

1 **Body size, reef area, and temperature predict global reef-fish species richness across**
2 **spatial scales**

3 Barneche DR¹, Rezende EL², Parravicini V^{3,4,5}, Maire E^{6,7}, Edgar GJ⁸, Stuart-Smith
4 RD⁸, Arias-González JE⁹, Ferreira CEL¹⁰, Friedlander AM^{11,12}, Green AL¹³, Luiz OJ¹⁴,
5 Rodríguez-Zaragoza FA¹⁵, Vigliola L¹⁶, Kulbicki M^{4,5}, Floeter SR¹⁷

6 ¹School of Life and Environmental Sciences, The University of Sydney, Sydney NSW 2006, Australia

7 ²Departamento de Ecología, Center of Applied Ecology and Sustainability, Facultad de Ciencias Biológicas,
8 Universidad Católica de Chile, Santiago 6513677, Chile

9 ³École Pratique des Hautes Etudes, CRIOBE, USR 3278 CNRS-EPHE-UPVD, Labex Corail, Université de Per-
10 pignan, Perpignan, France

11 ⁴Institut de Recherche pour le Développement, UMR Entropie, Labex Corail, Université de Perpignan, 66000,
12 Perpignan, France

13 ⁵CESAB-FRB, Immeuble Henri Poincaré, Domaine du Petit Arbois, 13857, Aix-en-Provence cedex 3, France

14 ⁶MARBEC, UMR IRD-CNRS-UM-IFREMER 9190, Université Montpellier, Montpellier Cedex, France

15 ⁷Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville
16 Qld 4811, Australia

17 ⁸Institute for Marine and Antarctic Studies, University of Tasmania, Hobart TAS 7001, Australia.

18 ⁹Departamento de Recursos del Mar, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico
19 Nacional, Unidad Mérida, Ant. Carr. A Progreso Km 6, 97310 Mérida, Yucatán, Mexico

20 ¹⁰Departamento de Biología Marinha, Universidade Federal Fluminense, 24001-970 Niterói, RJ, Brazil

21 ¹¹Pristine Seas-National Geographic, Washington, DC, USA

22 ¹²Department of Biology, University of Hawaii, Honolulu, 96822, USA

23 ¹³The Nature Conservancy, Brisbane, Qld 4101, Australia

24 ¹⁴Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin NT 0815, Aus-
25 tralia

26 ¹⁵Departamento de Ecología, CUCBA, Universidad de Guadalajara. Camino Ramón Padilla Sánchez 2100, Za-
27 popan, Jalisco, México. C.P. 45110

28 ¹⁶Institut de Recherche pour le Développement, UMR Entropie, Labex Laboratoire Excellence Récifs Coralliens,
29 98848 Noumea, New Caledonia

30 ¹⁷Departamento de Ecología e Zoología, Marine Macroecology and Biogeography Lab, Universidade Federal de
31 Santa Catarina, 88040-900 Florianópolis, SC, Brazil

32 **Running title:** Scaling of reef-fish species richness

33 **Keywords:** species-energy, neutral theory, biogeography, local and regional diversity, spatial
34 scale, community assembly

35 **Type of article:** Research paper; **Word Count:** 5500 (Text), 317 (Abstract)

36 **Number of figures:** 5; **Number of references:** 48

37 **Correspondence should be sent to:** Diego R. Barneche, School of Life and Environmental
38 Sciences, The University of Sydney, Sydney NSW 2006, Australia (Tel: +61 (2) 8627 0246;
39 Email: barnechedr@gmail.com).

40 **Abstract**

41 **Aim:** To investigate biotic and abiotic correlates of reef-fish species richness across multiple
42 spatial scales.

43 **Location:** Tropical reefs around the globe, including 485 sites in 109 sub-provinces spread
44 across 14 biogeographic provinces.

45 **Time period:** Present.

46 **Major taxa studied:** 2,523 species of reef fish.

47 **Methods:** We compiled a database encompassing 13,050 visual transects. We used hierar-
48 chical linear Bayesian models to investigate whether fish body size, reef area, isolation, tem-
49 perature, and anthropogenic impacts correlate with reef-fish species richness at each spatial
50 scale (i.e. sites, sub-provinces, provinces). Richness was estimated using coverage-based rar-
51 efaction. We also tested whether species packing (i.e. transect-level species richness m^{-2}) is
52 correlated with province-level richness.

53 **Results:** Body size had the strongest effect on species richness across all three spatial scales.
54 Reef area and temperature were both positively correlated with richness at all spatial scales.
55 At the site scale only, richness decreased with reef isolation. Species richness was not corre-
56 lated with proxies of human impacts. Species packing was correlated with species richness at
57 the province level following a sub-linear power function. Province-level differences in species
58 richness were also mirrored by patterns of body size distribution at the site scale. Species rich-
59 provinces exhibited heterogeneous assemblages of small-bodied species with small range
60 sizes, whereas species-poor provinces encompassed homogeneous assemblages composed
61 by larger species with greater dispersal capacity.

62 **Main conclusions:** Our findings suggest that body size distribution, reef area, and tempera-
63 ture are major predictors of species richness and accumulation across scales, consistent with
64 recent theories linking home range to species-area relationships as well as metabolic effects

65 on speciation rates. Based on our results, we hypothesise that in less diverse areas, species
66 are larger and likely more dispersive, leading to larger range sizes and less turnover between
67 sites. Our results indicate that changes in province-level (i.e. regional) richness should leave a
68 tractable fingerprint in local assemblages, and that detailed studies on local-scale assemblage
69 composition may be informative of responses occurring at larger scales.

70 **Introduction**

71 Determining the proximal and ultimate causes of species richness is fundamental to under-
72 stand why some regions can accommodate an extraordinary diversity whereas others contain
73 just a few species (Lessard *et al.*, 2012). Studies have postulated that the dynamics of dis-
74 persal, speciation and extinction over large temporal scales shapes regional pools of species,
75 whereas, locally, species composition is presumably influenced more strongly by local eco-
76 logical processes (e.g. mutualism, competition or predation; Srivastava, 1999; Ricklefs, 2004).
77 However, disentangling how regional versus local processes contribute to community compo-
78 sition and species richness remains challenging (e.g. Cornell & Harrison, 2014), and yet it is
79 crucial to understand why different regions exhibit such dramatic differences in biodiversity.
80 In tropical coral reef systems, for instance, species numbers increase rapidly from small tran-
81 sects to whole islands, culminating in a bewildering regional richness; whereas in more tem-
82 perate rocky reefs new species accumulate at a much slower pace as spatial scale increases, re-
83 sulting in a substantially lower regional richness (Witman *et al.*, 2004; Edgar *et al.*, 2017). We
84 postulate that this pattern reflects intrinsic (e.g. body size and life-history strategies) and ex-
85 trinsic factors (e.g. temperature and geographic barriers), and that their interaction ultimately
86 explains the contrasting levels of species richness and spatial structure across provinces.

