

---

## Scale-dependency of the environmental influence on fish $\beta$ -diversity: Implications for ecoregionalization and conservation

Garcia Jessica <sup>1,\*</sup>, Pelletier Dominique <sup>1</sup>, Carpentier Liliane <sup>1</sup>, Roman William <sup>1</sup>, Bockel Thomas <sup>1</sup>

<sup>1</sup> Unité de Recherche Lagons; Ecosystèmes et Aquaculture Durable en Nouvelle Calédonie (LEAD-NC), Ifremer; Nouméa New Caledonia

\* Corresponding author : Jessica Garcia, email address : [jessicagarcia.mpa@gmail.com](mailto:jessicagarcia.mpa@gmail.com)

---

### Abstract :

#### Aim

To determine the ecoregions (spatial marine areas with similar environmental and physical conditions associated with relatively homogeneous fish assemblages) for shallow reef fish assemblages based on predictive models of beta diversity ( $\beta$ -diversity) that account for both large-scale environmental factors and local habitat characteristics. We assessed the influence of a spatial scale to rank the importance of these factors.

#### Location

New Caledonian (south-west Pacific Ocean, 17–24° S, 158–172° W) Exclusive Economic Zone, Coral Sea Marine Park.

#### Taxon

Fish.

#### Methods

Fish and habitat data that were collected at 13 sites around New Caledonia using unbaited rotating underwater video (285 sampling stations) were analysed. Gradient forest modelling was used to predict the fish  $\beta$ -diversity along the gradients of environmental factors. Ecoregions were obtained by applying clustering methods to gradient forest predictions.

#### Results

The gradient forest models of  $\beta$ -diversity retained 59 species (total: 206 fish species) with  $R^2 > 0$ ,

---

including 19 fish species with  $R^2$  from 0.03% to 69%. For these 19 species, the models explained up to 26% of the variance. At a large scale,  $\beta$ -diversity was significantly explained by nutrient concentrations, sea surface salinity and temperature. Among the eight ecoregions that were delineated based on the  $\beta$ -diversity predictions, three regions corresponded to remote sites under oceanic influence where human pressures are low and the surface nutrient concentrations are high. On the local scale, the benthic habitat explained  $\beta$ -diversity better than the physical and chemical parameters, particularly in the areas subject to anthropogenic pressures.

#### Main conclusions

On the local scale, the respective importance of environmental factors (physical and chemical parameters versus benthic habitat) differed according to ecosystem health. Our findings suggest that nutrient enrichment due to avifauna may have a positive effect on fish  $\beta$ -diversity when an ecosystem is healthy. The ecoregions reflect fish species composition in relation to a large set of environmental parameters.

**Keywords** :  $\beta$ -diversity, ecoregion, gradient forest, marine biodiversity, New Caledonia, scale-dependency, species turnover, underwater video

17 **ACKNOWLEDGEMENTS**

18 We thank the many people who helped with the fieldwork and image analysis, particularly  
19 Gilles Hervé, Delphine Mallet, Kévin Leleu, Nicolas Guilpart, Miguel Clarke, Napoléon  
20 Colombani, Jérémie Drelon, and Matthieu Boureau. Most data were collected over the  
21 AMBIO project funded by both IFREMER, New Caledonia Government and Provinces, the  
22 Conservatoire des Espaces Naturels of New Caledonia, and the French Ministry of Ecology,  
23 the French Initiative for Coral Reefs (IFRECOR), and the French Marine Protected Area  
24 Agency. Part of the data was collected during the PAMPA project and with the help of the  
25 Research Institute for Development.

## 58 INTRODUCTION

59 Beta diversity ( $\beta$ -diversity) describes the spatial variation in species composition at different  
60 spatial and temporal scales, e.g., in relation to environmental gradients (Barton et al., 2013).  
61  $\beta$ -diversity is key for conservation science and understanding ecosystem dynamics, the  
62 scaling of diversity, and the delineation of biotic regions (McKnight et al., 2007).  $\beta$ -diversity  
63 been defined in multiple ways (Anderson et al., 2011). Commonly,  $\beta$ -diversity has two main  
64 components: the spatial turnover of species (species composition, which accounts for species  
65 replacement) and the nestedness of assemblages, which accounts for species loss (more details  
66 in Fig. 1 in Baselga (2010)).  $\beta$ -diversity can be calculated in different ways using  
67 similarity/dissimilarity measures, as summarized in Baselga (2010). These measures are based  
68 on presence/absence data and not abundance data. In this study, we use the term  $\beta$ -diversity to  
69 depict the spatial variation in fish assemblage based on a species abundance matrix. They are  
70 not used here; rather, the term  $\beta$ -diversity is used to depict the spatial variation in the species  
71 composition of an assemblage based on the per species abundance matrix.

72  $\beta$ -diversity is commonly used to define ecoregions, which are relatively homogeneous  
73 geographical units and distinct from one another (Groun, 2009a). Delineating marine  
74 ecoregions is difficult (Lourie & Vincent, 2004; Kulbicki et al., 2013a) because oceans are

75 vast and have no hard boundaries (Baker & Hollowed, 2014). In this paper, an ecoregion is  
76 defined as a spatial marine area with relatively homogeneous species composition that is  
77 clearly distinct from adjacent systems (Spalding et al. 2007). Ecoregions are determined  
78 within the New Caledonia Exclusive Economic Zone (EEZ), and this differs from Spalding et  
79 al. (2007) review paper on marine bioregionalization who defined the term of ecoregions at  
80 larger scales, i.e not restricted to one country. Determining ecoregions requires broad-scale  
81 analytical tools that can fully utilize ecological data sets and fill corresponding data gaps  
82 (Thomson et al., 2014). These requirements are commonly achieved using environmental  
83 surrogates, since many studies have shown a close link between the local benthic habitat and  
84 the spatial distribution and structure of fish communities (McCoy & Bell, 1991; Kulbicki,  
85 1997; Sale, 1998; Ferraris et al., 2005), particularly around live coral (Mumby & Steneck,  
86 2008; Chabanet et al., 2010). Conversely, ecoregions have rarely been determined from fish  
87 assemblages (Koubbi et al., 2011; Lasram et al., 2015). In this paper, ecoregions are  
88 determined from both environmental and physical conditions and fish species composition.  
89 Species distribution models have been developed over the past two decades to model and  
90 predict species distributions, primarily from environmental parameters (Guisan & Thuiller,  
91 2005). Generalized dissimilarity modelling (Ferrier et al., 2007) or gradient forest modelling  
92 (Ellis et al., 2012) can be used to analyse and predict  $\beta$ -diversity based on the entire species  
93 assemblage. In gradient forest models, as in other tree-based methods, explanatory factors are  
94 hierarchized according to their influence on the response. In our paper, gradient forest  
95 modelling is applied to the matrix of species abundances per station to analyse relationships  
96 between the spatial variations in abundances per species and environmental variables.  
97 Three important features should be considered when modelling the spatial variation of  $\beta$ -  
98 diversity and producing biogeographic predictions: i) the nature of the predictors, ii) the  
99 spatial scales and grain (i.e. resolution), and iii) the types of organisms (Barton et al., 2013).

100 Barton et al. (2013) noted that biologists often select spatial scales and grains depending on  
101 the taxon of interest, its general spatial scale of occurrence and the movement or dispersal  
102 capacity of the target species. For example, units were restricted to square centimetres for  
103 bacteria (Martiny et al., 2011), square metres for geographic studies on arthropods (Cardoso  
104 et al., 2009), and square kilometres for birds (Rahbek, 2005). For biogeographic studies,  
105 Spalding et al. (2007) proposed a nested system of spatial units corresponding to realms,  
106 provinces and ecoregions.

107 In terms of conservation and management, historical perceptions and political divisions of  
108 marine areas (e.g., EEZ) may not reflect ecological reality (Lourie & Vincent, 2004), and they  
109 may also reveal the technical constraints and costs encountered by scientists and local  
110 managers, e.g., when monitoring large areas. The spatial scope of our study is the entire New  
111 Caledonia EEZ, which corresponds to a single ecoregion in Spalding et al. (2007). Therefore,  
112 our study is conducted at a more local scale than in the cited paper. The study area is  
113 approximately 1,740,000 km<sup>2</sup> with an extremely long barrier reef (1,600 km long). The  
114 unbaited rotating underwater video (STAVIRO) system (Pelletier et al., 2012) used to sample  
115 fish and habitat in this large area provides sufficiently fine grain sampling. Moreover, this  
116 system is easy to implement and does not require extensive boat or human resources to collect  
117 the data.

118 In this paper, we first quantified the respective influences of environmental parameters (both  
119 physical and chemical parameters and local benthic habitat) on fish species composition.

120 Gradient forest modelling was used to conduct this analysis at three nested spatial scales: i)  
121 New Caledonia EEZ (>1,400,000 km<sup>2</sup>); ii) the intermediate scale of the mainland and  
122 Chesterfield-Bellona reefs (>1,300 km<sup>2</sup>), and iii) the local scale (>100 km<sup>2</sup>).

123 Our second focus in this paper was to delineate ecoregions across spatial environmental  
124 gradients by clustering the gradient forest predictions obtained at the New Caledonia EEZ

125 scale. We hypothesized that the reef fish assemblages would be distinct at the isolated reefs  
126 and lagoons that are characterized by oceanic influence and low human impact. Several  
127 studies suggested that in remote and uninhabited areas, fish assemblages are dominated by  
128 predator species, resulting in an inverted biomass pyramid (Stevenson et al., 2007;  
129 Friedlander et al., 2010; Mourier et al., 2016).

