1   **Title:** A simple spatially explicit neutral model explains range size distribution of reef fishes

2   **Short running title:** neutral model of range size distributions

3   **Authors**: Adriana Alzate[1,2,3], Thijs Janzen[4,5], Dries Bonte[3], James Rosindell[6*] & Rampal S.

4   Etienne[1*]

5   **Author affiliations:**

6   [1] Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 11103,

7   9700 CC Groningen, The Netherlands

8   [2] Terrestrial Ecology Unit, Ghent University, B-9000 Ghent, Belgium

9   [3] ECOMARES, Calle 39 Norte # 3CN-89, Cali, Colombia

10   [4] Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön,

11   Germany

12   [5] Institute for Biology and Environmental Sciences, Carl von Ossietzky University, 26111

13   Oldenburg, Germany

14   [6] Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst

15   Road, Ascot SL5 7PY, UK

16   *Joint last authors

17

18

Abstract

Aim

The great variation in range sizes among species has fascinated ecologists for decades. In reef-associated fish species, which live in fragmented habitats and adopt a wide range of dispersal strategies, we may expect species with greater dispersal ability to spread over larger ranges. However, empirical evidence for such a positive relationship between dispersal and range size in reef fishes remains scarce. Here, we unveil the more nuanced role of dispersal on the range size distribution of reef associated fishes using empirical data and a novel spatially explicit model.

Location

Tropical Eastern Pacific

Major taxa studied

Reef-associated fishes

Methods

We estimated range size distributions for six different guilds of all reef-associated fishes with different dispersal abilities. We used a one-dimensional spatially explicit neutral model, which simulates the distribution of species along a linear coastline to explored the effect of dispersal, speciation and sampling on the distribution of range sizes. Our model adopts a more realistic gradual speciation process (protracted speciation) and incorporates important long distance dispersal events with a fat-tail dispersal kernel. We simulated our model using a highly efficient coalescence approach, which guarantees the metacommunity, is sampled at dynamic equilibrium. We fitted the model to the empirical data using an approximate Bayesian computation approach, with a sequential Monte Carlo algorithm.

Results

43   Stochastic birth, death, speciation and dispersal events alone can accurately explain empirical

44   range size distributions for six different guilds of tropical, reef-associated fishes. Variation in

45   range size distributions among guilds are explained purely by differences in dispersal ability

46   with the best dispersers covering larger ranges.

47   *Main conclusions*

48   A simple combination of neutral processes with guild-specific dispersal ability provides a

49   general explanation for both within- and across-guild range size variation. Our results support

50   the theoretically expected, but empirically much debated, hypothesis that dispersal promotes

51   range size.

52

53   **Keywords:** spatially explicit, neutral model, dispersal, range size, range size distribution, reef

54   fishes.

55

56

Introduction

What is driving the large natural variation in the range size of species (Gaston 2003)? Answers to this long-standing question in macroecology were initially provided by investigating the effects of speciation and extinction processes (Anderson 1985, Gaston & Chown 1999). However, as suggested by Gaston and He (2002), these processes are not sufficient to explain range size distributions in nature, as they only affect the creation, division and removal of ranges. Among the other factors that could influence range size, dispersal ability of individuals is the one with the most important: dispersal is needed for the colonization of new habitats, and for persistence in existing habitats that are suboptimal, where demographic rescue can act to avoid local extinction (MacArthur & Wilson 1967, Brown & Kodric-Brown 1977). Dispersal also promotes gene flow, bringing the genetic variability necessary for adaptation, which is important for successful colonization and ultimately range expansion (Holt & Gomulkiewicz 1996). One group of organisms for which dispersal seems especially important is reef fishes because they live in habitats that are highly fragmented; making the ability to disperse key for habitat colonization, establishment, and range expansion. Despite theoretical expectations predicting a positive relationship between dispersal and range size, empirical evidence for this in reef fishes remains scarce (Lester & Ruttenberg 2005, Ruttenberg & Lester 2015, Mora *et al.* 2012, Luiz *et al.* 2013).

There are many possible explanations for the apparent lack of a positive range size-dispersal relationship; these reflect the many processes that potentially drive range size (reviewed in Gaston 2003) including speciation, local extinction, and range size changes during a species' lifetime (Webb & Gaston 2000). Firstly, range size is likely to vary with species age (Webb & Gaston 2000), i.e. older species might have attained larger ranges than newly formed species.