87 The dynamics of populations locally and along the meta-community continuum dictate
88 how species accumulate across spatial scales. Population dynamics across space and time are
89 fundamentally determined by resource demands and life history (Peters, 1983), dispersal ca-
90 pacity (Luiz *et al.*, 2013), and local abundance (Reuman *et al.*, 2014), all of which are strong
91 correlates of body size. In reef fishes, smaller species are more abundant (Ackerman *et al.*,
92 2004; Barneche *et al.*, 2016), have smaller home-ranges (Nash *et al.*, 2014), and have more
93 limited geographic range when compared to larger species (Stier *et al.*, 2014). Theory pre-

94 dicts, and evidence suggests that size-correlated traits, particularly abundance, home range,
95 dispersal capacity, and geographic range, should interact to define how the accumulation of
96 species plays out across spatial scales, from small transects to entire biogeographic provinces
97 (Brown & Nicoletto, 1991; Allen & White, 2003; Belmaker, 2009; Reuman *et al.*, 2014).
98 However, we still lack explicit tests of whether body size contributes to the accumulation of
99 reef-fish species richness across spatial scales around the globe (but see Belmaker, 2009 for
100 existing cross-scale comparisons).

101 In addition to body size, environmental and geographical factors are often invoked to
102 explain gradients in species richness. For instance, the Theory of Island Biogeography
103 (MacArthur & Wilson, 1967) states that species richness increases with habitat area, and
104 decreases with degree of isolation. In evolutionary time, the degree of isolation among
105 connected sites (Hubbell, 2001) will directly affect how new species accumulate in space
106 (i.e. from sites to entire provinces), since species ability to disperse is expected to affect gene
107 flow, vicariance, and ultimately speciation rates. Studies have shown that, in ectotherms, body
108 size and environmental temperature can directly affect both dispersal capacity (O'Connor
109 *et al.*, 2007; Luiz *et al.*, 2013) and speciation rates, via effects on individual metabolic rates
110 (Allen *et al.*, 2006). In fact, models that combine thermal effects on speciation rates (Allen
111 *et al.*, 2006) with meta-community dynamics can reproduce realistic latitudinal diversity
112 gradients currently observed in the world's oceans (Tittensor & Worm, 2016; Worm &
113 Tittensor, 2018).

114 Externally to "natural" factors, anthropogenic impacts can influence community composi-
115 tion due to over-harvesting, habitat degradation and introduction of non-native species, con-
116 tributing to multiple processes such as trophic cascades and biotic homogenisation (Jackson
117 *et al.*, 2001; Dornelas *et al.*, 2014). For reef ecosystems in particular, recent studies show that
118 population size of human settlements and accessibility to fish markets can have a negative ef-
119 fect on fish biomass (Cinner *et al.*, 2016). Yet, it is still unknown whether such anthropogenic
120 impacts exhibit consistent effects on species richness across spatial scales, which is important
121 because, at present, extinctions have been observed at small scales but rarely within and across
122 large biogeographic provinces (Kulbicki *et al.*, 2015).

123 To understand how these factors might affect species richness at different scales, here we

124 present a global analysis of how reef-fish species richness builds up in space, from local sites
125 to biogeographic provinces. We test whether body size, reef area (proxy for habitat avail-
126 ability) and isolation, human disturbance, and sea surface temperature (SST) correlate with
127 species richness across spatial scales. Reef fishes provide an ideal model for investigating this
128 problem because they are species rich (Parravicini *et al.*, 2013), globally widespread, and easy
129 to sample with a high level of accuracy. Reef fishes also represent a major food source for
130 > 500 million people (Teh *et al.*, 2013) and vary considerably in body size (Kulbicki *et al.*,
131 2015). We show that, after controlling for the sampling-related group effects, and consistent
132 with recent theoretical predictions (Allen & White, 2003; Reuman *et al.*, 2014; Tittensor &
133 Worm, 2016; Worm & Tittensor, 2018), body size, reef area, and temperature are systemati-
134 cally correlated with species richness and accumulation across spatial scales. In light of our
135 findings, we hypothesise that in less diverse areas, species are larger and likely more disper-
136 sive, leading to larger range sizes and less turnover between sites.

137 **Methods**

138 ***Database and Field sampling***

139 We compiled a global database that encapsulates several decades of field data collection by
140 several of the authors. It encompasses 13,050 belt transects across 485 sites (islands, atolls
141 and coastal contiguous reefs) spread through 14 tropical biogeographic provinces (Tropical
142 Eastern Pacific, Offshore Tropical Eastern Pacific, Easter Island, The Hawaiian archipelago,
143 Polynesia, Central Pacific, SW Pacific, Central Indo-West Pacific, Western Indian Ocean, NW
144 Indian Ocean (Red Sea), Eastern Atlantic, Offshore SW Atlantic, Coastal SW Atlantic and the
145 Caribbean; Fig. 1, Table S1). Sites span a 28-fold difference in species richness (Fig. 1e).

146 Reef-fish assemblages were surveyed through belt transects of different areas depend-
147 ing on the data source (Table S1). Data from Cuba, Bahamas and Belize were collected as
148 presence/absence data. Some sites along the Pacific coast of Mexico (area of 40 m²) and the
149 Caribbean (area of 100 m²) were collected as total abundance counts. At all remaining sites,
150 divers tallied the numbers and body lengths of all fish simultaneously. All transects started
151 and ended approximately at the same depth (within 3 m of depth variation) and were oriented

parallel to the reef. We only utilised transects conducted over hard-reef bottoms in order to make a representative comparison across provinces and reduce methodological bias generated by multi-habitat comparisons (Srivastava, 1999). Sites with less than 3 samples were excluded from the database. A total of 2,523 species were observed across all transects. To quantify overall species richness in these provinces and estimate the potential effects of sampling bias associated with field transects, we contrasted these data against an exhaustive checklist, i.e. a compilation of published lists of species occurrences at multiple sites. Each site-specific list is assembled by combining multiple methods of sampling in order to obtain a thorough assessment of species richness, including that of small-bodied species. This checklist encompasses a total of 5,410 species (Parravicini *et al.*, 2013) (Table S2). Below we explain how we standardised the different transects for the purposes of estimating species richness at multiple scales.

We also compiled information on body size distribution, reef area and isolation, human disturbance (gravity of human settlement and gravity of human market, based on population size and travel time) and sea surface temperature (SST) for each site, sub-province and biogeographic province. Succinctly, species-level maximum adult body size was obtained from the published literature and online databases, and the modal (i.e. the most frequent) size was calculated for each spatial scale (for visualisation purposes, we sometimes divide these estimates into one of six classes following Parravicini *et al.* (2013): 0–7 cm, 7–15 cm, 15–30 cm, 30–50 cm, 50–80 cm and > 80 cm). Reef area (obtained from the Coral Reef Millennium Census Project, and Halpern & Floeter (2008); see “Model predictors” section in the Supplementary Information) was estimated in a 12-km radius around each site, and summed across sites within sub-provinces and sub-provinces within provinces to obtain estimates at higher scales, while reef isolation was calculated as the distance to the coast or the distance to the nearest reef. Human disturbance was estimated, following Maire *et al.* (2016) and Cinner *et al.* (2016), as the ratio between the population of the nearest major market or the nearest human settlement divided by the squared travel time to reach each site (i.e. gravity), averaged for analyses across localities and then provinces. And finally, SST was obtained from daily time-series data from NOAA covering a 5-year period (°C; 0.25° resolution) with the R package *noaaErddap* (available on <https://github.com/dbarneche/noaaErddap>). We describe in de-

182 tail how these variables were estimated, the diagnostic analyses performed to ensure that our
183 statistical models are unbiased, and results robust to different sets of assumptions, in the Sup-
184 plementary Information (see also Table A1).