## 130 **MATERIALS AND METHODS**

### 131 **Study area**

132 The study area comprises the entire EEZ of New Caledonia (17-24° S, 158-172° E; Fig. 1),  
133 which is characterized by a variety of oceanic and continental coral reefs (Andréfouët et al.,  
134 2009). This area comprises (i) a main island (16,500 km<sup>2</sup>) that is surrounded by a large, highly  
135 diversified lagoon subject to a range of anthropogenic pressures, particularly those from  
136 Nouméa City; and (ii) remote reefs and atolls under oceanic influence, such as the  
137 Chesterfield-Bellona reefs and the Matthew, Hunter and Walpole Islands, which are well-  
138 preserved areas and host large populations of several bird species (Borsa et al., 2010). A large  
139 number of lagoons in New Caledonia were included on the UNESCO World Heritage List in  
140 2008 (World Heritage Committee, 2008) because of the exceptionally high diversity of their  
141 coral reef ecosystems.

### 142 **Ecological data collection**

143 Benthic habitat and fish were sampled by using an unbaited rotating underwater video  
144 (STAVIRO system) (Pelletier et al., 2012). These cameras were set to record images for 9  
145 min (3 rotations of 3 min each). During each rotation, the camera rotates 60° every 30 seconds  
146 (6 sectors per rotation). The sampling was performed at 13 sites that are representative of the  
147 ecosystems encountered in the Coral Sea Marine Park (CSMP) and on the World Heritage  
148 property, both outside of the coastal lagoon (Chesterfield and Bellona, Matthew, Hunter and  
149 Walpole) and around the mainland (Nouméa, Koné, Pouebo, Hyenghène, Borendy, Merlet,

150 Corne Sud and Ouano) (Fig. 1). We selected a subset of 285 sampling stations that are  
151 exclusively located in hard substrate reef habitats and correspond to depths between 2 and 36  
152 m. The data from the majority of the stations (88%) in the subset were collected between May  
153 2012 and September 2013. Based on the results of Mallet et al. (2016) and Powell et al.  
154 (2016), we selected the 285 sampling stations that were collected between 8 a.m. and 3 p.m.  
155 to reduce the possible short-term temporal variations in fish abundance. At each station, each  
156 fish that was recorded in a 5 m radius around the camera was identified to the finest  
157 taxonomic level (see Pelletier et al. (2012) for a complete description of the procedure). In our  
158 analysis, we retained only individual fish that were identified at the species level. Fish  
159 abundance was estimated as the mean number of fish per rotation, which was averaged over  
160 the three rotations and scaled to a 100 m<sup>2</sup> surface area. Fish abundance data resulting from  
161 image analysis were log-transformed before the analysis to correct for distribution skewness.

162

### 163 **Environmental predictor parameters**

#### 164 *Local habitat*

165 At each station, local habitat descriptors were estimated from the same video footage used for  
166 the fish, following a method derived from Clua et al. (2006) (see Pelletier et al. 2012 for  
167 details). The descriptors consisted of abiotic descriptors (substrate nature and granulometry)  
168 and biotic percent covers (living coral, seagrass and macroalgae). The coral cover categories  
169 included dead coral and live coral, distinguishing massive and encrusting coral, digitate coral  
170 and branching coral. The other habitat categories that were considered in the present study  
171 were seagrass, macroalgae, sand, gravel, rock and slab cover. The topography and complexity  
172 were scored on a scale that ranged from 0 to 5 (Pelletier et al., 2012), while other habitat  
173 categories were percent values. The grain of these local habitat data strictly matched the  
174 resolution of the fish data, i.e., it corresponded to the surface of a disc with a radius of 5 m

175 around the STAVIRO system. The local habitat data formed a non-continuous data layer over  
176 the study area. To obtain a map of  $\beta$ -diversity for the entire area, these data had to be  
177 interpolated at each cell of the grid that was used to model and predict fish  $\beta$ -diversity. The  
178 interpolation of the local habitat data relied on the inverse distance weighting interpolation  
179 (IDW) method (ArcView 10.2 with Spatial Analyst extension, ESRI, Redlands, USA) that  
180 conserved the local variations between sample points (Torres et al., 2008). The interpolation  
181 was limited to depths shallower than 250 m. Although no fish data were collected at such  
182 depths, the 250 m value was retained to allow for the visualization of external slopes on the  
183 map.

#### 184 *Chlorophyll a, physical factors and nutrients*

185 Twenty-one environmental parameters were considered (Table 1). The sea surface  
186 temperature (SST), sea surface salinity (SSS), bathymetry (Bathy), distance to coast (Coast  
187 dist), latitude (NSAsp) and longitude (EWAsp) aspects (30 arc-second) were obtained from the  
188 MARSPEC dataset (Sbrocco & Barber, 2013). For the SST and SSS, we considered the  
189 annual mean (MeanAnsst, MeanAnsst), annual variance (VrAnsst, VrAnsst) and annual range  
190 (RgAnsst, RgAnsst). The maximum and minimum SST values (Maxsst and Minsst) were also  
191 considered. The distance to the coast was computed with respect to the closest island, which  
192 was either the mainland or the Loyalty Islands. Chlorophyll a concentrations were obtained by  
193 using the GES-DISC Interactive Online Visualization and Analysis Infrastructure (Giovanni)  
194 at a 9 km spatial resolution. These data were available for 2007, 2008, 2010 and 2012.  
195 Nutrient concentrations (phosphates, nitrates and silicates) were extracted from NOAA's  
196 National Center for Environmental information datasets (<http://www.noaa.gov/>). The  
197 corresponding ground truth data were either relatively close (<50 km) for remote reefs or  
198 inside the lagoon for mainland sites. We also utilized the nutrient data from the external slope  
199 to characterize the nutrient conditions in this zone. For each nutrient, we retained the surface

200 concentration and the mean concentration between 5 m and 35 m, which were consistent with  
 201 the depth range of the stations. All environmental variables were mapped on a common grid  
 202 of 0.024° in latitude by 0.024° in longitude.

### 203 **Gradient forest modelling**

204 Gradient forest (Ellis et al., 2012) is a non-parametric statistical method that is derived from  
 205 random forest (RF) modelling (Breiman, 2001; Cutler et al., 2007). Based on a machine  
 206 learning algorithm, RF evaluates the importance of environmental predictors for each species  
 207 and determines where significant changes in species composition and abundance occur along  
 208 the environmental gradients. Gradient forest generalizes the RF method by extending the  
 209 analysis to the entire species assemblage across a gradient of environmental factors. Similar to  
 210 RF, gradient forest is statistically robust and involves a large set of regression trees (>1,000  
 211 trees). Each tree is built by successive dichotomic partitioning of sampling station sets, such  
 212 that species abundance is as homogeneous as possible within each set. At each stage,  
 213 partitioning is achieved by selecting a specific split value  $v$  for one of the predictors to reduce  
 214 impurity, i.e., the sum of squared deviations around the set mean (Breiman, 2001; Cutler et  
 215 al., 2007). The partitions are successively divided in this way until a partition results in a  
 216 terminal node. The trees are built from bootstrap samples of the response variable (fish  
 217 abundance), and the nodes are determined by the environmental predictors that better  
 218 discriminate between the tree branches the differences in the tree branches. For each tree, the  
 219 data that are not retained in the bootstrap sample are named out-of-bag (OOB) data; they  
 220 contribute to the cross-validated estimate of the generalization error. In this study, this  
 221 resampling procedure with replacement was repeated for each tree, and the tree-like structure  
 222 was recovered. The goodness-of-fit of the model was measured through the proportion of  
 223 variance explained,  $R_s^2$ , for species  $s$  as follows.

$$R_s^2 = \sum_i (X_{si} - \hat{X}_{si})^2 / (X_{si} - \bar{X}_s)^2$$

224 where  $X_{si}$  is the  $i$ th abundance observation,  $\hat{X}_{si}$  is the OOB prediction, and  $\bar{X}_s$  is the mean  
 225 abundance.

226 Each RF was divided among the predictor variables as a function of their importance,  $I_{sp}$ , and  
 227 among the raw importance values,  $I_{spvt}$ . Thus,  $I_{sp}$  quantifies the contribution of a predictor to  
 228 the goodness-of-fit of the model. In practice,  $I_{sp}$  is the increase in the prediction error when a  
 229 predictor  $p$  is randomly permuted. The contribution of predictor  $p$  to the goodness-of-fit of a  
 230 model for species  $s$  ( $R_{sp}^2$ ) is defined as

$$R_{sp}^2 = R_S^2 I_{sp} / \sum_{p'} I_{sp'}$$

231 In this paper, gradient forest was used to examine the responses of fish assemblages along  
 232 environmental gradients, hierarchize the environmental parameters based on their influence  
 233 on the assemblages and identify the threshold values of the environmental parameters that  
 234 determine the changes in  $\beta$ -diversity described by changes in species composition.  
 235 Such changes may result from nestedness or spatial turnover, as explained by Baselga (2010).  
 236 Nestedness and turnover were calculated following the approach proposed in the cited paper.  
 237 The results showed that 99.6% of the dissimilarity in our dataset was due to turnover, and  
 238 thus, nestedness was negligible in our study. Therefore, most of the variation analysed here is  
 239 considered compositional turnover.

240 The distribution of the  $R^2$  values from all fish species was analysed among the predictors as a  
 241 function of predictor importance,  $I_{sp}$ , and along the gradient of values for each predictor as a  
 242 function of the density of  $I_{spvt}$ . This procedure allowed the  $\beta$ -diversity along the gradient of  
 243 each environmental predictor  $p$  to be determined (see Breiman (2001), Cutler et al. (2007),  
 244 and Ellis et al. (2012) for the full details of the modelling procedures).

245 **Prediction**

246 Where fish data were not available, predictions were obtained at all locations where  
247 environmental covariates were available. To this end, we used the cumulative importance  
248 distributions as empirical functions (see `gradientForest` predict function). Model-predicted  $\beta$ -  
249 diversity values were then ordinated and visualized through principal component analysis  
250 (PCA). At each grid point, a colour was determined from a red-green-blue (RGB) colour  
251 palette based on the first three component coordinates. The outcome was mapped over  
252 geographical space. Out of the thirteen sites in the initial data set, local gradient forest  
253 modelling was performed for eight sites (Chesterfield, Bellona, Merlet, Corne Sud, Pouebo,  
254 Koné, Ouano and Nouméa). These sites were selected as they were characterized by  
255 contrasting local habitats.

