717 Gubanova A., Kress N., and Protopapa, M. (2014). Regional and seasonal characteristics of epipelagic  
718 mesozooplankton in the Mediterranean Sea based on an artificial neural network analysis. *Journal of*  
719 *Marine Systems*, 135, 64-80.  
720  
721 Merow, C., Smith M. J., Edwards T. C., Guisan A., McMahon S. M., Normand S., Thuiller W., Wüest R.  
722 O., Zimmermann N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species  
723 distribution models? *Ecography*, 37(12), 1267-1281.  
724  
725 Mouchet, M., Guilhaumon F., Villéger S., Mason N. W., Tomasini J. A., & Mouillot, D. (2008). Towards  
726 a consensus for calculating dendrogram-based functional diversity indices. *Oikos*, 117(5), 794-800.  
727  
728 Nishibe, Y., Takahashi K., Ichikawa T., Hidaka K., Kurogi H., Segawa K., & Saito, H. (2015).  
729 Degradation of discarded appendicularian houses by oncaeid copepods. *Limnology and Oceanography*,  
730 60(3), 967-976.  
731  
732 O'Brien, T. (2005). COPEPOD: A global plankton database. *NOAA Technical Memorandum NMFS-*  
733 *F/SPO-73*, 19.  
734  
735 Paffenhöfer, G-A. (1993). On the ecology of marine cyclopoid copepods (Crustacea, Copepoda). *Journal*  
736 *of Plankton Research*, 15(1), 37-55.  
737  
738 Pomerleau, C., Sastri A. R., & Beisner, B. E. (2015). Evaluation of functional trait diversity for marine  
739 zooplankton communities in the Northeast subarctic Pacific Ocean. *Journal of Plankton Research*, 37(4),  
740 712-726.  
741  
742 Phillips, S. J., Anderson R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species  
743 geographic distributions. *Ecological Modelling*, 190(3), 231-259.  
744  
745 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for  
746 Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>  
747  
748 Razouls C., de Bovée F., Kouwenberg J. & Desreumaux, N. (2005-2017). Diversity and Geographic  
749 Distribution of Marine Planktonic Copepods. Available at <http://copepodes.obs-banyuls.fr/en>.  
750  
751 Reygondeau, G., & Beaugrand, G. (2011). Future climate-driven shifts in distribution of *Calanus*  
752 *finmarchicus*. *Global Change Biology*, 17(2), 756-766.  
753  
754 Robinson, L., Elith J., Hobday A., Pearson R., Kendall B., Possingham H., & Richardson, A. (2011).  
755 Pushing the limits in marine species distribution modelling: lessons from the land present challenges and  
756 opportunities. *Global Ecology and Biogeography*, 20(6), 789-802.  
757

758 Sieburth, J. M., Smetacek V., & Lenz, J. (1978). Pelagic ecosystem structure: heterotrophic compartments  
759 of the plankton and their relationship to plankton size fractions. *Limnology and Oceanography*, 23(6),  
760 1256-1263.

761

762 Siokou-Frangou, I., Christaki U., Mazzocchi M. G., Montresor, M., Ribera d'Alcala M., Vaqué D., &  
763 Zingone, A. (2010). Plankton in the open Mediterranean Sea: a review. *Biogeosciences*, 7(5), 1543-1586.

764

765 The MerMex Group (2011). Marine ecosystems' responses to climatic and anthropogenic forcings in the  
766 Mediterranean. *Progress in Oceanography*, 91(2), 97-166.

767

768 Thuiller, W., Brotons L., Araùjo M. B., & Lavorel, S. (2004). Effects of restricting environmental range  
769 of data to project current and future species distributions. *Ecography*, 27(2), 165-172.

770

771 Turner, J. T. (2002). Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquatic*  
772 *Microbial Ecology*, 27(1), 57-102.

773

774 Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological  
775 pump. *Progress in Oceanography*, 130, 205-248.