81   Secondly, species range dynamics are affected by biological interactions, eco-evolutionary

82   dynamics and by their behavioral and functional traits (Stahl *et al.* 2014). Thirdly, sampling

83   intensity and detection probability vary across space and across species (Dennis *et al.* 1999,

84   Alzate *et al.* 2014), and such sampling biases could also drive variation in range size. Finally,

85   stochastic events, especially during early life, may bring additional noise to the final range

86   size, making it difficult to find general patterns.

87

88   The dispersal component of range size-dispersal relationships is also problematic: dispersal is

89   a complex trait, varying at several life stages, e.g. during departure, transfer and settlement

90   phases (Bonte 2012), in ways that are not easily quantifiable. This may influence the outcome

91   of studies examining the role of dispersal. For example, many studies of dispersal on reef

92   fishes have focused primarily on the larval stage (Lester & Ruttenberg 2005, Lester *et al.*

93   2007, Mora *et al.* 2012), despite evidence that dispersal also occurs in earlier life stages as

94   eggs and in late life stages as adult fishes (Leis 1978, Kaunda-Arara & Rose 2004,

95   Appeldoorn *et al.* 1994, Addis *et al.* 2013).

96

97   Given the complexity of the problem, a promising approach for understanding the drivers of

98   range sizes (in contrast to the many correlative studies) is to model the process, including one

99   or several possible factors affecting range sizes. Although some previous studies have

100   attempted to explain range sizes using colonization-extinction models (Hanski 1982) or

101   population models (Gaston & He 2002), they were not developed to explain variation in range

102   size across many species exploring several factors. Here, we apply a variation of the unified

103   neutral theory of biodiversity and biogeography (Hubbell 2001), originally used to explain

104   other macroecological patterns such as species abundance distributions, species area

105   relationships and beta-diversity. We extend the neutral model to include spatially explicit

106 dynamics and a more realistic speciation process (Rosindell *et al.* 2008, Rosindell *et al.*

107 2011), both of which we expect to be important for a study of interspecific variation in range

108 sizes. This mechanistic model provides a way to quantitatively assess how dispersal can

109 influence species range size distributions, while at the same time considering other interacting

110 factors, including both sampling and speciation, that are known to affect range size (Gaston

111 2003). We tested the ability of our model to explain variation in range sizes by comparing its

112 predictions against empirical range size distributions of a complete reef fish assemblage in a

113 well-defined region: The Tropical Eastern Pacific (TEP). We made predictions of range size

114 distributions for each of six distinct guilds with different dispersal characteristics in the early

115 (egg and larval) as well as the later adult life stages. Our model is neutral and so excludes any

116 within-guild niche-based processes and individual differences. Crucially, by applying

117 independent neutral models to each of the six guilds we were able to focus on studying the

118 effects of different dispersal abilities for each guild in isolation from other complicating

119 factors such as environmental preference. With our spatially explicit neutral model, we tested

120 firstly whether range size distributions within guilds of reef fishes can be explained by neutral

121 factors alone and secondly whether variation in range size distribution across guilds can be

122 explained by differences in dispersal ability.

123

124 **Methods**

125 *Reef-associated fish data*

126 From the online database "Shorefishes of the Tropical Eastern Pacific - SFTEP" (Robertson &

127 Allen 2016), we collated spatial coordinates of species occurrences (45.860 records) for all

128 bony fishes (575 species) associated to reef habitats reported in the TEP. We used only

129 records inside the TEP region: 24° N (outer coast of California gulf, including all the inner

130 coast) and 4° S (SFTEP, Robertson & Allen 2016).

131   Reef fish species were classified in six different dispersal guilds according to traits related to

132   dispersal: spawning mode and adult mobility. We classified spawning mode in two types:

133   pelagic and non-pelagic. The differences in this early life history might confer diverse

134   capacities for dispersal (Riginos *et al*. 2011, Leis *et al*. 2013). Pelagic spawners release their

135   eggs in the water column, which are passively transported by water currents until the larvae

136   hatch and are able to better control active swimming (Stobutzki 1997, Leis *et al.* 2013). This

137   increase in the pre-hatching dispersal period might have strong and broader effect on dispersal

138   in the pelagic environment (Leis *et al*. 2013). Contrary to pelagic spawners, for which both

139   the egg and larval phases are pelagic, non-pelagic spawners either attach their eggs to the

140   substrate, are livebearers, or keep their eggs in the mouth or pouch until they hatch. Their

141   larvae usually emerge at larger sizes and are more mature than the larvae of non-pelagic

142   spawners (Wootton 1992, Leis *et al.* 2013), resulting in an early control of active swimming,

143   therefore limiting dispersal (Munday & Jones 1998, Leis 2006, Leis *et al.* 2013). We

144   classified adult mobility following Floeter and colleagues (2004) as low, medium and high.