256 Gradient forest was applied at i) a large scale, including the entire data set (New Caledonian  
257 EEZ, 1,400,000 km<sup>2</sup>), ii) an intermediate scale by running separate models for the mainland  
258 and for the Chesterfield-Bellona reefs (1,300 km<sup>2</sup>), and iii) a local scale (100 km<sup>2</sup>) by  
259 modelling the data separately for each survey site. Gradient forest modelling was performed  
260 by using R 3.1.2 (`gradientForest` package in R Core Team, 2014). The gradient forest model  
261 outputs for the entire data set were further analysed to define the ecoregions. These outputs  
262 correspond to a prediction of fish species composition in each cell of a grid covering the study  
263 area. The grid was constructed from the coordinates of the environmental variables (36648  
264 grid points). The gradient forest predicted the fish abundance per species at each grid point  
265 where abundance data were not available based on the non-linear response of the fish  
266 abundance data to the gradients in the environmental predictors. This way, the environmental  
267 predictors were transformed to a multidimensional biological space that could be mapped in  
268 geographic space (more details in (Pitcher et al., 2012)). As expressed in other terms by  
269 Leaper et al. (2011), “gradient forest provides a transformation from environmental space into  
270 a space where distance represents compositional differences”.

271 Ecoregions were then obtained by clustering predictions through a K-means partitioning  
272 technique (vegan package in R (Oksanen et al., 2007)). The partitions were determined based  
273 on the simple structure index (ssi) criterion. Therefore, each cluster comprised stations that  
274 had similar patterns in biological community composition, as predicted by the gradient forest  
275 from the environmental parameters. At the regional scale, local habitat cover was visualized  
276 for each site by using radar plots (fmsb package in R; (Nakazawa, 2014)).

### 277 **Spatial autocorrelation**

278 To evaluate the importance of spatial autocorrelation in our results, we used the method of  
279 Leaper et al. (2011). These authors used generalized dissymmetry modelling (GDM) with an  
280 implementation of the Ferrier et al. (2007) GDM link function. We modelled the Bray-Curtis  
281 dissimilarity between the sample stations as a function of the Euclidean distance between the  
282 biologically transformed environmental variables at sampling stations, both with and without  
283 geographic distance between sampling stations.

284 Spatial autocorrelation was found to have little influence on the gradient forest results since  
285 the difference in variance explained by the model when accounting for geographic distance  
286 was small (0.9%). Therefore, we ran the model without accounting for the distances between  
287 stations.

## 288 **RESULTS**

### 289 **Delineation of ecoregions**

290 A total of 206 species, 56 genera, and 28 families of fish were observed at the 285 sampling  
291 stations. The gradient forest model was built from all fish species, but it retained only 59  
292 species with  $R^2 > 0$ , of which 19 fish species had  $R^2$  values that ranged between 0.03% and  
293 69%. For these 19 species, the model accounted for up to 26% of the total variance. These  
294 species in decreasing importance were *Scarus niger*, *Paracanthurus hepatus*, *Siganus*  
295 *doliatus*, *Chlorurus sordidus*, *Lethrinus miniatus*, *Ctenochaetus striatus*, *Scarus ghobban*,

296 *Parupeneus multifasciatus*, *Siganus vulpinus*, *Plectropomus laevis*, *Acanthurus olivaceus*,  
297 *Hemigymnus melapterus*, *Aphareus furca*, *Bodianus axillaris*, *Pterocaesio trilineata*,  
298 *Lethrinus variegatus*, *Carcharhinus amblyrhynchos*, *Chaetodon lunulatus*, and *Ctenochaetus*  
299 *binotatus*. The total variance explained (TVE) was 77% for the physical and chemical  
300 parameters and 23% for the local habitat. At the large scale, the most influential  
301 environmental predictors were phosphate concentrations (surface and 5-35 m), with 15% of  
302 the TVE (Fig. 2a). The next most important factors were the maximum SSS and the annual  
303 variance of temperature, which explained 6% and 5% of the TVE, respectively. Two other  
304 significant parameters were surface silicate (5% of the TVE) and nitrate (5% of the TVE)  
305 concentrations. Local habitat parameters were less influential than the above predictors.

306         The model was used to predict the  $\beta$ -diversity from the 35 environmental and habitat  
307 parameters (Fig. 3a). Eight ecoregions were then determined by clustering the  $\beta$ -diversity  
308 from these predictions (Fig. 3b). Ecoregion 1 included the isolated oceanic reefs of  
309 Chesterfield (the lagoon and western reef), Entrecasteaux, Hunter and small parts of  
310 northwestern Bellona (Fig. 3). This ecoregion was characterized to high surface  
311 concentrations of phosphate and nitrate. (Fig. 4; Table 2). Ecoregion 2 corresponded to the  
312 deep external slopes of the lagoon were subject to a strong oceanic influence (Fig. 3). This  
313 region was the most heterogeneous with large variations in i) surface phosphate, ii)  
314 temperature, iii) salinity, and iv) nitrates (Fig. 4; Table 2).

315 The lagoon that surrounds the main island included ecoregions 3, 8 and 4 (Fig. 4), which are  
316 located on a north-south gradient. Ecoregion 3 corresponded to the large Northern Lagoon far  
317 from the coastline that is open to oceanic influence, while the southern part of this lagoon was  
318 assigned to ecoregion 8. The latter also comprised lagoon areas on the western and eastern  
319 coasts in the northern half of New Caledonia mainland. Compared with ecoregion 3,  
320 ecoregion 8 was characterized by relatively high values in salinity, silicate and nitrates (Table

321 2). Ecoregion 4 corresponded to the most southern area, including half of the lagoon around  
322 the mainland and Pine and Walpole islands. In this ecoregion, nitrates were particularly low  
323 (Fig. 4, Table 2).

324 The other ecoregions corresponded to areas outside the main lagoon. Ecoregion 5 comprised  
325 only the north Bellona lagoon and several reef patches that were south of Chesterfield. It was  
326 characterized by the same nutrient parameters as ecoregion 1 but had a higher salinity (Table  
327 2). Ecoregion 6 exclusively constituted the Loyalty Islands, and it was identified by low  
328 surface concentrations in phosphate and silicate (Fig. 4; Table 2). Ecoregion 7 was  
329 represented only by south Bellona and was on average the furthest region from the coastline.  
330 Overall, at the large scale, the local habitat did not influence much  $\beta$ -diversity prediction and  
331 ecoregion construction and thus appeared as a secondary explanatory factor compared to  
332 nutrients and physical parameters.

### 333 **Influence of the environmental parameters at the intermediate scale**

334 Gradient forest modelling was also run separately for two distinct areas, namely, the mainland  
335 and Chesterfield-Bellona. For the mainland (Fig. 2b), the physical and chemical predictors  
336 explained 66% of the TVE, while the local habitat variables explained 34% of it. Phosphate  
337 (18% of the TVE), salinity (7% of the TVE) and branching coral cover (6% of the TVE) were  
338 the most significant factors that explained  $\beta$ -diversity (Fig. 2c). For Chesterfield-Bellona (Fig.  
339 2c), the local habitat explained 28% of the TVE, whereas the physical and chemical predictors  
340 explained up to 72% of the TVE, in particular, the nitrate concentration between 5 m and 35  
341 m (11% of the TVE), maximum salinity (9% of the TVE), mean temperature (9% of the  
342 TVE), and massive and encrusting coral cover (8% of the TVE).

### 343 **Influence of the environmental parameters at the local scale**

344 Out of the thirteen sites in the initial data set, gradient forest modelling was performed for  
345 eight sites (Chesterfield, Bellona, Merlet, Corne Sud, Pouebo, Koné, Ouano and Nouméa) that

346 were characterized by contrasted local habitats (Fig. 5). The average live coral cover was high  
347 at the Chesterfield (79%), Bellona (48%), Merlet (45%), Corne Sud (47%) and Pouebo (53%)  
348 sites (Fig. 5), where a large diversity of coral forms (massive and encrusting coral, digitate  
349 coral, branching coral, and hard coral) was observed. At these five sites, the local habitat  
350 parameters explained a relatively low proportion of the TVE, with 27% for Chesterfield, 18%  
351 for Bellona, 12% for Merlet, 35% for Corne Sud and 23% for Pouebo. In contrast, the average  
352 live coral cover was lower on the western coast of New Caledonia, i.e., Koné (33%), Ouano  
353 (20%) and Nouméa (20%) (Fig. 5). At these three sites,  $\beta$ -diversity was mainly explained by  
354 the local habitat parameters, with 89%, 50% and 98% of the TVE at Koné, Ouano and  
355 Nouméa, respectively. At Ouano, the macroalgae cover was the main predictor.