185 ***Species richness at different scales***

186 For the purposes of our study, we adopted three nested spatial scales: sites, sub-provinces
187 and biogeographic provinces (see Table S1 for geographical coordinates). Sites are defined
188 as small islands or stretches of continuous reefs in larger islands or coastlines (e.g. Arvoredo
189 Island in southern Brazil, or reefs around Noumea in New Caledonia, i.e. ~10's of km). Fol-
190 lowing Edgar *et al.* (2017), we aggregated sites if they were closer than 12 km in linear dis-
191 tance from one another. Sub-provinces encompass sites that belong to the same biogeographic
192 sub-provinces (e.g. Cape Verde, Fiji, Meso-American Barrier in Mexico, i.e. ~100's of km).
193 Finally, biogeographic provinces are well-defined regions (i.e. > 1,000's of km) that follow
194 the recent classification proposed by Kulbicki *et al.* (2013) based on hierarchical analyses of
195 reef-fish species composition. However, we further separated the SW Atlantic into continental
196 (i.e. Brazilian coast) and Offshore (oceanic islands) following Floeter *et al.* (2008).

197 We estimated species richness at the site, sub-province and province scales using coverage-
198 based rarefaction (i.e. sampling completeness). Sampling coverage represents the proportion
199 of the total number of individuals in a community that belong to the species represented in
200 the sample (Chao & Jost, 2012). In contrast to classic sample-based or individual-based rar-
201 efaction methods, this coverage-based sampling technique provides more reliable estimates
202 of species richness across communities. This is because while a fixed area or number of indi-
203 viduals may suffice to represent low-richness communities, it may be insufficient to represent
204 species-rich communities (see details in Chao & Jost, 2012). Calculations were done using the
205 R package *iNEXT* version 2.0.12 (Hsieh *et al.*, 2016). Species richness (i.e. Hill number qD
206 = 0) was estimated by transforming all transect information into presence/absence data. This
207 procedure, concomitantly with the hierarchical models accounting for differences in sampling
208 method described below (see *Statistical analyses*), ensured that all species richness estimates
209 are directly comparable. Estimates at the site, sub-province, and province scales in these mod-
210 els were based on sampling coverages of 0.83, 0.89, and 0.98, respectively, that correspond

211 to the highest coverage values yielding robust, unbiased estimations across all spatial scales
212 according to the package's algorithm (Chao & Jost, 2012; Hsieh *et al.*, 2016). In other words,
213 with this approach, we minimise much of the potential variation in species richness estimates
214 due to differences in relative abundance or coverage sampling.

215 **Statistical analyses**

216 At the site and sub-province scales, we test whether modal body size, B , reef area, A , distance
217 to coast, C , distance to nearest reef, R , gravity of human markets, M , gravity of human settle-
218 ment, H , and temperature, T , explain reef-fish species richness, S , using a hierarchical linear
219 model,

$$\ln S = (\beta_0 + \Delta_g) + \beta_B \ln B + \beta_A \ln A + \beta_C \ln C + \beta_R \ln R + \beta_M \ln M + \beta_H \ln H + \beta_T T + \varepsilon, \quad (1)$$

220 where β_0 is the estimated model intercept, $\beta_{[B,A,C,R,M,H,T]}$ are fixed-effect slopes, Δ_g represent
221 random deviations from the model intercept (on log scale) attributable to a grouping random
222 variable that represents a higher spatial scale – i.e. sub-provinces nested in provinces for the
223 site scale, and provinces for the sub-province scale – and ε represents the residual variation.

224 Note that by including Δ_g we are able to control for other unmeasured variables that might
225 otherwise lead to correlated residuals at higher spatial levels (e.g. for spatial autocorrelation or
226 sampling effects given that sub-provinces were generally studied by the same research group).

227 Accordingly, the model residuals obtained at the site scale, controlling for the effects of sub-
228 province nested within provinces, were not spatially autocorrelated according to Moran's I
229 test ($P = 0.09$). Therefore, our hierarchical model adequately controls for the spatial structure
230 of our sites. We note that species richness is functionally dependent on sampling area, and
231 that perhaps our results could be biased by not including this covariate in the model. In the
232 Supplementary Information we show that our main findings with regards to the effects of body
233 size, reef area, and temperature are maintained even after including total sampling area as a
234 covariate (Table A4).

235 At the province scale, due to the small number of independent samples ($n = 14$), we fit sep-

236 arate linear regressions of the form

$$\ln S = \beta_0 + \beta_B \ln B + \varepsilon. \quad (2)$$

237 In each regression, we respectively used modal body size, reef area, and temperature as pre-
238 dictors because they were the only variables with consistently strong effect sizes (i.e. slopes)
239 both at the site and sub-province scales (see *Results* section below). Given that we only have
240 14 observations at this scale, adding multiple covariates at once would overfit the model.

241 We fit eqns 1 and 2 in a Bayesian framework by using the R package *brms* to derive pos-
242 terior distributions and associated 95% credible intervals (CIs) for the fitted parameters. Both
243 fixed ($\beta_{[B,A,C,R,M,H,T]}$) and random (Δ_g) effects were assigned normally-distributed priors that
244 were vague (i.e. locally uniform over the region supported by the likelihood), with means of
245 zero. Model residuals (ε), and standard deviations for both the random effects ($\sigma [\Delta_g]$) and
246 model residuals ($\sigma [\varepsilon]$), were also assigned weakly informative priors following a Student-*t*
247 distribution. The posterior distributions of model parameters were estimated using Markov
248 chain Monte Carlo (MCMC) methods by constructing three chains of 2,000 steps, including
249 1,000-step warm-up periods, so a total of 3,000 steps were retained to estimate posterior dis-
250 tributions (i.e. $3 \times (2,000 - 1,000) = 3,000$). Fixed effects were considered statistically sig-
251 nificant if their posterior 95% credible intervals did not overlap zero. We use Bayesian R^2 in
252 order to estimate the amount of explained variation of each model (Gelman *et al.*, 2017). Pos-
253 terior predictive checks for all three models are provided in the Supplementary Information
254 (Fig. S1).