145   Low adult mobility denotes site-attached species with a restricted home range ($< 10m^2$).

146   Medium adult mobility denotes species that are weakly mobile, relatively sedentary, with

147   close association to the substrate and that can be distributed over the entire reef area ($<$

148   ~$1000m^2$). High adult mobility denotes species that are highly mobile with wide horizontal

149   displacement and that occur in the water column (Floeter *et al.* 2004). Mobility for each

150   species was assigned depending on the taxonomical level at which information was reported:

151   species, genus or family adult mobility. In some cases, mobility information was not

152   available, but could be assigned according to the biology of the species, e.g. pearlfishes

153   (Family Carapidae) that are known to live inside the anal pore of sea cucumbers were all

154   classified as having low adult mobility. Information on adult mobility was obtained from

155   several sources (data base in Suppl. Mat). Information on spawning mode was obtained from

156   the SFTEP online database (Robertson & Allen 2016). Pelagic larvae duration, although often

157   used when studying range size of reef fishes, is not known for the majority (69%) of species

158   in the TEP region, making it unsuitable for this study.

159

160   *Measuring range size*

161   The range size of each species was calculated using a novel metric, developed for maximizing

162   comparability between simulated and observed range sizes: coastline distance. In contrast

163   with other traditional metrics, e.g. maximum linear distance, latitudinal and longitudinal

164   extent (Gaston 1994), coastline distance does not underestimate or overestimate range size

165   due to the particular spatial configuration of the TEP (Fig. S1). We defined coastline distance

166   as the contour distance (measured using units of 100 km) between the most distant points

167   along the coast line where the species was reported. However, the east and west coast of the

168   Californian gulf are treated as a single coast because the distance between opposing coasts is

169   likely too small to substantially restrict dispersal at similar latitudes (Fig. S1). All distance

170   measurements were calculated in kilometers using the function geodist from the R package

171   gmt (Magnusson 2015) and transformed in relative values, where 100% is the coastline

172   distance between the latitudes 24N and 4S.

173

174   *Spatially explicit neutral model*

175   We used a one-dimensional spatially explicit neutral model to simulate the spatial distribution

176   of species along a linear coastline. This configuration best reflects the particular geographical

177   distribution of reefs (coral and rocky) in the TEP region: a long coastline with a narrow

178   continental platform. As in the original neutral model (Hubbell 2001), the habitat is saturated

179   (zero-sum dynamics) and the species identity of an individual has no bearing on its chances of

180   dispersal, mortality, reproduction, the initiation of speciation or the completion of speciation

181 (see below). At every time step one individual, chosen at random according to a uniform

182 distribution, dies and is replaced by the newborn offspring of an existing individual

183 determined by a Pareto dispersal kernel:

184
$$f(x) = \begin{cases} \frac{\alpha X_m{}^{\alpha}}{X^{\alpha+1}}, & x \geq X_m \\ 0, & x \leq X_m \end{cases}$$

185 where X_m is a scale parameter (mode) and α is a shape parameter, that changes the distribution

186 from an exponential-like distribution (large value of α) to a very fat-tailed distribution (lower

187 values of α), i.e. many short distance dispersal events are combined with an occasional very

188 long-distance dispersal event. Random samples from the distribution can be calculated using

189 the inverse random sampling formula for the range size T:

$$T = \frac{X_m}{U^{1/\alpha}}$$

190 where U is a random variate drawn from a uniform distribution between 0 and 1. To separate

191 the effects of the shape of the distribution and the mean dispersal distance (X_{mean}), we rescaled

192 the Pareto distribution such that $X_m = X_{mean}$, i.e. that X_m reflects the mean dispersal distance,

193 and α still reflects the shape (see Suppl. Mat. for full derivation). The Pareto distribution

194 considers the possibility of long distance dispersal, in line with empirical dispersal

195 distributions of reef fishes (Jones 2015).