## 356 **DISCUSSION**

357 The first objective of this study was to model the relationships between environmental  
358 parameters and fish species composition over a large area encompassing all the reef areas of  
359 New Caledonia, which stand together away from other surrounding islands in the Western  
360 Pacific. By conducting the analysis at several scales, we showed that these relationships and  
361 the respective influences of environmental factors on  $\beta$ -diversity depended on the geographic  
362 scale. To the best of our knowledge, this is the first study considering environmental  
363 parameters from both large-scale data and local information about benthic habitat to model  
364 and understand  $\beta$ -diversity patterns in coral reef ecosystems. This method is important  
365 because, in such ecosystems, lagoon and reef area habitats are very heterogeneous at small  
366 scales, which results in large spatial variations in fish distribution. By accounting for local  
367 variations due to habitat, spatial patterns at larger scales, particularly those linked to gradients  
368 in nutrients and physical parameters, could be better understood. This understanding was also  
369 made possible because data were available for a wide range of benthic habitat conditions with  
370 differences in rocky and sandy substrate covers. These data were an important asset for the

371 second objective of our study, i.e., determining ecoregions at the study area scale using model  
372 predictions obtained at the largest scale. Ecoregions were thus determined from both  
373 environmental and physical conditions and fish species composition. Our model explained  
374 26% of the total variance, which may not seem large but is indeed consistent with results from  
375 other studies using gradient forest modelling, with e.g., 30% (Beazley et al. 2015), 32%  
376 (Booker et al. 2014), 12.4% (Compton et al. 2012) of the TVE by their respective models.  
377 Booker et al. (2014) used an approach by phylum that tends to reduce data variability, and  
378 they studied invertebrate communities, which are less mobile than fish. Commonly, local  
379 habitat structure highly determined richness and abundance of small sedentary reef fishes  
380 (Belmaker et al., 2008; Belmaker, 2009; Messmer et al., 2011). These studies mainly focused  
381 on fish species closely associated with coral (e.g: gobies, damselfishes and juveniles) and at a  
382 really small range. The 19 species retained in the model are mobile within their range, but  
383 most of them are generally considered as living in a restricted reef area (in contrast e.g. to  
384 pelagic species) but still exhibit a larger range than species closely associated with live coral.  
385 Consequently it is likely that the fish species in our study are less sensitive to live coral cover,  
386 one of the local habitat variables, than the species studied in these three studies.

387 A major outcome of our models is that the respective influences of environmental factors on  
388  $\beta$ -diversity depend on the scale of analysis. This finding is discussed below, first through the  
389 link between local habitat and ecosystem health in coastal areas and second through natural  
390 enhancement in nutrients on oceanic reefs. The last part of the discussion addresses the  
391 determination of ecoregions.

392

393 **Hierarchical importance of environmental factors depends on both scale and ecosystem**  
394 **health**

395 In the models per site, the proportion of variance explained (TVE) by environmental factors  
396 strongly differed among the sites, and in particular, this variance ranged from 12% to 98% for  
397 the local habitat factor. The results obtained at the site scale demonstrated that the links  
398 between  $\beta$ -diversity and local habitat (coral, seagrass and macroalgae) depended on the health  
399 of the local ecosystem. Live corals play a crucial role in the functioning of coral reef  
400 ecosystems, and healthy coral reef ecosystems are characterized by a dominance of live coral.  
401 Coral reef ecosystems may be impacted by many drivers and events, including global  
402 warming, ocean acidification, storms, trophic-level dysfunction, eutrophication and other  
403 anthropogenic factors. These ecosystems are vulnerable to sudden events that can induce a  
404 phase shift in the community structure from coral-dominated to algal-dominated systems  
405 (Mumby & Steneck, 2008). Macroalgae biomass hinders and obstructs coral recruitment  
406 (Kuffner et al., 2006) and inhibits coral growth (Tanner, 1995) and fecundity (Foster et al.,  
407 2008). The ecoregions obtained here correspond to three health statuses for coral ecosystems  
408 (Fig. 4): i) atoll reefs with no human impacts that are dominated by live coral (the  
409 Chesterfield and Bellona reefs), ii) reefs closer to the coast that are dominated by live coral  
410 (Merlet, Corne Sud and Pouebo), and iii) coastal reefs where the macroalgae cover was  
411 higher, such as Ouano and Nouméa and, to a lesser extent, Koné.

412 In moderate areas of eutrophication, where coral-dominated ecosystems influenced fish  
413 diversity by increasing phytoplankton biomass and macroalgae, the herbivory biomass was  
414 sufficient to regulate algae production (Burkepile & Hay, 2006). In contrast, on macroalgae-  
415 dominated reefs, a similar eutrophication level explained less species turnover than the local  
416 habitat. This finding was consistent with the fact that in the absence of suitable habitat, fish  
417 abundance and diversity are significantly reduced, particularly for the species that strongly  
418 depend on live coral for settlement habitat or food (Mumby & Steneck, 2008). Airoidi et al.  
419 (2008) showed that anthropogenic disturbances reduced habitat complexity and diversity and

420 resulted in the decline of resident species and food resources, and such environmental  
421 homogenization can greatly impact fish diversity.  
422 Therefore, the influence of nutrients on  $\beta$ -diversity appeared to be dependent on ecosystem  
423 health. Local habitat was a critical factor in explaining  $\beta$ -diversity in areas where  
424 anthropogenic pressures were high, while the nutrient influence was low in these areas. We  
425 hypothesized that in these situations, ecosystem eutrophication is a secondary factor that  
426 influences fish  $\beta$ -diversity. At a local scale, the preservation of a healthy habitat may thus  
427 increase the resistance of an ecosystem to eutrophication, with a moderate nutrient enrichment  
428 (of a natural or anthropogenic origin) having a positive impact on fish diversity in healthy  
429 habitats.

#### 430 **Natural enhancement of nutrients on oceanic reefs**

431 On oceanic reefs, the outcomes of models contrasting remote reefs with sites located within  
432 the mainland lagoon indicated a strong influence of nutrients on fish  $\beta$ -diversity. Phosphate  
433 and nitrate concentrations were particularly high (Fig. 5) in the ecoregions that corresponded  
434 to remote reefs and islands, such as Chesterfield and Bellona, Entrecasteaux, Matthew, and  
435 Hunter. The corresponding nutrient data points were close to the reef. These results are  
436 consistent with the hypothesis of an “island mass effect” (IME) (Gove et al., 2016), which  
437 depicts the increase of phytoplankton biomass near island-reef ecosystems in oligotrophic  
438 tropical oceans. In the lagoons of oceanic atolls, an IME may result from waves and tidal-  
439 driven flushing into enclosed waters (Gove et al., 2016). These isolated reefs and atolls are  
440 uninhabited and have low anthropogenic impacts. However, these locations are home to  
441 abundant sea bird populations that produce large quantities of guano, which is rich in both  
442 organic nitrate and phosphate. In fact, Chesterfield, Bellona and Entrecasteaux were exploited  
443 for guano extraction (Clua et al., 2015) in the 1870s. The phosphate origin may be mineral or  
444 organic (including anthropogenic). Although several studies have been conducted since the

445 beginning of phosphate exploitation at New Caledonian reefs and islands, none of these  
446 studies found evidence of mineral phosphate at these reefs (Delvignier & Jegat, 2001). In  
447 addition to the well-known sources of nutrients, animal waste products (e.g., from sea birds,  
448 reef-associated fish and mobile marine invertebrates) may substantially contribute to the  
449 nutrients in coral reef ecosystems (Gove et al., 2016). McCauley et al. (2012) also observed  
450 the fertilization of soil and an increase in coastal nutrients due to sea birds. Oceanic  
451 ecoregions (i.e., ecoregions 1, 5 and 7) correspond to coral reef ecosystems in very good  
452 health, with high proportions of living coral, high diversities of coral shapes and high fish  
453 diversities. These regions may thus be considered quasi-pristine areas, where continuous  
454 eutrophication naturally occurs from avifauna defecation and where such eutrophication is  
455 likely to be dispersed or absorbed by the coral reef ecosystem. Complex ecological  
456 mechanisms ruling the dynamics of coral reefs represent a major scientific challenge. Our  
457 results are consistent with the hypothesis that where ecosystem health is good, substantial  
458 nutrient enrichment and oceanic current may be integrated by the ecosystem.

459 The results for the Chesterfield and Bellona islands show that terrestrial and marine fauna  
460 were highly linked, which demonstrates the importance of preserving the entire ecosystem.  
461 Many studies showed that anthropogenic activities, i.e., eutrophication, led to strong impacts  
462 on benthic communities with an increase in algal cover, which is detrimental to live coral  
463 (Fichez et al., 2005; Littler et al., 2006; Allgeier et al., 2017). This habitat degradation  
464 impacts a fish assemblage via a rise of detritivores and herbivorous fishes (Khalaf &  
465 Kochzius, 2002) such as scarids, pomacentrids and acanthurids. Eutrophication can also  
466 impact juvenile fish sensitive to chemical cues (Chabanet et al., 1995) or adults by favouring  
467 parasites. In our study, we showed that  $\beta$ -diversity is influenced by multiple factors and could  
468 be inducing a strong impact (positive or negative) on the functions of the coral reef  
469 ecosystem.

## 470 **Ecoregion delineation and implications for conservation**

471 This study provides the first delineation of ecoregions based on both fish species turnover and  
472 environmental parameters at the scale of an entire archipelago encompassing both remote and  
473 coastal areas. Ecoregion identification is important for conservation planning, and in this area  
474 included in the CSMP,  $\beta$ -diversity was identified as a gap in the initial knowledge base  
475 assessment (Ceccarelli et al., 2013). Eight ecoregions were determined from both fish  
476 community data and many environmental parameters, including physical, chemical and  
477 nutrient parameters, as well as local habitat data. The data that were used pertained to areas  
478 with contrasting fish assemblages and natural and anthropogenic pressures, which allowed  
479 better model predictions. The nearly pristine areas that were considered here provided original  
480 information for a better understanding of how environmental factors naturally shape fish  
481 diversity. The first study to define biogeographical delimitations was that of Ekman (1953),  
482 which was based on zoogeographical parameters, environmental barriers and levels of  
483 endemism. Other studies have demonstrated that the presence of natural barriers, such as the  
484 Red Sea land bridge and the Isthmus of Panama, results in biogeographical delimitations  
485 (Bellwood & Wainwright, 2002). Recently, Kulbicki et al. (2013) proposed six large marine  
486 regions in the world based on reef fish assemblage data. These studies brought important  
487 insights into the spatial patterns and dynamics of these ecosystems. In Spalding et al. (2007),  
488 New Caledonia is a single ecoregion. In the present study, ecoregions were determined within  
489 New Caledonia EEZ, and thus, the spatial extent is smaller than in the references above.  
490 However, the definition of an ecoregion remains that of Spalding et al. (2007): “a spatial  
491 marine area with relatively homogeneous species composition that is clearly distinct from  
492 adjacent systems”. The spatial extent of our study corresponds to a well-identified natural  
493 entity in the Western Pacific and applies to a management setting that is relevant for both the  
494 New Caledonian archipelago and the CSMP. Furthermore, this study is located in marine