255 We also tested whether reef-fish species richness at the transect scale (i.e. species packing
256 = species m^{-2}) was correlated with richness at the province level. To do so, we used a two-
257 step approach. First, we ran a Bayesian hierarchical linear model with richness at the transect
258 scale as a function of transect area on a log-log scale. We do so given the non-linear nature
259 of the species-area relationship (Rosenzweig, 1995). Sites nested in sub-provinces nested in
260 provinces were used as a random effect. Model fitting (number of MCMC samples, chains,
261 warm-up periods, and prior distributions) follow the same as the model in eqn 1. Because the
262 random-effect coefficients represent deviations from the estimated model intercept, at each

263 one of the spatial scales we summed each random-effect coefficient with the model intercept
264 in order to obtain a measure of species packing (species m^{-2}). Second, we ran ordinary least
265 squares for each of the 3,000 posterior samples from our random-effect coefficients at the
266 province scale (i.e. species packing on the log scale) against province-level species richness
267 (also on the log scale) – as estimated using coverage-based rarefaction (coverage = 0.98). This
268 approach allowed us to estimate a mean slope between species packing and province-level
269 richness, and R^2 values with associated 95% credible intervals. A slope of 1 would indicate
270 that species packing is directly proportional to province-level richness (i.e. a linear relation-
271 ship), whereas a slope > 0 and < 1 would indicate that species packing scales sub-linearly with
272 province-level richness.

273 All analyses were repeated removing the smallest size class ($< 7 \text{ cm}$) to circumvent po-
274 tential problems of under-detection during sampling (Table A2 in the Supplementary Infor-
275 mation). We also fit the model at the site scale separately for the Atlantic and Americas as a
276 group (i.e. including the tropical eastern Pacific), and the rest of the Pacific as another group
277 (Table A6 in the Supplementary Information).

278 Results

279 Site scale

280 Our model explained 73.6% of the variation in reef-fish species richness across sites (eqn 1),
281 which ranged from 10 to 288 species. After controlling for all covariates, reef-fish species
282 richness decreased with modal maximum body size (Fig. 2a). After binning species richness
283 based on maximum body sizes, we observe that small-bodied species prevail in species-rich
284 sites, particularly in the Pacific region (Figs. 3b and 3c). We observed that the rate of in-
285 crease in biodiversity with sampled area (i.e. the number of transects) is substantially higher
286 at species-rich provinces (Fig. 4a). However, species packing increased sub-linearly with
287 province-level richness (slope = 0.26, 95% C.I. = 0.15 – 0.36; R^2 = 0.55, 95% C.I. = 0.28 –
288 0.78; Fig. S2). Moreover, the observed modal body size is consistently smaller (i.e. low mean
289 value and lower variances, Fig. 4b-c) within the richest provinces. In particular, the estimated
290 slope for the mode of the body size distribution was the strongest among all estimated slopes

291 (β_B : -0.70; 95% C.I.: -0.87– -0.51), implying a 74.0% decrease in richness over the observed
292 range of modal body sizes (from 67 cm to 9.65 cm).

293 Species richness increased with reef area (β_A : 0.05; 95% C.I.: 0.02–0.08; Fig. 2d), and
294 decreased with distance to the closest reef (β_R : -0.02; 95% C.I.: -0.03– -0.00). Our model also
295 indicated that species richness increases with SST (β_T : 0.05; 95% C.I.: 0.01–0.09; Fig. 2g),
296 implying a 49.3% increase in richness going from 21.3°C to 29.73°C.

297 Variables related to human disturbance (gravity of markets and gravity of nearest human
298 settlement) and distance to coast were not significantly correlated with species richness
299 (respectively β_M : 0.00; 95% C.I.: -0.02–0.02; β_H : 0.00; 95% C.I.: -0.01–0.01; β_C : 0.02;
300 95% C.I.: -0.00–0.04). The hierarchical structure of our model, with sub-provinces nested in
301 provinces as our random-effect group, was able to capture considerable variation in species
302 richness at different scales. Particularly, after accounting for the fixed effects, we observed a
303 1.55-fold variation among sub-provinces on average (i.e. $\approx e^{0.22 \times 2}$ with $\sigma_{\Delta_g} = 0.22$).

304 The direction and significance of fixed-effect coefficients were all similar after removing
305 the smallest size class (< 7 cm; Table A2 in the Supplementary Information). Results for the
306 Atlantic + Americas, and the rest of the Pacific show that the mean model coefficients and
307 their directions (i.e. positive or negative) are maintained within the Pacific, but are not signifi-
308 cant within the Atlantic + Americas (Table A6 in the Supplementary Information).

309 *Sub-province scale*

310 At the sub-province scale, reef-fish species richness ranged from 15 to 364 species. Our
311 model (eqn 1) explained 78.8% of the variation in species richness. Consistent with the model
312 at the site scale, after controlling for all covariates, reef-fish species richness decreased with
313 modal maximum body size (Fig. 2b). Species turnover moving from sites to sub-provinces
314 was most pronounced for small-bodied species (Fig. S3). The modal-size estimated slope
315 was again the strongest among all estimated slopes (β_B : -0.77; 95% C.I.: -1.11– -0.43). Apart
316 from modal maximum body size, reef area (Fig. 2e) and temperature (Fig. 2h) were the only
317 other significant variables in our model (β_A : 0.08; 95% C.I.: 0.04–0.12; β_T : 0.05; 95% C.I.:
318 0.00–0.11). Our model indicates that after accounting for the fixed effects, species richness
319 varies by 2.41-fold among provinces (i.e. $\approx e^{0.44 \times 2}$ with $\sigma_{\Delta_g} = 0.44$).

320 After the removal of the smallest size class, modal body size and temperature were still sig-
321 nificant and in the same direction as the main model. However, distance from the nearest reef
322 was not significant although it had the same mean negative direction (β_R : -0.01; 95% C.I.:
323 -0.04–0.02; Table A2 in the Supplementary Information).

324 *Province scale*

325 Consistent with our findings at the site and sub-province scales, body size (eqn 2) explained
326 61.2% of the variation in species richness (ranging from 44 to 806 species) at the province
327 scale. Particularly, province-level richness decreased with increasing modal maximum body
328 size (β_B : -2.79; 95% C.I.: -4.14– -1.43; Fig. 2c). Similar effects were detected after remov-
329 ing the smallest size class (β_B : -2.68; 95% C.I.: -4.04– -1.35; Table A2 in the Supplementary
330 Information). We also analysed the relationship between species richness and area (Fig. 2f)
331 or temperature (Fig. 2i) in two separate linear regressions (Table A5 in the Supplementary In-
332 formation), and both variables were positively correlated with species richness. Reef area was
333 positively correlated with richness (β_A : 0.22; 95% C.I.: 0.08–0.37; Fig. 2f), and explained
334 46.5% of the variation in species richness. Temperature was also positively correlated with
335 richness (β_T : 0.31; 95% C.I.: 0.00–0.62; Fig. 2f), and explained 30.7%.