196

197 In contrast to the classical neutral model, we assumed that speciation is a gradual process

198 rather than an instantaneous event (Rosindell *et al.* 2010). When a birth event takes place, an

199 incipient species can form with probability μ; the newborn is still observed in the model as

200 being conspecific to its parent, but if sufficient time passes and descendants of the newborn

201 individual survive, those descendants will be considered a new good species rather than an

202 incipient one. This protracted speciation model entails one extra parameter τ: 'protractedness',

203 the number of generations required for an incipient species to become a real species, where

204   one generation means half of the turnover of the community because generations overlap.

205   Both speciation probability and protractedness influence the generation of new species, the

206   true speciation rate is a function of both parameters ($\mu/1 + \tau$) as described by Rosindell *et al.*

207   (2010). We simulated the spatially explicit neutral model using a coalescence approach

208   (Rosindell *et al.* 2008), which improves simulation efficiency while guaranteeing the

209   metacommunity is sampled at dynamic equilibrium and thus eliminating the problem of

210   determining an appropriate 'burn-in time' for the simulations.

211

212   ***Model behavior***

213   We explored the effect of dispersal on the distribution of range sizes by running simulations

214   using various dispersal kernels, which differ in their X_{mean} and α parameter values. We used a

215   linear lattice composed of 100,000 'units' which could be thought of as individual organisms

216   or larger cohorts of individuals behaving in a similar manner (Harfoot *et al.* 2014). We found

217   that larger lattices produce similar results (Fig. S2), but are computationally intractable for

218   parameter fitting exercises that require many successive simulation' runs. As in the real world

219   not all individuals are sampled, the proportion of sampled individuals (sampling percentage)

220   could therefore affect the observed distribution of ranges. Sampling was performed by

221   randomly choosing individuals along the linear lattice, and only sampled individuals were

222   used to quantify range sizes. Although sample areas along the TEP are not random, sampling

223   in a realistic manner produces virtually similar results as with random sampling (Fig. S3). We

224   examined the effect of dispersal (X_{mean} and α), speciation, protractedness and sampling

225   percentage on the distribution of species' range sizes. As species age is also suggested to be

226   positively related with range size (Gaston 2003), we also explored the effect of interspecific

227   variation in speciation rates on the distribution of range sizes. When speciation rate is high,

228   species are in average younger, thus affecting the final range size distribution.

229

230   In our default scenario, we used the following parameter values: X_{mean} = 0.02, α = 3.0,

231   sampling percentage s = 100%, speciation probability μ = 0.0005, protractedness τ = 10. We

232   then performed 5 sets of alternative scenarios, in which either values of X_{mean}, α, sampling

233   percentage, speciation probability or protractedness were altered. We explored 5 different

234   X_{mean} and α values (X_{mean} = [2%, 5%, 10%, 20%, 40%], α = [1.5, 2.0, 2.5, 3.0, 3.5]), 5

235   different sampling percentages (s = [1%, 5%, 20%, 50%, 100% of all individuals]) and 4

236   different speciation probability and protractedness values (μ= [5 x 10^{-2}, 5 x 10^{-3}, 5 x 10^{-4}, 5

237   x 10^{-5}], τ = [0, 10, 100, 1000]).

238

239   At the end of our simulations we estimated the range size for each species as the linear

240   distance (which is equivalent to coastline distance in a one-dimensional model) between the

241   most distant points where the species is recorded. The range size was measured in relative

242   terms, relative to the total lattice size. We replicated the simulations 100 times and calculated

243   mean and 95% CI values. Range sizes were transformed to percentages (100 % total size of

244   the linear lattice).

245

246   ***Model fitting***

247   In order to estimate dispersal (X_{mean} and α), sampling, speciation and protractedness values

248   that produced range size distributions matching those of empirical data, we used an

249   approximate Bayesian computation approach, with a sequential Monte Carlo algorithm (ABC-

250   SMC) as described by Toni *et al.* (2009). To assess the similarity between the data and

251   simulation outcomes, we calculated the sum of squares between the inverse cumulative

252   distribution for the simulated and empirical data, based on the differences in both the range

253   size distributions and species richness levels. Progression of the acceptance threshold was

254   modeled as an exponentially decreasing function, where the threshold at iteration t of the

255   ABC-SMC algorithm was: $500 \exp\left(-\frac{1}{4}t\right)$. We assumed the following prior distributions for

256   each parameter (on a $\log_{10}$ scale, e.g. $U_{10}(0,1) = 10^{U(0,1)}$, where U is a uniform distribution),

257   X_{mean}: $U_{10}(-4, -0.25)$, α: $U_{10}(0,1)$, speciation initiation rate: $U_{10}(-5, 0)$, protractedness: $U_{10}(0,5)$