495 areas that are among the most preserved in the world. The exceptional diversity of coral and  
 496 fish species, healthy populations of large predators, and considerable number and diversity of  
 497 large fish in this area were the main criteria for inscribing 80% of the New Caledonian reef  
 498 and lagoon areas on the World Heritage List (<http://whc.unesco.org/en/documents/115088/>).  
 499 Management and conservation issues always imply actions at several scales (Kittinger et al.,  
 500 2011 for a Pacific case study). The different management types are interconnected at local and  
 501 regional scales, and they rely on the same scientific evidence for policy making; the findings  
 502 of this study provide original knowledge at three distinct scales. The results per site pertain to  
 503 local management by New Caledonian provinces, corresponding to the implementation and  
 504 assessment of marine reserves. The other results are relevant for regional management, and  
 505 they are also useful for putting all the sites into perspective and providing contrast between  
 506 remote sites that can be considered reference health statuses and coastal areas that are subject  
 507 to stronger anthropogenic pressures. This analysis brings new insight into understanding  
 508 human impacts on marine ecosystems by comparing almost pristine areas with urbanized  
 509 coastal areas. At the scale of the lagoon surrounding the main island, our results may  
 510 document conservation planning for the establishment or consolidation of a network of  
 511 marine protected areas including the maximal diversity of species or ecosystems (Growth,  
 512 2009).

513 Table 1 Environmental predictor variables used for the gradient forest modelling.

Parameters	Environmental parameters	Abbrev.	Units	Source
Local habitat	Live coral	--	%	AMBIO project
	Dead coral	--		
	Hard coral	--		
	Massive and encrusting coral	Mass_Enc_Cor		

	Digitate coral	--		
	Branch coral	--		
	Seagrass	--		
	Macroalgae	--		
	Sand	--		
	Slab	--		
	Rocks	--		
	Gravel	--		
	Topography	--	Range: 1-5	
	Complexity	--		
Bathymetry	Depth	Bathy	Metres	MARSPEC
Latitude and longitude aspects	East/west orientation	EWAsp	Radians	
	North/south orientation	NSAsp		
Coastal line	Distance to coastline	Coastal Dist	Kilometres	
Sea surface temperature (SST)	Mean annual surface	MeanAnsst	°C	
	Annual range	RgAnsst		
	Annual variance	VrAnsst		
Sea surface salinity (SSS)	Mean annual surface	MeanAnsSS	Psu	
	Annual range	RgAnsSS		
	Annual variance	VrAnsSS		
	Minimum monthly	MinSSS		
	Maximum monthly	MaxSSS		
Chlorophyll a	Mean chlorophyll 2007	Chloa07	mg.m <sup>3</sup>	Giovanni_NASA
	Mean chlorophyll 2008	Chloa08		
	Mean chlorophyll 2010	Chloa10		
	Mean chlorophyll 2012	Chloa12		
Silicate	Surface	Sil surface	μmol.l <sup>-1</sup>	NOAA's National
Phosphate	Surface	Phos surface		Center for

	Mean 5 to 35 m	Phos 535	Environmental
Nitrate	Surface	Nit surface	Information
	Mean 5 to 35 m	Nit 535	

514

515

516 Table 2 Values of environmental parameters used in gradient forest modelling for each ecoregion.

Ecoregion	Phosphate ( $\mu\text{mol.l}^{-1}$ )		Salinity (psu)			VrAnsst ( $^{\circ}\text{C}$ )	Silicate ( $\mu\text{mol.l}^{-1}$ )	Nitrate ( $\mu\text{mol.l}^{-1}$ )		Distance to coastline
	Surf.	Mean 5-35 m	Mean	Minsss	Maxsss			Surf.	Surf.	
1	0.15± 0.12- 0.15	0.14± 0.11- 0.16	35.1± 35-35.4	34.9± 34.8- 35.2	35.3± 35.17- 35.53	20.1± 17- 27.8	1.55± 1.31-1.7	0.22± 0.04- 0.25	0.14± 0.05- 0.17	15± 1-166
2	0.13± 0.01- 0.16	0.15± 0.1- 0.18	35.3± 35-35.5	35.1± 34.8- 35.3	35.43± 35.17- 35.67	22.3± 16.9-35.1	1.5± 1.11-1.7	0.14± 0.04- 0.25	0.11± 0.04- 0.18	30± 1-278
3	0.12± 0.11- 0.13	0.16± 0.13- 0.18	35.2± 35-35.5	35± 34.9- 35.3	35.32± 35.12- 35.62	26.2± 17.6-28.9	1.39± 1.2-1.66	0.04± 0.036- 0.1	0.04± 0.03- 0.11	20± 1-81
4	0.12± 0.12- 0.13	0.16± 0.14- 0.18	35.2± 35.1- 35.4	35.1± 35-35.2	35.45± 35.27- 35.55	33± 20.2- 37.5	1.32± 1.23-1.57	0.04± 0.04- 0.08	0.04± 0.03- 0.09	7± 1-71
5	0.15± 0.12- 0.16	0.15± 0.13- 0.16	35.3± 35.1- 35.5	35.1± 34.9- 35.3	35.45± 35.29- 35	26.1± 18.4-28.7	1.46± 1.31-1.67	0.23± 0.08- 0.25	0.15± 0.09- 0.17	95± 3-241
6	0.1± 0.1- 0.12	0.11± 0.1- 0.12	35.2± 35.2- 35.4	35.1± 35.1- 35.1	35.53±35.47- 35.67	22.6± 16.8-28.5	1.18± 1.01-1.45	0.052± 0.05- 0.08	0.06± 0.06- 0.1	3± 1-19
7	0.16± 0.15- 0.16	0.15± 0.15- 0.16	35.3± 35.3- 35.4	35.2± 35.1- 35.3	35.49±35.42- 35.56	28± 24.5- 29.4	1.46± 1.31-1.49	0.21± 0.15- 0.25	0.15± 0.14- 0.18	173.8± 9- 275
8	0.12± 0.11- 0.15	0.14± 0.12- 0.16	35.5± 35.3- 35.5	35.2± 35.1- 35.3	35.6±35.45- 35.67	25.6± 17.8-30.5	1.57± 1.34-1.69	0.85± 0.06- 0.25	0.1± 0.06- 0.16	9.5± 1-167

517

518 Figure 1 Location of sites sampled for fish and local benthic habitat.

519

520 Figure 2 Overall conditional importance of the environmental predictors of fish  $\beta$ -diversity on

521 different geographic scales at New Caledonia (a), mainland (b), and Chesterfield-Bellona (c).

522 The local habitat variables are displayed in green (live coral, dead coral, hard coral, massive  
523 and encrusting coral (Mass\_Enc\_Cor), digitate coral, branching coral, seagrass, macroalgae,  
524 sand, gravel, rocks and slab covers).

525

526 Figure 3 a) Mapped gradient forest predictions of fish species turnover in the study area.

527 Insert: Biplot of the biological space of the PCA of assemblage composition with the

528 projection of the main environmental parameters. The colour codes were obtained by

529 matching the colours between the maps and the biplots that display the first factorial plan of

530 the PCA of assemblage composition and the vectors that show the direction and intensity of

531 the seven environmental variables. b) Ecoregions obtained from the clustering of gradient

532 forest outputs. 1: Isolated oceanic reefs, 2: External slopes of the barrier reefs, 3: Remote

533 northern lagoon of the mainland, 4: Southern lagoon of the mainland, 5: South of Chesterfield

534 and north of Bellona, 6: Loyalty Islands, 7: South of Bellona, and 8: Northern lagoon of the

535 mainland.

536 Figure 4 Distribution of nutrient concentrations and physical parameters in each ecoregion.

537 The colour of each boxplot corresponds to the colours of the ecoregions in Fig. 3b. The

538 middle black line in each box corresponds to the median value of the parameter.

539

540 Figure 5 Left side of each plot: Importance of the environmental predictors in explaining the

541 fish  $\beta$ -diversity at the site. Right side of each plot: Radar plots that indicate the relative mean