336 **Discussion**

337 Our results indicate that reef-fish modal body size, reef area, and environmental temperature
338 are consistently correlated with reef-fish species richness across spatial scales (Fig. 2, Ta-
339 ble A5, Supplementary Information). The fact that body size was negatively correlated with
340 species richness at all spatial scales may suggest that the observed accumulation of species
341 across space is strongly associated with smaller body sizes (Fig. 3). This interpretation is con-
342 sistent with previous studies showing that beta diversity across scales is driven by the accu-
343 mulation of smaller species in mammals (e.g. Brown & Nicoletto, 1991) and reef fishes (e.g.
344 Belmaker, 2009). Moreover, our empirical results lend support to the ‘marine diversity spec-
345 trum’ theory proposed for pelagic marine ecosystems (Reuman *et al.*, 2014), whereby the neg-
346 ative linear relationship between richness and body size arises due to the size dependence of
347 home range and dispersal capacity (Allen & White, 2003; Reuman *et al.*, 2014) that has been
348 previously observed in reef fishes (Luiz *et al.*, 2013; Nash *et al.*, 2014).

349 We note that sites that are rich in smaller-bodied reef fishes also encompass a range of suit-
350 able, heterogeneous habitats (e.g. hard and soft corals, sponges, algae; Messmer *et al.*, 2011),
351 and our results indicate that increasing habitat area will also yield higher species richness
352 across scales. We note that our analysis cannot resolve the causality of these relationships
353 given that shallow-reef organisms (e.g. reef-building corals and reef fishes) show a parallel
354 history of diversification through time (Bellwood *et al.*, 2017). However, the fact that body
355 size varies with richness even within a gradient of species-rich sites and provinces (e.g. from
356 Indo-West Pacific to Polynesia) suggests that the evolutionary processes associated with the
357 diversification of reef fishes have aided speciation within smaller-bodied species. In partic-
358 ular, the centre of origin and accumulation hypotheses states that the richer Indo-West Pa-
359 cific is a product of a series of events that facilitated both the speciation (“cradle”) as well as
360 the accumulation (“museum”) of existing species through time (Bellwood *et al.*, 2017). Our
361 findings could be consistent with this hypothesis by noting that Indo-West Pacific reefs have
362 gone through a series of sea-level changes leading to temporally-variable geographic isolation
363 (Bellwood *et al.*, 2017). Therefore, speciation may have been further promoted by vicariance
364 because small-bodied species have, on average, lower realised dispersal capacity (Luiz *et al.*,
365 2013).

366 After accounting for the effect of sampling area, species richness at the transect scale
367 scaled sub-linearly with province-level richness (Fig. S2). The sub-linear nature of this
368 relationship corroborates the suggestion of recent studies that species richness estimated
369 at small spatial scales might not necessarily resemble that at the province scale (Vellend *et*
370 *al.*, 2013; Dornelas *et al.*, 2014). We speculate that this observed phenomenon reflects the
371 interaction between body size, local abundance and geographic range (Fig. S4a), since smaller
372 species are generally more abundant (Ackerman *et al.*, 2004; Barneche *et al.*, 2016) and
373 often exhibit smaller geographic range sizes than larger species (Belmaker, 2009; Luiz *et al.*,
374 2013). Consequently, it is expected that species at the extremes of the body size distribution
375 are detected in only a small fraction of transects: for small species this may relate to smaller
376 home range sizes and difficulties in visual detection, whereas for large species low abundance
377 should result in a small representation across those transects within their geographic range
378 (*Preston’s veil*; Preston, 1948). Our data strongly support this interpretation, with both

379 the smallest and largest size classes being less frequently sampled in nearly all provinces,
380 except in Easter Island (Fig. S4b). Accordingly, while population abundance was negatively
381 correlated with body size across sites in a subset of the data used here (Barneche *et al.*, 2016),
382 the number of provinces in which each species was detected was positively correlated with
383 their maximum size (Fig. S4c), indicating that smaller species tend to have smaller geographic
384 ranges (Fig. S4d).

385 Because of the interactions between size, abundance and distribution, species of intermedi-
386 ate body size constitute the core of most local assemblages in all provinces (Fig. S4b). How-
387 ever, body size distributions also differ significantly among provinces concomitantly with
388 species richness. The linear regression between total species richness against modal body size
389 at the province scale (eqn 2b, Figs. 2c and 5a) gives rise to a continuum, with species-poor
390 peripheral provinces of the Atlantic falling at one extreme and highly complex reefs such as
391 those in the Indo-West Pacific, Polynesia and, to a lesser degree, the Caribbean at the other
392 (Fig. 5a). Thus, the body size distribution within local assemblages is indicative of the over-
393 all reef-fish richness of their respective provinces. This can be readily understood because
394 medium to large fishes are expected to contribute to species richness at the transect level,
395 without a proportional contribution to species richness at higher spatial scales, if they are
396 widely spread (see also Soininen *et al.*, 2007; Reuman *et al.*, 2014). As a consequence, in
397 provinces where larger fishes predominate, local assemblages should be more homogeneous
398 in their composition and exhibit body size distributions shifting towards higher values (Fig.
399 5b). The opposite is expected in provinces occupied predominantly by small species, which
400 include a multitude of species with small geographic ranges. Again, this implies that the enor-
401 mous diversity of species-rich provinces emerges primarily from the accumulation of smaller
402 species with restricted distribution (Fig. 5b).

403 Our models at all scales indicate that temperature was positively correlated with species
404 richness, consistent with the idea that the latitudinal diversity gradient is in part driven by
405 the temperature dependence of ectothermic metabolism and speciation rates (Allen *et al.*,
406 2006; Tittensor & Worm, 2016). Additionally, our results add to a vast number of studies that
407 support predictions from the Theory of Island Biogeography (MacArthur & Wilson, 1967),
408 whereby species richness increases with increasing habitat area, and decreases with isola-

409 tion. For instance, it has been shown recently that smaller-bodied reef fish prey declines more
410 rapidly with increasing isolation when compared to larger predators across Pacific reefs (Stier
411 *et al.*, 2014). We note though that reef isolation was not significant at the sub-province scale.
412 This discrepancy between scales could be due to our choice of measurement for isolation
413 (mean distance to the nearest reef across sites). Given this caveat, our results are consistent
414 with recent theoretical predictions (Tittensor & Worm, 2016; Worm & Tittensor, 2018) which
415 suggest that the combined effects of area, traits in the species pool (e.g. small body size), and
416 warmer temperatures will yield higher beta diversity across spatial scales, perhaps by promot-
417 ing higher speciation rates over time, and higher species turnover across space (MacArthur &
418 Wilson, 1967; Allen *et al.*, 2006; Tittensor & Worm, 2016; Worm & Tittensor, 2018).