258   and sampling: $U_{10}(-4, 0)$. Per ABC-SMC iteration, we used 10,000 particles. The ABC-SMC

259   algorithm ran for 20 iterations, or until the acceptance rate dropped below 1 in 1,000,000

260   proposed parameter combinations. Perturbation of the parameters was performed on a $\log_{10}$

261   scale, to avoid parameters reaching a negative value. Parameters were perturbed by first

262   taking the $\log_{10}$, then adding a random number drawn from a normal distribution with mean

263   zero and standard deviation 0.05, after which we exponentiated the parameter again. After

264   exponentiation, the parameter values were checked whether they still lay within the prior

265   ranges; if not, the particle was rejected. For each dataset we performed 10 replicate fits.

266

267   To assess the accuracy of our inference method, we generated artificial datasets using known

268   parameters, and performed the same ABC-SMC inference procedure as used on the empirical

269   data. If our method is accurate, inferred parameter values should be identical to the known

270   parameters used to generate the artificial data. Artificial data was generated using values for

271   X_{mean} of 0.001, 0.01, 0.1 or 0.2, α of 2, 4, 6 or 8, s of 0.025 or 0.25, and two different

272   speciation regimes: one with high speciation (0.01) and high protractedness (2500), and one

273   with low speciation (0.001) and low protractedness (25). For each parameter combination we

274   generated 10 artificial datasets. In total we performed (10 x 4 x 4 x 2 x 2) = 640 ABC-SMC

275   inferences to assess accuracy.

276

277   The one-dimensionality of our neutral model means the coastline distance metric treats the

278   coast of the TEP as also being one-dimensional (distance is only measured along the coast,

279   not as a birds-flight distance); this maximises the comparability of empirically observed range

280   sizes with those simulated by our one-dimensional, spatially explicit neutral model. In

281   addition, we excluded observations from oceanic islands when quantifying range sizes, again

282   to maximize comparability with simulated ranges. Our model was written in C++ and all post

283   simulation analyses were performed with R, version 3.3.1 (R core team 2016).

284

285   **Results (883)**

286   ***Range size distribution of reef associated fishes in the TEP***

287   Irrespective of their adult mobility, all three guilds of pelagic spawners have a relatively high

288   proportion of species with large ranges (Fig. 1a). The range size distributions of pelagic

289   spawners are qualitatively similar, with more than 70% of the species having ranges larger

290   than 50% of the maximum possible range or our sampling region. In contrast, the range size

291   distribution of non-pelagic spawners depends strongly on the capacity of adult fishes to

292   disperse. Within the non-pelagic spawners, the lowest dispersive guild has the highest

293   proportion of species with small ranges and the lowest proportion of species with large ranges

294   (Fig. 1a). While more than half of the species with medium or high adult mobility have ranges

295   larger than 80% of the maximum range, for species with low mobility only a fifth of species

296   have ranges larger than 80% of the maximum.

297

298   ***Spatially explicit neutral model***

299   The strongest effects on the distribution of range sizes are caused by variation in mean

300   dispersal distance (X_{mean}), speciation rate, and protractedness (Fig. 2). Dispersal (X_{mean} and α)

301   has a strong effect on the shape of the range size distribution. The contributions of X_{mean} and α

302   to the effect of dispersal on the range size distribution are not equal however, with the

303   majority of the dispersal effect resulting from X_{mean} (Fig. 2a). As X_{mean} increases, the

304   proportion of species with large ranges increases as well. In contrast, the shape parameter of

305   the dispersal kernel (α) has limited influence over the distribution of range sizes (Fig. 2b).

306   Speciation exerts a strong effect on the distribution of ranges, with a higher proportion of

307   species having a large range size when speciation rate is low. A high speciation rate produces

308   more new species, which initially have small ranges, thus a decrease in the number of species

309   with large ranges, and a (potentially unrealistically) high number of species in total (Fig. 2d).

310   The effect of protractedness is similar to that of speciation, as it modifies the number of

311   species and the rate at which these are created. The higher the protractedness, the longer the

312   time before an incipient species becomes a good species, and as a result fewer species have

313   small ranges (Fig. 2e). Sampling affects the distribution of ranges in a different way to

314   dispersal, speciation or protractedness: a lower sampling effort leads to more species with few

315   individuals and thus a higher proportion of species with apparently small ranges (Fig. 2c).