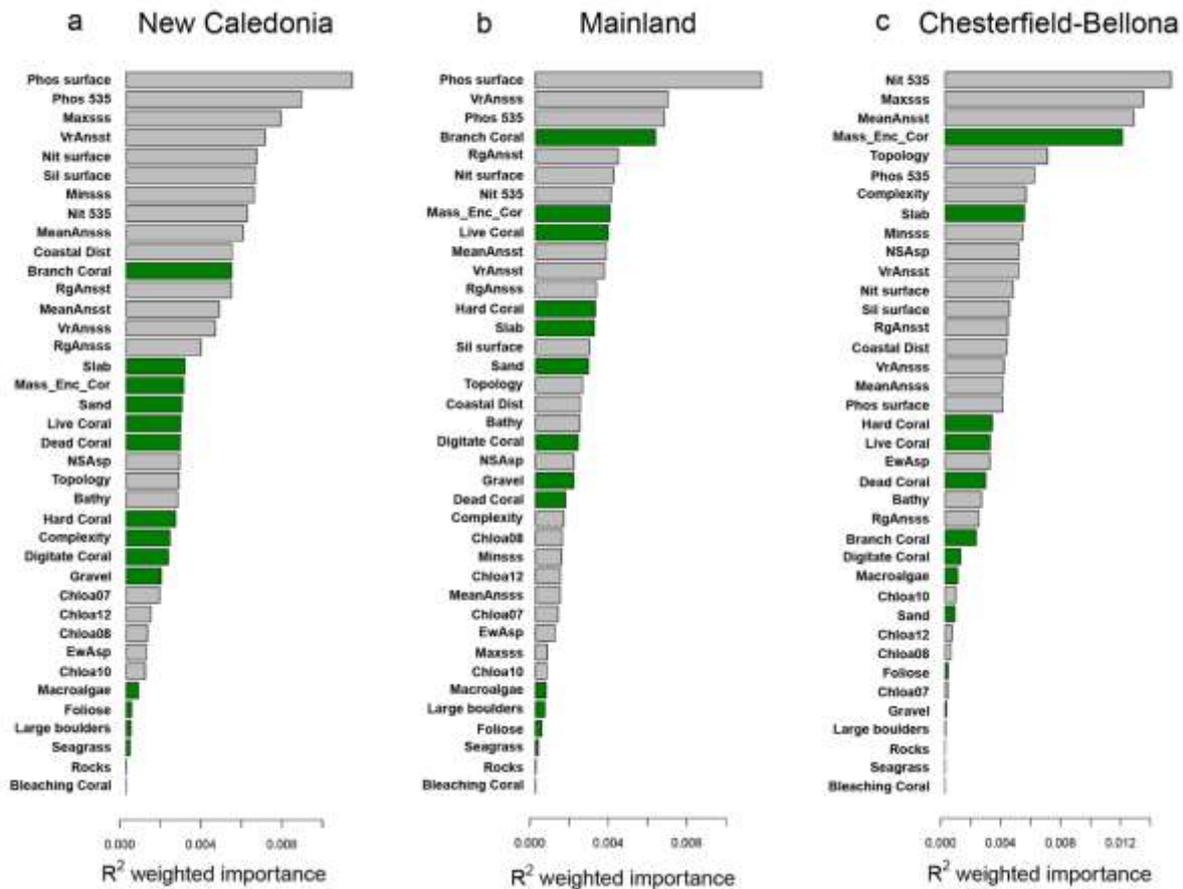
542 value of each local habitat variable, which were obtained by rescaling the variable to a 0% to

543 100% range based on its absolute range of values in the overall data set. This approach

544 enabled the comparison of the thirteen variables across the eight sites.