776

777 Verity, P., & Smetacek, V. (1996). Organism life cycles, predation, and the structure of marine pelagic  
778 ecosystems. *Marine Ecology Progress Series*, 130, 277-293.

779

780 Villarino, E., Chust G., Licandro P., Butenschön M., Ibaibarriaga L., Larrañaga A., & Irigoien, X. (2015).  
781 Modelling the future biogeography of North Atlantic zooplankton communities in response to climate  
782 change. *Marine Ecology Progress Series*, 531, 121-142.

783

784 Violle, C., Navas M. L., Vile D., Kazakou E., Fortunel C., Hummel I., & Garnier, E. (2007). Let the  
785 concept of trait be functional! *Oikos*, 116(5), 882-892.

786

787 Visser, A. W., Grønning J., & Jónasdóttir, S. H. (2017). *Calanus hyperboreus* and the lipid pump. *Limnology and*  
788 *Oceanography*, 62(3), 1155-1165.

789

790 Woodd-Walker, R. S., Ward P., & Clarke, A. (2002). Large-scale patterns in diversity and community  
791 structure of surface water copepods from the Atlantic Ocean. *Marine Ecology Progress Series*, 236, 189-  
792 203.

793 **Biosketch**

794 **Fabio Benedetti** is currently a post-doc at the Observatoire Océanologique de Villefranche-sur-  
795 Mer where he is studying zooplankton trait biogeography at multiple scales. His main research  
796 interests focus on the co-variation of biodiversity components with environmental gradients and  
797 how these affect ecosystem processes simultaneously.

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

## 801 **Figure Legends**

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

810  
811 **Figure 2:** Functional dendrogram showing the ordination of the 106 copepod species according  
812 to the combination of functional traits. A Multivariate Correspondence Analysis (MCA) based of  
813 the functional traits was performed to ordinate the species in a reduced functional space. The  
814 coordinates of the species along the four retained axes of the functional space were used to  
815 compute a Euclidean distance matrix. Agglomerative hierarchical clustering was used to draw  
816 the functional dendrogram. The chosen cutting-level is shown and the species are coloured  
817 according to the seven derived functional groups.

818  
819 **Figure 3:** Principal Component Analysis (PCA) based on the selected eleven niche  
820 characteristics of the 106 copepod species. The contribution of the eleven niche parameters to the  
821 niche space, as well as their correlation (i.e. the angles between the arrows) are evidenced. The  
822 abbreviations of the environmental predictors are as follows: sea surface temperature (SST), sea  
823 surface salinity (SSS), seasonal range of SST values ( $\Delta$ SST), mixed layer depth (MLD), active  
824 radiation (PAR) integrated over the MLD (MLPAR), and logarithm of surface chlorophyll-a  
825 concentration (logChla).

826  
827 **Figure 4:** Position of the seven functional groups in niche space. Smaller symbols correspond to  
828 the 106 copepod species with the colors and shapes varying according to the functional group  
829 they belong to. Larger symbols correspond to the average position of the functional groups,  
830 based on the coordinates of the species they comprise. The first principal component (PC1)  
831 represents a gradient characterizing oligotrophic conditions on the positive side (higher centers  
832 of SST, SSS, MLPAR) and more productive conditions on the negative side (broader SST and  
833 SSS niches, and higher MLD and logChla centers). The second principal component (PC2)  
834 distinguishes species with broader tolerance to SST, logChla variations (positive side) from  
835 species affiliated to habitats of lesser seasonal variability (negative side). The standard error of  
836 each FG' coordinates along PC1 and PC2 are illustrated with the error bars. The abbreviations of  
837 the environmental predictors are as follows: sea surface temperature (SST), sea surface salinity  
838 (SSS), seasonal range of SST values ( $\Delta$ SST), mixed layer depth (MLD), active radiation (PAR)  
839 integrated over the MLD (MLPAR), and the logarithm of surface chlorophyll-a concentration  
840 (logChla).