419 After accounting for the effects of reef-fish body size, reef area and isolation, temperature,
420 sampling artifacts, and other unmeasured potential biogeographical effects, human-related dis-
421 turbance metrics did not show any significant effect on reef-fish species richness at the site
422 and sub-province scales. This finding is particularly relevant because it suggests that, at least
423 looking at the studied spatial scales, species richness, contrary to assemblage size structure
424 and biomass (e.g. Jackson *et al.*, 2001; Cinner *et al.*, 2016), is not explained by the “gravity”
425 variables, which are based on population size and travel time. This may in large part be a re-
426 sult of the relatively low proportion of species that are exploited, with the rich array of small
427 fishes less likely to only be affected by humans through indirect mechanisms (such as habi-
428 tat alteration). It may also be attributable to the fact that our data collection is recent in time,
429 and it is possible that the majority of sampled reefs are already affected by human disturbance
430 to some degree. However, our results might suggest that habitat degradation associated with
431 loss of coral species in species-rich provinces may have profound impacts on species richness
432 (Alvarez-Filip *et al.*, 2009) via its impact on smaller species. On the other hand, larger fishes
433 might be able to disperse more easily towards other areas most likely because of their better
434 capacity to survive and establish reproductive populations (Luiz *et al.*, 2013). Comparisons
435 between Pacific provinces, the Tropical Eastern Pacific and the Caribbean support this conjec-
436 ture. The expansion of a mosaic of reef habitats in the Indo-Australian Archipelago during the
437 Oligocene/Miocene was a significant driver of cladogenesis for coral-reef-fish taxa (Bellwood
438 *et al.*, 2017), with smaller-sized lineages with low mobility and small home ranges radiating in

multiple provinces (Munday & Jones, 1998). Before the closure of the Isthmus of Panama, the Caribbean had a reef-building coral fauna that more closely resembled modern Indo-Pacific species. It further underwent a period of extensive faunal loss that has resulted in the modern day Atlantic and Eastern Pacific fish faunas (Budd, 2000; Bellwood & Wainwright, 2002; Di Martino *et al.*, 2018). The predominance of intermediate to large species across both sides of the Isthmus of Panama (Fig. 4b) may have resulted both from long-term and more recent loss of habitat (Alvarez-Filip *et al.*, 2009). This provides circumstantial evidence that assemblage structure is altered by habitat loss in a predictable manner (Wilson *et al.*, 2006). Although speculative at present, this hypothesis can readily be tested in future studies. In fact, recent evidence indicates that small species might suffer the most pronounced lags to tracking global changes (Graham *et al.*, 2007).

Here we demonstrate that in reef fishes, fish body size, reef area, and temperature can empirically predict species richness across spatial scales remarkably well. Therefore, our study lends strong support to theories that predict geographical gradients in species richness based on the combination of demographic processes which depend on habitat availability, size-dependent traits such as home range, dispersal and geographic range, and physiological processes such as the temperature-dependence of metabolism and speciation rates (Allen & White, 2003; Allen *et al.*, 2006; Reuman *et al.*, 2014; Tittensor & Worm, 2016). We hypothesize that species richness at small spatial scales builds up into province-level species richness at rates that are inversely related with the geographic range size of its constituent species, everything else being equal; in other words, reefs with larger, better dispersing species should exhibit less beta diversity because these species are found in all localities. In the absence of good proxies of abundance and range size in other taxonomic groups, these patterns that might be quite general across systems may remain unnoticed. Finally, at short temporal scales, activities that impact both small- (e.g. habitat degradation) and large-sized (e.g. overfishing) species might affect gradients in species richness in predictable ways. At longer temporal scales, the above effects combined with novel temperature regimes due to climate change should set the fate of environmental gradients in species richness.

467 **Acknowledgements**

468 We thank the French Foundation for Research on Biodiversity (FRB) and the Center of
469 Synthesis and Analysis of Biodiversity (CESAB) for funding the GASPAR working group
470 from which this study has been generated. We also thank AP Allen, LC Strotz, J Belmaker,
471 JE Williamson and M Westoby for comments on earlier drafts. Thanks also to D Tittensor, FA
472 Zapata, JP Quimbayo and JP Krajewski who shared their data, and all field assistants. ELR
473 was partly supported by a Bolsa Jovem Talento (CNPq 402053/2012-5). Financial support
474 from Brazil: SISBIOTA-Mar (PI: SR Floeter CNPq 563276/2010-0 and FAPESC 6308/2011-
475 8), Pro-Africa (PI: CEL Ferreira CNPq 490531/2007-5), CAPES, Marinha do Brasil, Instituto
476 Laje Viva; France: Institut de Recherche pour le Développement (IRD), PRISTINE project
477 (Fondation Total); Mexico: LE Calderón-Aguilera (CICESE), AL Cupul-Magaña, CM
478 Galván-Villa (Universidad de Guadalajara), HR Reyes-Bonilla, A Ayala-Bocos (UABCs),
479 RA López-Perez (UMAR), E Nuñez (UNACAR), fieldwork supported by CONABIO (PI:
480 LE Calderón-Aguilera project HJ026), PROMEP (PI: FA Rodríguez-Zaragoza, projects
481 103.5/08/2919 and 103.5/10/927) and SEP-CONACYT programs (PI: JE Arias-González);
482 Honduras: A Medina, P Rico (WWF-CA), M Aronne (Fundación Cayo Cochinos), fieldwork
483 supported by WWF-CA, USAID, FFEM and Universidad de Guadalajara (PI: FA Rodríguez-
484 Zaragoza); Nicaragua: R Barnutti (INPESCA), fieldwork supported by WWF-CA, USAID,
485 and Universidad de Guadalajara (PI: FA Rodríguez-Zaragoza); USA: National Geographic
486 Society (PI: SR Floeter; AM Friedlander).

487 **Statement of Authorship**

488 DRB, SRF and ELR conceived the study. All authors helped to collect and compile data.
489 DRB and ELR analysed the data and wrote the first draft. All authors contributed to substan-
490 tial revisions.

491 **List of brief titles of items in the Supplementary Material**

- 492 1. Appendix_Tables.xlsx (Contains Tables S1 and S2)
493 2. Supplementary_information.pdf (Contains Figures S1–4)

494 **Data accessibility statement**

495 All data and R code (data manipulation, analyses, figures and tables) can be downloaded from
496 a GitHub repository (<https://github.com/dbarneche/speciesPacking>), which will be made pub-
497 licly available upon publication.

498 **Biosketches**

499 **Diego Barneche** studies theoretical and empirical approaches that aim to unify evolutionary
500 ecology across levels of biological organisation. He is passionate about data, code, and repro-
501 ducible science.