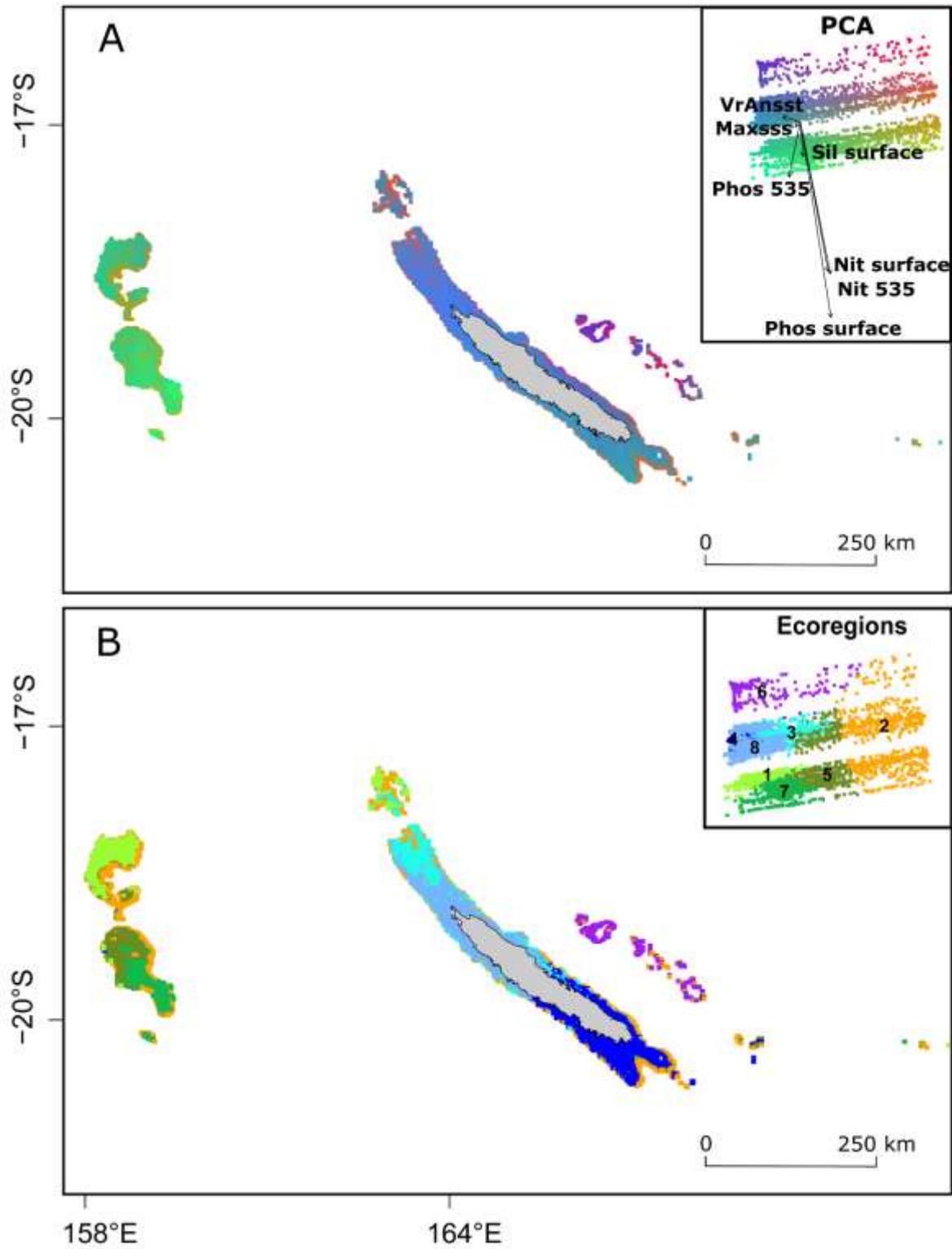
545





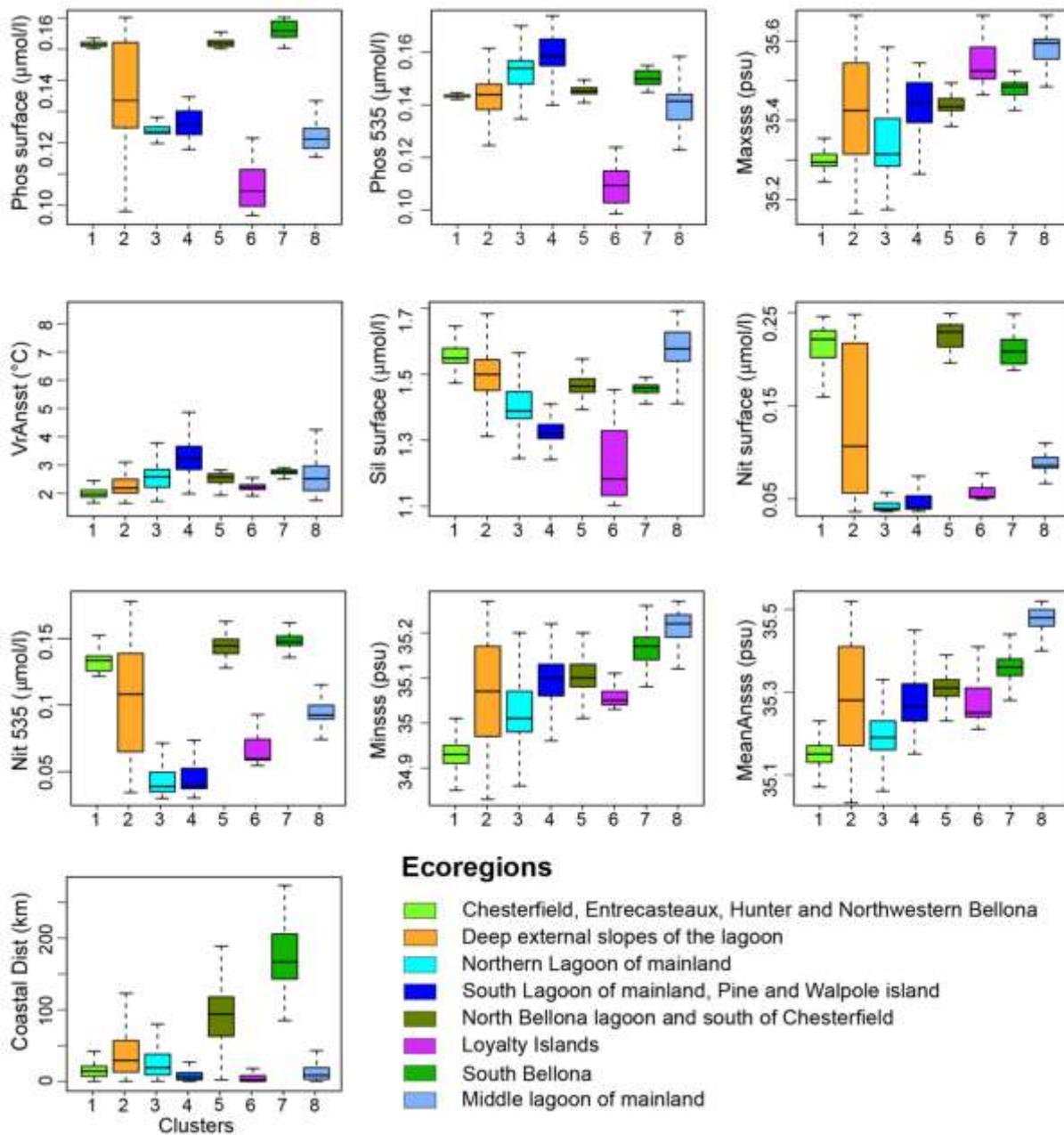
548

549 Figure 2



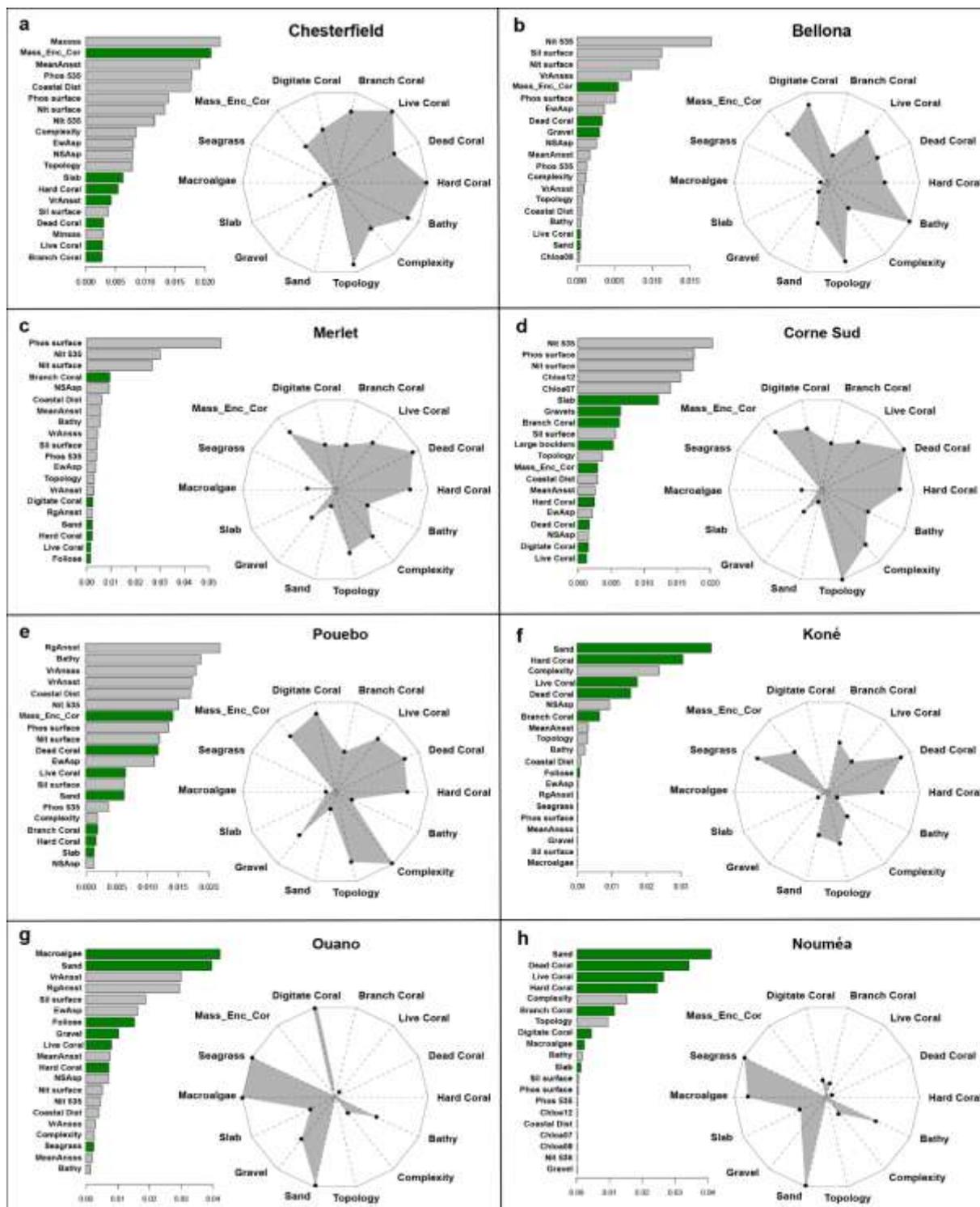
550

551 Figure 3



552

553 Figure 4



554

555 Figure 5

556 REFERENCES

557 Airoidi, L., Balata, D., & Beck, M.W. (2008) The Gray Zone: Relationships between habitat loss and  
 558 marine diversity and their applications in conservation. *Journal of Experimental Marine*  
 559 *Biology and Ecology*, **366**, 8–15.

- 560 Allgeier, J.E., Burkepille, D.E., & Layman, C.A. (2017) Animal pee in the sea: consumer-mediated  
561 nutrient dynamics in the world's changing oceans. *Global Change Biology*, **23**, 2166–2178.
- 562 Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J.,  
563 Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C., & Swenson,  
564 N.G. (2011) Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing  
565 ecologist: Roadmap for beta diversity. *Ecology Letters*, **14**, 19–28.
- 566 Andréfouët, S., Cabioch, G., Flamand, B., & Pelletier, B. (2009) A reappraisal of the diversity of  
567 geomorphological and genetic processes of New Caledonian coral reefs: a synthesis from  
568 optical remote sensing, coring and acoustic multibeam observations. *Coral Reefs*, **28**, 691–  
569 707.
- 570 Baker, M.R. & Hollowed, A.B. (2014) Delineating ecological regions in marine systems: Integrating  
571 physical structure and community composition to inform spatial management in the eastern  
572 Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, **109**, 215–240.
- 573 Barton, P.S., Cunningham, S.A., Manning, A.D., Gibb, H., Lindenmayer, D.B., & Didham, R.K. (2013)  
574 The spatial scaling of beta diversity: Spatial scaling of beta diversity. *Global Ecology and  
575 Biogeography*, **22**, 639–647.
- 576 Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity:  
577 Partitioning beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.
- 578 Bellwood, D.R. & Wainwright, P.C. (2002) The history and biogeography of fishes on coral reefs. *Coral  
579 Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* pp. 5–32. Academic Press,  
580 Boston, MA, USA.
- 581 Belmaker, J. (2009) Species richness of resident and transient coral-dwelling fish responds  
582 differentially to regional diversity. *Global Ecology and Biogeography*, **18**, 426–436.
- 583 Belmaker, J., Ziv, Y., Shashar, N., & Connolly, S.R. (2008) Regional variation in the hierarchical  
584 partitioning of diversity in coral-dwelling. *Ecology*, **89**, 2829–2840.
- 585 Borsa, P., Pandolfi, M., Andréfouët, S., & Bretagnolle, V. (2010) Breeding Avifauna of the Chesterfield  
586 Islands, Coral Sea: Current Population Sizes, Trends, and Threats. *Pacific Science*, **64**, 297–  
587 314.
- 588 Breiman, L. (2001) Random forests. *Machine learning*, **45**, 5–32.
- 589 Burkepille, D.E. & Hay, M.E. (2006) Herbivore vs. nutrient control of marine primary producers:  
590 context-dependent effects. *Ecology*, **87**, 3128–3139.
- 591 Cardoso, P., Aranda, S.C., Lobo, J.M., Dinis, F., Gaspar, C., & Borges, P.A.V. (2009) A spatial scale  
592 assessment of habitat effects on arthropod communities of an oceanic island. *Acta  
593 Oecologica*, **35**, 590–597.
- 594 Ceccarelli, D.M., McKinnon, A.D., Andréfouët, S., et al. (2013) The Coral Sea. *Advances in Marine  
595 Biology* pp. 213–290. Elsevier,
- 596 Chabanet, P., Dufour, V., & Galzin, R. (1995) Disturbance impact on reef fish communities in Reunion  
597 Island (Indian Ocean). *Journal of Experimental Marine Biology and Ecology*, **188**, 29–48.

- 598 Clua, E., Chauvet, C., Mourier, J., Werry, J.M., & Randall, J.E. (2015) Pattern of movements within a  
599 home reef in the Chesterfield Islands (Coral Sea) by the endangered Giant Grouper,  
600 *Epinephelus lanceolatus*. *Aquatic Living Resources*, **28**, 53–58.
- 601 Clua, E., Legendre, P., Vigliola, L., Magron, F., Kulbicki, M., Sarramegna, S., Labrosse, P., & Galzin, R.  
602 (2006) Medium scale approach (MSA) for improved assessment of coral reef fish habitat.  
603 *Journal of Experimental Marine Biology and Ecology*, **333**, 219–230.
- 604 Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007) Random  
605 forest for classification in ecology. *Ecology*, **88**, 2783–2792.
- 606 Delvignier & Jegat (2001) Historique des mines de phosphate en Nouvelle-Calédonie. *Bulletin de la*  
607 *Société de l'Etudes Historiques de la Nouvelle Calédonie*, **128**, 41–80.
- 608 Ekman, S., P. (1953) *Zoogeography of The Sea*. Sidgwick and Jackson, London,
- 609 Ellis, N., Smith, S.J., & Pitcher, C.R. (2012) Gradient forests: calculating importance gradients on  
610 physical predictors. *Ecology*, **93**, 156–168.
- 611 Ferraris, J., Pelletier, D., Kulbicki, M., & Chauvet, C. (2005) Assessing the impact of removing reserve  
612 status on the Abore Reef fish assemblage in New Caledonia. *Marine Ecology Progress Series*,  
613 **292**, 271–286.
- 614 Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007) Using generalized dissimilarity modelling to  
615 analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity*  
616 *and Distributions*, **13**, 252–264.
- 617 Fichez, R., Adjeroud, M., Bozec, Y.-M., Breau, L., Chancerelle, Y., Chevillon, C., Douillet, P., Fernandez,  
618 J.-M., Frouin, P., Kulbicki, M., Moreton, B., Ouillon, S., Payri, C., Perez, T., Sasal, P., &  
619 Thébault, J. (2005) A review of selected indicators of particle, nutrient and metal inputs in  
620 coral reef lagoon systems. *Aquatic Living Resources*, **18**, 125–147.
- 621 Foster, N., Box, S., & Mumby, P. (2008) Competitive effects of macroalgae on the fecundity of the  
622 reef-building coral *Montastraea annularis*. *Marine Ecology Progress Series*, **367**, 143–152.
- 623 Friedlander, A., Sandin, S., DeMartini, E., & Sala, E. (2010) Spatial patterns of the structure of reef fish  
624 assemblages at a pristine atoll in the central Pacific. *Marine Ecology Progress Series*, **410**,  
625 219–231.
- 626 Gove, J.M., McManus, M.A., Neuheimer, A.B., Polovina, J.J., Drazen, J.C., Smith, C.R., Merrifield, M.A.,  
627 Friedlander, A.M., Ehses, J.S., Young, C.W., Dillon, A.K., & Williams, G.J. (2016) Near-island  
628 biological hotspots in barren ocean basins. *Nature Communications*, **7**, 10581.
- 629 Grouns, I. (2009) Differences in bioregional classifications among four aquatic biotic groups:  
630 Implications for conservation reserve design and monitoring programs. *Journal of*  
631 *Environmental Management*, **90**, 2652–2658.
- 632 Khalaf, M.. & Kochzius, M. (2002) Changes in trophic community structure of shore fishes at an  
633 industrial site in the Gulf of Aqaba, Red Sea. *Marine Ecology Progress Series*, **239**, 287–299.
- 634 Kittinger, J.N., Dowling, A., Purves, A.R., Milne, N.A., & Olsson, P. (2011) Marine Protected Areas,  
635 Multiple-Agency Management, and Monumental Surprise in the Northwestern Hawaiian  
636 Islands. *Journal of Marine Biology*, **2011**, 1–17.