841 **Supporting Information**

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

845  
846 **Supporting Information S2:** Table summarising for every 106 copepod species: the number of  
847 observations (presences at the global scale, after re-sampling on the predictors cell grid); the  
848 average TSS value of the multivariate niche models and its standard deviation; the average niche  
849 characteristics derived from the univariate models and their associated standard deviations  
850 computed thanks to the bootstrapping procedure.

851  
852 **Supporting Information S3:** Pair-wise Spearman's correlation coefficient ( $\rho$ ) values of all the  
853 univariate niche characteristics ( $n = 24$ ).

854  
855 **Supporting Information S4:** Maps of the sampling effort (i.e. number of species presences) for  
856 each of the seven copepod functional group defined in the study. Presences are at a  $1^\circ \times 1^\circ$  cell  
857 grid resolution after re-sampling the observed occurrences on the cell grid of the environmental  
858 predictors.

859  
860 **Supporting Information S5:** Distribution of the latitude and the six environmental predictors'  
861 values for each of the seven copepod functional groups defined in the study. Values correspond  
862 to the environmental monthly data that were fitted on each species monthly occurrence.

863  
864 **Supporting Information S6:** Examples of response curves to Sea Surface Temperature (SST)  
865 variations for *Calanus helgolandicus*, *Acartia Acartiura clausi* and *Clausocalanus*  
866 *mastigophorus*, according to nine different types of environmental niche models (ENMs):  
867 Generalized Additive Models (GAM), Generalized Linear Models (GLM), Multi-Adaptive  
868 Regression Splines (MARS), Flexible Discriminant Analysis (FDA), Maximum Entropy  
869 (MAXENT), Boosted Regression Trees (BRT), Random Forest (RF), Classification Tree  
870 Analysis (CTA), and Neural Network Analysis (ANN). Additionally, the niche spaces (based on  
871 a Principal Component Analysis (PCA) performed on the chosen eleven niche characteristics)  
872 and positions of the seven copepod functional groups according to various parameters of the  
873 niche modelling framework designed for the present study: the choice of the ENM, the choice of  
874 the Generalized Additive Models (GAMs) smoothing parameter value ( $k$ ), and the percentage of  
875 the species' monthly occurrence data considered to train (and test) the GAMs.

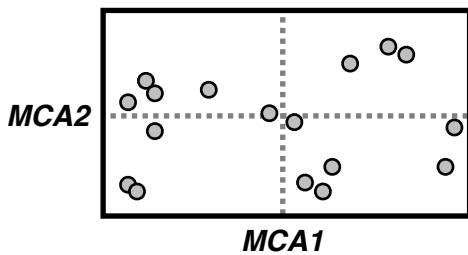
## Functional traits and groups

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

Categorize  
all traits

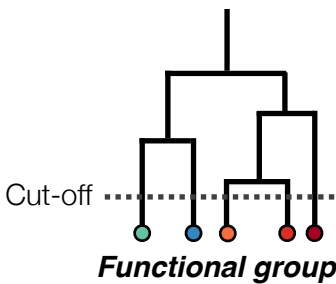
### Multiple Correspondance Analysis (MCA)