316

317   Prior to fitting the model to empirical data, we used the ABC-SMC fitting procedure on

318   simulated range size distributions with a known set of parameters (known values for X_{mean}, α,

319   speciation, sampling and protractedness). We found that posterior distributions of parameter

320   values were generally closely matching the real values (Fig. S4), indicating that our fitting

321   procedure was appropriate for estimating the parameter values of our neutral model. Only in

322   the case of the α parameter (measuring the shape of the dispersal kernel), were estimates were

323   not accurate, likely due the low strength of α in explaining range size variation (see above).

324

325   The same fitting procedure on empirical range size distributions, for the six dispersal guilds of

326   reef fishes, showed adequate fit between observed and predicted range size distributions (Fig.

327   3). Furthermore, in line with expectations, the estimated mean dispersal distances for each

328   guild were largest for the guilds with the highest proportion of large ranges which were

329   pelagic spawners and guilds with high adult mobility as expected. α values were similar for all

330   dispersal guilds (between 3.4 and 4.7). Estimated sampling completeness was lowest for the

331   guilds of non-pelagic spawners with high and medium mobility (0.76 and 0.48%

332   respectively), similarly low for the guild of pelagic spawners (3 - 9%) and very high for the

333   guild of non-pelagic spawners with low adult mobility (38%). Protractedness (the time it takes

334   for an incipient species to become a true species) values were the lowest for non-pelagic

335   fishes, low mobility species (13 generations), while values were intermediate for pelagic

336   spawners (160-730 generations) and highest for non-pelagic spawners with high and

337   intermediate mobility (3500 and 7000 generations respectively). The speciation probability

338   parameter (giving probability for an individual to become a new incipient species) was similar

339   to protractedness being low for pelagic spawners (0.02-0.03), similarly high for non-pelagic

340   spawners with high and medium adult mobility (0.08, 0.06) and the lowest for non-pelagic

341   spawners with low adult mobility (0.0007). See Table S1 for a complete description of the

342   model estimates.

343

344   For two dispersal guilds (pelagic spawners with high and medium adult mobility), our neutral

345   model could not fully explain the bimodality in their range size distribution. This mismatch

346   was strongest for pelagic spawners with medium adult mobility (Fig. 3). To explore what

347   caused these mismatches, we performed further analyses, in which we plotted the distribution

348   of ranges for fishes that are endemic to the TEP and one for the non-endemics (following

349   Robertson & Allen 2016). The distribution of ranges in the TEP for these two groups showed

350   differences for all guilds, but especially for the guild of pelagic spawners with medium

351   mobility (Fig. 4). In this case, the bimodality does not appear in either endemics or non-

352   endemics when separated, the combination of these two different distributions thus explains

353   the observed bimodality in the overall distribution.

354

355   **Discussion**

356   For decades, macroecologists have tried to understand the large variation in range sizes across

357   species. Using a new approach comprised of several spatially explicit neutral models, we have

358   shown that range size variation can emerge from stochastic birth, death, speciation and

359   variable dispersal abilities. Due to the mixed results of other studies, the importance of

360   dispersal ability in explaining range size variation has often been questioned (Lester &

361   Ruttenberg 2005, Luiz *et al.* 2013, Ruttenberg & Lester 2015). Here, we show that dispersal is

362   really an important factor shaping the range size distribution of species, but that our detailed

363   analyses were required to see this. For example, a study of only species with pelagic eggs may

364   not have revealed any clear effect of dispersal. High dispersal produces distributions with a

365   large proportion of species with large ranges, whereas low dispersal produces a large

366   proportion of small ranged species, consistent with a positive relationship between dispersal

367   and range size. Our model, however, also shows that range size variation can be large within

368   dispersal guilds, as dispersal only affects the *probability* to have large or small ranges. Thus,

369   although low dispersal produces distributions with a large proportion of small ranges, there

370   are also some species with large ranges, and vice versa for high dispersal. This also helps

371   explain why it has been challenging for empirical studies to show clear links between

372   dispersal ability and range size: for each level of dispersal ability, a large variation in range

373   sizes is still possible. Our neutral model predicts range size distributions with a close fit to the

374   empirical distributions for six different dispersal guilds of reef fishes in the TEP, and for each

375   guild estimated mean dispersal distance was in line with expectations, indicating that despite

376   their simplicity, neutral models still capture the most important processes for driving range

377   size variation within such guilds. Importantly, the neutral models we used were originally

378   developed to understand other macroecological patterns (Hubbell 2001), and thus can be seen

379   as an independent mechanistic tool, rather than a phenomenological construct tailored to fit

380   one pattern only.