- 637 Koubbi, P., Moteki, M., Duhamel, G., Goarant, A., Hulley, P.-A., O'Driscoll, R., Ishimaru, T., Pruvost, P.,  
638 Tavernier, E., & Hosie, G. (2011) Ecoregionalization of myctophid fish in the Indian sector of  
639 the Southern Ocean: Results from generalized dissimilarity models. *Deep Sea Research Part*  
640 *II: Topical Studies in Oceanography*, **58**, 170–180.
- 641 Kuffner, I., Walters, L., Becerro, M., Paul, V., Ritson-Williams, R., & Beach, K. (2006) Inhibition of coral  
642 recruitment by macroalgae and cyanobacteria. *Marine Ecology Progress Series*, **323**, 107–  
643 117.
- 644 Kulbicki, M. (1997) Results of ten years of research (1985-1995) by ORSTOM on diversity, density,  
645 biomass, and community trophic structure of the lagoon and reef fishes in New Caledonia.  
646 *Cybiurn*, **21**, 47–79.
- 647 Kulbicki, M., Parravicini, V., Bellwood, D.R., Arias-González, E., Chabanet, P., Floeter, S.R., Friedlander,  
648 A., McPherson, J., Myers, R.E., Vigliola, L., & Mouillot, D. (2013) Global Biogeography of Reef  
649 Fishes: A Hierarchical Quantitative Delineation of Regions. *PLoS ONE*, **8**, e81847.
- 650 Lasram, F.B.R., Hattab, T., Halouani, G., Romdhane, M.S., & Le Loc'h, F. (2015) Modeling of Beta  
651 Diversity in Tunisian Waters: Predictions Using Generalized Dissimilarity Modeling and  
652 Bioregionalisation Using Fuzzy Clustering. *PLOS ONE*, **10**, e0131728.
- 653 Leaper, R., Hill, N.A., Edgar, G.J., Ellis, N., Lawrence, E., Pitcher, C.R., Barrett, N.S., & Thomson, R.  
654 (2011) Predictions of beta diversity for reef macroalgae across southeastern Australia.  
655 *Ecosphere*, **2**, art73.
- 656 Littler, M.M., Littler, D.S., & Brooks, B.L. (2006) Harmful algae on tropical coral reefs: Bottom-up  
657 eutrophication and top-down herbivory. *Harmful Algae*, **5**, 565–585.
- 658 Lourie, S.A. & Vincent, A.C.J. (2004) Using Biogeography to Help Set Priorities in Marine Conservation.  
659 *Conservation Biology*, **18**, 1004–1020.
- 660 Mallet, D., Vigliola, L., Wantiez, L., & Pelletier, D. (2016) Diurnal temporal patterns of the diversity  
661 and the abundance of reef fishes in a branching coral patch in New Caledonia. *Austral*  
662 *Ecology*, **41**, 733–744.
- 663 Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., & Horner-Devine, M.C. (2011) Drivers of bacterial  $\beta$ -  
664 diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*, **108**,  
665 7850–7854.
- 666 McCauley, D.J., DeSalles, P.A., Young, H.S., Dunbar, R.B., Dirzo, R., Mills, M.M., & Micheli, F. (2012)  
667 From wing to wing: the persistence of long ecological interaction chains in less-disturbed  
668 ecosystems. *Scientific Reports*, **2**, 409.
- 669 McCoy, E.D. & Bell, S.S. (1991) Habitat structure: The evolution and diversification of a complex topic.  
670 *Habitat Structure* (ed. by S.S. Bell, E.D. McCoy, and H.R. Mushinsky), pp. 3–27. Springer  
671 Netherlands, Dordrecht.
- 672 McKnight, M.W., White, P.S., McDonald, R.I., Lamoreux, J.F., Sechrest, W., Ridgely, R.S., & Stuart, S.N.  
673 (2007) Putting Beta-Diversity on the Map: Broad-Scale Congruence and Coincidence in the  
674 Extremes. *PLoS Biology*, **5**, e272.

- 675 Messmer, V., Jones, G.P., Munday, P.L., Holbrook, S.J., Schmitt, R.J., & Brooks, A.J. (2011) Habitat  
676 biodiversity as a determinant of fish community structure on coral reefs. *Ecology*, **92**, 2285–  
677 2298.
- 678 Mourier, J., Maynard, J., Parravicini, V., Ballesta, L., Clua, E., Domeier, M.L., & Planes, S. (2016)  
679 Extreme Inverted Trophic Pyramid of Reef Sharks Supported by Spawning Groupers. *Current*  
680 *Biology*, **26**, 2011–2016.
- 681 Mumby, P. & Steneck, R. (2008) Coral reef management and conservation in light of rapidly evolving  
682 ecological paradigms. *Trends in Ecology & Evolution*, **23**, 555–563.
- 683 Nakazawa, M. (2014) Functions for medical statistics book with some demographic data.  
684 <http://CRAN.R-project.org/package=fmsb>. R package version 0.5.1, .
- 685 Pelletier, D., Leleu, K., Mallet, D., Mou-Tham, G., Hervé, G., Boureau, M., & Guilpart, N. (2012)  
686 Remote High-Definition Rotating Video Enables Fast Spatial Survey of Marine Underwater  
687 Macrofauna and Habitats. *PLoS ONE*, **7**, e30536.
- 688 Pitcher, R.C., Lawton, P., Ellis, N., Smith, S.J., Incze, L.S., Wei, C.-L., Greenlaw, M.E., Wolff, N.H.,  
689 Sameoto, J.A., & Snelgrove, P.V.R. (2012) Exploring the role of environmental variables in  
690 shaping patterns of seabed biodiversity composition in regional-scale ecosystems:  
691 *Environmental shaping of seabed biodiversity*. *Journal of Applied Ecology*, **49**, 670–679.
- 692 Powell, A., Pelletier, D., Jones, T., & Mallet, D. (2016) The impacts of short-term temporal factors on  
693 the magnitude and direction of marine protected area effects detected in reef fish  
694 monitoring. *Global Ecology and Conservation*, **8**, 263–276.
- 695 Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness  
696 patterns. *Ecology Letters*, **8**, 224–239.
- 697 Sale, P.F. (1998) Appropriate spatial scales for studies of reef-fish ecology. *Austral Ecology*, **23**, 202–  
698 208.
- 699 Sbrocco, E.J. & Barber, P.H. (2013) MARSPEC: ocean climate layers for marine spatial ecology:  
700 *Ecological Archives* E094-086. *Ecology*, **94**, 979–979.
- 701 Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge,  
702 M.A., Lombana, A., Lourie, S.A., Martin, K.D., Mcmanus, E., Molnar, J., Recchia, C.A., &  
703 Robertson, J. (2007) Marine Ecoregions of the World: A Bioregionalization of Coastal and  
704 Shelf Areas. *BioScience*, **57**, 573.
- 705 Stevenson, C., Katz, L.S., Micheli, F., Block, B., Heiman, K.W., Perle, C., Weng, K., Dunbar, R., &  
706 Witting, J. (2007) High apex predator biomass on remote Pacific islands. *Coral Reefs*, **26**, 47–  
707 51.
- 708 Tanner, J.E. (1995) Competition between scleractinian corals and macroalgae: An experimental  
709 investigation of coral growth, survival and reproduction. *Journal of Experimental Marine*  
710 *Biology and Ecology*, **190**, 151–168.
- 711 Thomson, R.J., Hill, N.A., Leaper, R., Ellis, N., Pitcher, C.R., Barrett, N.S., & J. Edgar, G. (2014)  
712 Congruence in demersal fish, macroinvertebrate, and macroalgal community turnover on  
713 shallow temperate reefs. *Ecological Applications*, **24**, 287–299.

714 **BIOSKETCH**

715 The AMBIO project aims to provide data and methodologies for a comprehensive assessment  
716 of biodiversity and resources (mostly fish and habitats) in New Caledonia, at both local and  
717 large scales. A large data set of more than 3,000 observations that covers most New Caledonia  
718 reefs and associated habitats has been collected using remote unbaited rotating underwater  
719 video (STAVIRO). The project is conducted in close collaboration with management  
720 practitioners and has a strong capacity-building component.

721 Author contributions: J.G. and D.P. conceived the study. D.P., L.C. and W.R. collected field  
722 data. L.C. and W.R. extracted data from video samples. J.G. and T.B. conducted statistical  
723 analyses of the data. J.G. interpreted the results and wrote the manuscript with the assistance  
724 of D.P.

725 Editor: Luiz Rocha

726