*Functional space*



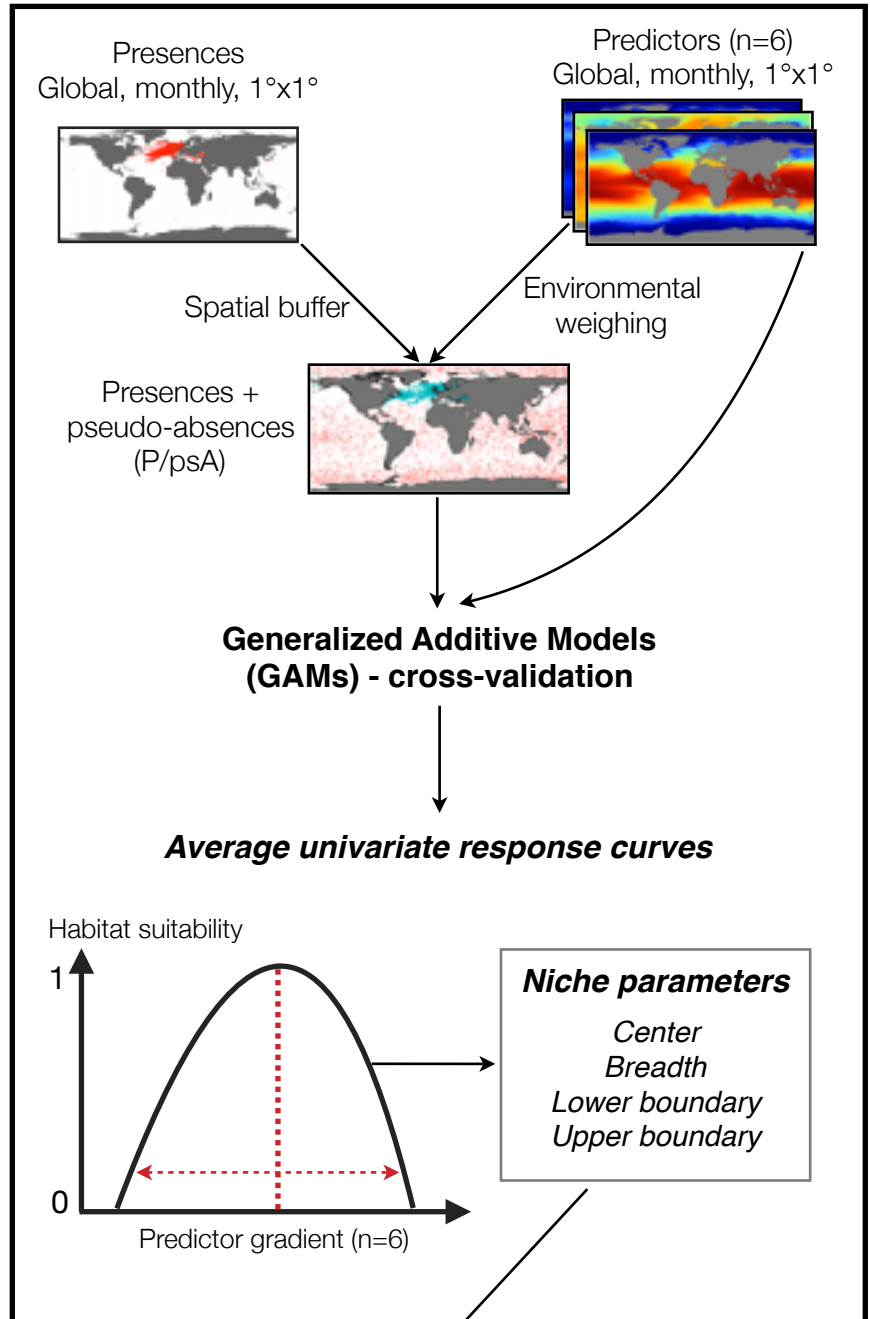
Hierarchical clustering  
on MCA coordinates

*Functional dendrogram*



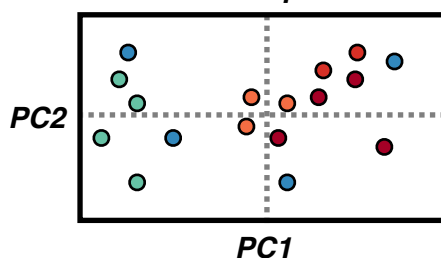
Project functional  
groups in niche space

## Environmental niche modelling



### Principal Component Analysis (PCA)

*Niche space*



Variance analysis  
on PC scores

*Are functional groups  
differently distributed  
in niche space ?*

Parameter filtering  
(Spearman's rank correlation)