381

382   Although our models generally fitted empirical range size distributions adequately, there were

383   some exceptions. Within guilds of pelagic spawners with high and medium adult mobility,

384   range size distribution tended to be bimodal, something that could not be explained by neutral

385   processes alone. We found that this bimodality primarily resulted from the combination of

386   two different background distributions: TEP endemics vs. TEP non-endemics, with the

387   endemics generally having larger ranges within the TEP. We hypothesize that the former have

388   had a longer time to increase their ranges in the region whilst the latter are biased by including

389   the edges of many wider ranged species that mostly occupy areas outside the TEP. We also

390   found that the range size distribution of non-endemic, pelagic spawners with medium mobility

391   was bimodal (Fig. 4e). A possible explanation is that this is due to their origin, with some

392   species coming originally from temperate regions (North and South America), and others

393   from tropical areas outside the TEP. We conjecture that the majority of species with large

394   ranges are trans-Pacific species, already adapted to tropical conditions. In contrast, 22 out of

395   the 24 species with very small ranges come from temperate regions, and it is likely that their

396   adaptations to a temperate climate and asymmetrical dispersal made these species less able to

397   expand their ranges into areas with more tropical conditions (Holt 2003). In fact, species

398   coming from the temperate north do not go down to the south and vice versa, whereas

399   transpacific species are well distributed along the coast (Fig. S5).

400

401   Our results showed that in addition to dispersal, speciation and sampling intensity can also

402   play an important role in shaping the distribution of range sizes. When sampling effort was

403   low, only a single individual was detected for many species (hence they were treated as

404   singletons, even if more individuals were present but not observed), leading to a high

405   proportion of species with very small ranges. The proportion of species with small ranges also

406   increased when speciation rates were high, or when speciation was a fast, non-gradual process

407   (low protractedness). In these cases, new species emerged continuously with low abundance

408   and restricted range. This outcome is in line with hypotheses attempting to explain why range

409   sizes in the tropics are usually smaller than in temperate regions such as 'Rapoport's rule'

410   (Rapoport 1982), which proposes that higher speciation rates in the tropics have caused this

411   pattern (Stevens 1989). Future empirical studies may potentially provide better tests of the

412   validity of our model outcomes. For instance, our predictions of how observed range size

413   distributions change when communities are increasingly intensively sampled, leading to larger

414   ranges as second conspecific individuals are seen for many singleton species.

415

416   While we could explain range size distributions using neutral models within guilds, average

417   range size varied across guilds, and observed species characteristics: both differences in adult

418   mobility and spawning mode. The estimated dispersal abilities from our models suggest that

419   differences in average range size are strongly influenced by dispersal. Consistent with

420   previous studies on neutral models with guild structure (using predictions for abundance

421   instead of range size, Janzen *et al.* 2015, Aduse-Poku *et al.* 2017), our results show that while

422   community dynamics within guilds may be captured by a neutral model, across guilds niche-

423   based processes drive variation in range size. Neutral theory was originally proposed to

424   describe community assembly within guilds (Hubbell 2001), and our results are consistent

425   with this philosophy. We take the concept further however, and show that across guilds,

426   niche-based processes, in this case differing dispersal strategies, play a larger role in driving

427   ecological patterns.

428

429 We have shown here how variation in range size across species can be explained by a

430 combination of neutral processes and guild-specific differences in dispersal. Our findings thus

431 make substantial progress towards settling a long-standing debate about the underlying causes

432 of variation in range size, and the role of dispersal in this pattern.

433

434

435 **Acknowledgments**

References

1. Addis, D. T., Patterson III, W. F., Dance, M. A. & Ingram Jr, G. W. (2013). Implication of reef fish movement from unreported artificial reef sites in the northern Gulf of Mexico. *Fisheries Research*. 147, 349-358.

2. Aduse-Poku, K., Molleman, F., Oduro, W., Oppong, S.K., Lohman, D.J. & Etienne, R.S. (2017). Relative contribution of neutral and deterministic processes in shaping fruit-feeding butterfly assemblages in Afrotropical forests. *Ecol. Evol.* In press. https://doi.org/10.1002/ece3.3618

3. Alzate, A. Zapata, F.A. & Giraldo, A. (2014). A comparison of visual and collection-based methods for assessing for community structure of coral reef fishes in the Tropical Eastern Pacific. *Rev. Biol. Trop.* 62(1), 359-371.