502 **References**

- 503 Ackerman, J.L., Bellwood, D.R. & Brown, J.H. (2004) The contribution of small individuals
504 to density–body size relationships: Examination of energetic equivalence in reef fishes. *Oe-*
505 *cologia*, **139**, 568–571.
- 506 Allen, A.P. & White, E.P. (2003) Effects of range size on species–area relationships. *Evolu-*
507 *tionary Ecology Research*, **5**, 493–499.
- 508 Allen, A.P., Gillooly, J.F., Savage, V.M. & Brown, J.H. (2006) Kinetic effects of temperature
509 on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sci-*
510 *ences*, **103**, 9130–9135.
- 511 Alvarez-Filip, L., Dulvy, N.K., Gill, J.A., Côté, I.M. & Watkinson, A.R. (2009) Flattening of
512 Caribbean coral reefs: Region-wide declines in architectural complexity. *Proceedings of the*
513 *Royal Society B: Biological Sciences*, **276**, 3019–3025.
- 514 Barneche, D.R., Kulbicki, M., Floeter, S.R., Friedlander, A.M. & Allen, A.P. (2016) Energetic
515 and ecological constraints on population density of reef fishes. *Proceedings of the Royal Soci-*
516 *ety of London B: Biological Sciences*, **283**.
- 517 Bellwood, D.R. & Wainwright, P.C. (2002) *The history and biogeography of fishes on coral*
518 *reefs. Coral reef fishes: Dynamics and diversity in a complex ecosystem* (ed. by P.F. Sale), pp.
519 5–32. Academic Press, San Diego.
- 520 Bellwood, D.R., Goatley, C.H.R. & Bellwood, O. (2017) The evolution of fishes and corals on
521 reefs: Form, function and interdependence. *Biological Reviews*, **92**, 878–901.
- 522 Belmaker, J. (2009) Species richness of resident and transient coral-dwelling fish responds
523 differentially to regional diversity. *Global Ecology and Biogeography*, **18**, 426–436.
- 524 Brown, J.H. & Nicoletto, P.F. (1991) Spatial scaling of species composition: Body masses of
525 North American land mammals. *The American Naturalist*, **138**, 1478–1512.
- 526 Budd, A.F. (2000) Diversity and extinction in the Cenozoic history of Caribbean reefs. *Coral*
527 *Reefs*, **19**, 25–35.
- 528 Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: Standardizing sam-

- 529 ples by completeness rather than size. *Ecology*, **93**, 2533–2547.
- 530 Cinner, J.E., Huchery, C., MacNeil, M.A., Graham, N.A.J., McClanahan, T.R., Maina, J.,
531 Maire, E., Kittinger, J.N., Hicks, C.C., Mora, C., Allison, E.H., D'Agata, S., Hoey, A., Fearnly,
532 D.A., Crowder, L., Williams, I.D., Kulbicki, M., Vigliola, L., Wantiez, L., Edgar, G., Stuart-
533 Smith, R.D., Sandin, S.A., Green, A.L., Hardt, M.J., Beger, M., Friedlander, A., Campbell,
534 S.J., Holmes, K.E., Wilson, S.K., Brokovich, E., Brooks, A.J., Cruz-Motta, J.J., Booth, D.J.,
535 Chabanet, P., Gough, C., Tupper, M., Ferse, S.C.A., Sumaila, U.R. & Mouillot, D. (2016)
536 Bright spots among the world's coral reefs. *Nature*, **535**, 416–419.
- 537 Cornell, H.V. & Harrison, S.P. (2014) What are species pools and when are they important?
538 *Annual Review of Ecology, Evolution, and Systematics*, **45**, 45–67.
- 539 Di Martino, E., Jackson, J.B.C., Taylor, P.D. & Johnson, K.G. (2018) Differences in extinction
540 rates drove modern biogeographic patterns of tropical marine biodiversity. *Science Advances*,
541 **4**.
- 542 Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. & Magurran,
543 A.E. (2014) Assemblage time series reveal biodiversity change but not systematic loss. *Sci-
544 ence*, **344**, 296–299.
- 545 Edgar, G.J., Alexander, T.J., Lefcheck, J.S., Bates, A.E., Kininmonth, S.J., Thomson, R.J.,
546 Duffy, J.E., Costello, M.J. & Stuart-Smith, R.D. (2017) Abundance and local-scale processes
547 contribute to multi-phyla gradients in global marine diversity. *Science Advances*, **3**.
- 548 Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J.C., Smith-Vaniz, W.F., Wirtz, P., Ed-
549 wards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcón, J.M., Bowen,
550 B.W. & Bernardi, G. (2008) Atlantic reef fish biogeography and evolution. *Journal of Bio-
551 geography*, **35**, 22–47.
- 552 Gelman, A., Goodman, B., Gabry, J. & Ali, I. (2017) R-squared for bayesian regression mod-
553 els.
- 554 Graham, N.A.J., Wilson, S.K., Jennings, S., Polunin, N.V.C., Robinson, J., Bijoux, J.P. &
555 Daw, T.M. (2007) Lag effects in the impacts of mass coral bleaching on coral reef fish, fish-

- 556 eries, and ecosystems. *Conservation Biology*, **21**, 1291–1300.
- 557 Halpern, B.S. & Floeter, S.R. (2008) Functional diversity responses to changing species richness
558 in reef fish communities. *Marine Ecology Progress Series*, **364**, 147–156.
- 559 Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: An r package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, **7**, 1451–1461.
- 560
- 561 Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*, Princeton
562 University Press, Princeton, New Jersey, USA.
- 563 Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J.,
564 Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B.,
565 Lenihan, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J. & Warner, R.R.
566 (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science*, **293**,
567 629–637.
- 568 Kulbicki, M., Parravicini, V. & Mouillot, D. (2015) Patterns and processes in reef fish body
569 size. *Ecology of fishes on coral reefs* (ed. by C. Mora), pp. 104–115. Cambridge University
570 Press, Cambridge.
- 571 Kulbicki, M., Parravicini, V., Bellwood, D.R., Arias-González, E., Chabanet, P., Floeter, S.R.,
572 Friedlander, A., McPherson, J., Myers, R.E., Vigliola, L. & Mouillot, D. (2013) Global bio-
573 geography of reef fishes: A hierarchical quantitative delineation of regions. *PLoS ONE*, **8**,
574 e81847.
- 575 Lessard, J.-P., Belmaker, J., Myers, J.A., Chase, J.M. & Rahbek, C. (2012) Inferring local eco-
576 logical processes amid species pool influences. *Trends in Ecology & Evolution*, **27**, 600–607.
- 577 Luiz, O.J., Allen, A.P., Robertson, D.R., Floeter, S.R., Kulbicki, M., Vigliola, L., Becheler, R.
578 & Madin, J.S. (2013) Adult and larval traits as determinants of geographic range size among
579 tropical reef fishes. *Proceedings of the National Academy of Sciences*, **110**, 16498–16502.
- 580 MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*, Princeton Uni-
581 versity Press, Princeton, NJ, USA.
- 582 Maire, E., Cinner, J., Velez, L., Huchery, C., Mora, C., Dagata, S., Vigliola, L., Wantiez, L.,
583 Kulbicki, M. & Mouillot, D. (2016) How accessible are coral reefs to people? A global assess-

- 584 ment based on travel time. *Ecology Letters*, **19**, 351–360.
- 585 Messmer, V., Jones, G.P., Munday, P.L., Holbrook, S.J., Schmitt, R.J. & Brooks, A.J. (2011)
- 586 Habitat biodiversity as a determinant of fish community structure on coral reefs. *Ecology*, **92**,
- 587 2285–2298.
- 588 Munday, P.L. & Jones, G.P. (1998) The ecological implications of small body size among
- 589 coral-reef fishes. *Oceanography and Marine Biology: An Annual Review*, **36**, 373–411.
- 590 Nash, K.L., Welsh, J.Q., Graham, N.A.J. & Bellwood, D.R. (2014) Home-range allometry in
- 591 coral reef fishes: Comparison to other vertebrates, methodological issues and management
- 592 implications. *Oecologia*, **177**, 73–83.
- 593 O'Connor, M.I., Bruno, J.F., Gaines, S.D., Halpern, B.S., Lester, S.E., Kinlan, B.P. & Weiss,
- 594 J.M. (2007) Temperature control of larval dispersal and the implications for marine ecology,
- 595 evolution, and conservation. *Proceedings of the National Academy of Sciences*, **104**, 1266–
- 596 1271.
- 597 Parravicini, V., Kulbicki, M., Bellwood, D.R., Friedlander, A.M., Arias-González, J.E., Cha-
- 598 banet, P., Floeter, S.R., Myers, R., Vigliola, L., D'Agata, S. & Mouillot, D. (2013) Global
- 599 patterns and predictors of tropical reef fish species richness. *Ecography*, **36**, 1254–1262.
- 600 Peters, R.H. (1983) *The ecological implications of body size*, Cambridge University Press,
- 601 Cambridge, UK.
- 602 Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- 603 Reuman, D.C., Gislason, H., Barnes, C., Mélin, F. & Jennings, S. (2014) The marine diversity
- 604 spectrum. *Journal of Animal Ecology*, **83**, 963–979.
- 605 Ricklefs, R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology*
- 606 *Letters*, **7**, 1–15.
- 607 Rosenzweig, M.L. (1995) *Species diversity in space and time*, Cambridge University Press,
- 608 Cambridge, UK.
- 609 Soininen, J., Lennon, J.J. & Hillebrand, H. (2007) A multivariate analysis of beta diversity
- 610 across organisms and environments. *Ecology*, **88**, 2830–2838.
- 611 Srivastava, D.S. (1999) Using local-regional richness plots to test for species saturation: Pit-

- 612 falls and potentials. *Journal of Animal Ecology*, **68**, 1–16.
- 613 Stier, A.C., Hein, A.M., Parravicini, V. & Kulbicki, M. (2014) Larval dispersal drives trophic
614 structure across Pacific coral reefs. *Nature Communications*, **5**, 5575.
- 615 Teh, L.S.L., Teh, L.C.L. & Sumaila, U.R. (2013) A global estimate of the number of coral reef
616 fishers. *PLoS ONE*, **8**, e65397.
- 617 Tittensor, D.P. & Worm, B. (2016) A neutral-metabolic theory of latitudinal biodiversity.
618 *Global Ecology and Biogeography*, **25**, 630–641.
- 619 Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D.,
620 De Frenne, P., Verheyen, K. & Wipf, S. (2013) Global meta-analysis reveals no net change in
621 local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*,
622 **110**, 19456–19459.
- 623 Wilson, S.K., Graham, N.A.J., Pratchett, M.S., Jones, G.P. & Polunin, N.V.C. (2006) Multi-
624 ple disturbances and the global degradation of coral reefs: Are reef fishes at risk or resilient?
625 *Global Change Biology*, **12**, 2220–2234.
- 626 Witman, J.D., Etter, R.J. & Smith, F. (2004) The relationship between regional and local
627 species diversity in marine benthic communities: A global perspective. *Proceedings of the
628 National Academy of Sciences of the United States of America*, **101**, 15664–15669.
- 629 Worm, B. & Tittensor, D.P. (2018) *A theory of global biodiversity*, Princeton University Press.

630 **Figure legends**

631 **Figure 1.** Reef-fish species richness across scales. For clarity, the world map in (a) was
632 divided into major geographical realms (dotted lines) that accommodate multiple sites (points)
633 within biogeographical provinces (different colours and symbols). In (b–e), we show the
634 species richness at the checklist (i.e. list of species occurrences; $n = 132$ sites, Table S2),
635 province, sub-province and site scales respectively, with points being ordered on the x axis
636 based on their longitudes in (a). In (c–e), species richness was estimated using coverage-based
637 rarefaction, and the mean coverage among points (\pm S.D.) is given at the top right corner.

638 **Figure 2.** Fitted data based on our Bayesian hierarchical linear model showing the effects
639 of modal body size (a–c), reef area (d–f), and temperature (g–i) on reef-fish species richness
640 (from Fig. 1c–e) at the site (left column, eqn 1), sub-province (middle column, eqn 1), and
641 province (right column, eqn 2) scales. Panels at the site (a, d, g) and sub-province (b, e, h)
642 scales depict the partial effects of each variable after correcting species richness for the ef-
643 fects of all remaining fixed-effect covariates. At the province scale, three separate models
644 have been fitted, each for a different predictor. At the site scale, species richness has been
645 corrected for the random effects at the sub-province nested in provinces. At the sub-province
646 scale, species richness has been corrected for the random effects at the province scale. Bottom
647 left values represent mean (\pm S.D.) coverage among data points, which are the same for plots
648 within the same column. Bayesian R^2 are displayed with associated 95% credible intervals.
649 Panels (d) and (g) are generated from the same model as panel (a), and therefore contain the
650 same Bayesian R^2 (similarly for panels (b), (e) and (h)). Panels (c), (f), and (i) are generated
651 from three separate models as explained in the text, and therefore have different Bayesian R^2
652 values. Thin dashed lines show 95% Bayesian credible intervals around the mean model pre-
653 diction. Colours and symbols as in Fig. 1.

654 **Figure 3.** Violin plots showing the difference in species richness among provinces for each of
655 the six body size classes considered in our study, which are respectively indicated at the top
656 right corner in italics. Each violin represents a distribution of sites within a given province
657 (bottom points as in Fig. 1). We filtered our data for each class, and estimated the species
658 richness using coverage-based rarefaction, and the mean coverage among points (\pm S.D.) is

659 given at the top right corner. Black dotted lines separate major geographical realms.

660 **Figure 4.** Richness and body size across different spatial scales. In (a), sample-based rar-
661 efaction curves (based on 500 permutations) of reef-fish species richness show how reef-fish
662 species richness accumulates as sampling effort increases in different provinces. Mean rich-
663 ness is shown (\pm 95% confidence intervals). Mean modal body size (b) and variance in size
664 (c) of the species pool across different sampling areas, and across sites within our checklist
665 (from Fig. 1b; Table S2). Body size means and variances at each province were estimated
666 from randomly permuted subsets of transects based on 1,000 permutations (1, 5, 10, or 15
667 transects), or across all transects (All), or across all sites for the checklist data (Checklist).
668 Colours and symbols as in Fig. 1.

669 **Figure 5.** The negative relationship between modal body size and species richness across
670 provinces (a). In (b), we propose conceptually that the pattern in (a) arises from the different
671 contribution of large (top) vs. small (bottom) species to the observed mean body size and local
672 (but not province-level) richness. Distributions (blue for large species, red for small species)
673 represent hypothetical species-specific local abundances across space. Local richness can be
674 sampled at any point along this continuum, whereas the province-level richness arises from
675 the sum of all species (i.e. distributions). For the purposes of illustration, we represent the low
676 and high richness provinces respectively with large and small species only, but notice that in
677 reality provinces are made of a multitude of species with varying abundance and geographic
678 ranges. Colours and symbols in (a) as in Fig. 1.