4. Anderson, S. (1985). The theory of range-size (RS) distributions. *Am. Mus. Novitates*. 2833, 1-20.

5. Appeldoorn, R. S., Hensley, D. A., Shapiro, D. G., Kioroglou, S. & Sanderson, B. G. (1994). Egg dispersal in a Caribbean coral reef fish, Thalassoma bifasciatum. II. Dispersal off the reef platform. *Bull. Mar. Sci.* 54(1), 271-280.

6. Bonte, D., Van Dyck, H., Bullock, J. M., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E., Mustin, K., Saastamoinen, M., Schtickzelle, N., Stevens, V. M., Vandewoestijne, S., Baguette, M., Barton, K., Benton, T. G., Chaput-Bardy, A., Clobert, J., Dytham, C., Hovestadt, T., Meier, C. M., Palmer, S. C. F., Turlure, C. & Travis, J. M. J. (2012). Costs of dispersal. *Biol. Rev*. 87, 290–312.

7. Brown, J.H. & Kondric-Brown, A. (1977). Turnover rates in insular biogeography: effects of immigration on extinction. *Ecology*. 58, 445-449.

8. Dennis, R.L.H. & Hardy, P.B. (1999). Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. *Global. Ecol. Biogeogr*. 8(6), 443-454.

9. Gaston, K.J. (1996). Species-range-size distributions: patterns, mechanisms and implications. *TREE*. 11, 197-201.

10. Gaston, K.J. (2003). The structure and dynamics of geographic ranges. Oxford University Press. New York, USA, pp. 266.

11. Gaston, K.J. & Chown, S.L. (1999). Geographic range size and speciation. In *Evolution of biological diversity*. (ed. A.E. Magurran & R.M. May), pp. 236-259. Oxford University Press.

12. Gaston, K.J. & He, F. (2002). The distribution of species range size: a stochastic process. *Proc. R. Soc. B.* 269, 1079-1086.

13. Lester, S. E. & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: a synthetic analysis. *Proc. R. Soc. B.* 72, 585-591.

14. Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., Kulbicki, M., Vigliola, L., Becheler, R. & Madin, J. S. (2013). Adult and larval traits as determinants of geographic range size among tropical reef fishes. *PNAS.* 110(41), 16498-16502.

15. MacArthur, R.H. & Wilson, E.O. (1967). *The theory of island biogeography.* Princeton University Press, pp. 203.

16. Mora, C., Treml, E., Robert. J., Crosby, K., Roy, D. & Titterson, D. P. (2012). High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. *Ecography.* 35(1), 89–96.

17. Holt, R.D. & Gomulkiewicz, R. (1996). The evolution of species' niches: a population dynamic perspective. In *Case studies in mathematical modelling* (ed. H.G. Othmer., F.R. Adler., M.A. Lewis. & L.C. Dallon), pp. 25-50. Saddle River, NJ: Prentice-Hall.

18. Ruttenberg, B.I. & Lester, S.E. (2015). Patterns and processes in geographic range size in coral reef fishes. In *Ecology of fishes on coral reefs* (ed. Mora, C.), pp 97-103. Cambridge University Press, Cambridge, United Kingdom.

19. Webb, T.J. & Gaston, K.J. (2000). Geographic range size and evolutionary age in birds. *Proc. R. Soc. B.* 267 (1455), 1843-1850.

20. Stahl, U., Reu, B. & Wirth, C. (2014). Predicting species range limits from functional traits for the tree flora of North America. *PNAS.* 111, 13739-13744.

21. Dennis, R.L.H., Sparks, T.H. & Hardy, B.P. (1999). Bias in butterfly distribution maps: the effect of sampling effort. *J. Insect. Conserv.* 3, 33–42.

22. Lester, S. E., Ruttenberg, B. I., Gaines, S. D. & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecol. Lett.* 10, 745–758.

23. Leis, J.M. (1978). Systematics and zoogeography of the porcupinefishes (Diodon, Diodontidae, Tetraodontiformes), with comments on egg and larvae development. *Fish. Bull.* 76 (3), 535-567.

24. Kaunda-Arara, B. & Rose, G. A. (2004). Long-distance movements of coral reef fishes. *Coral Reefs.* 23, 410-412.

25. Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos.* 38, 210-221.

513   26. Hubbell, S. P. (2001) The Unified Neutral Theory of Biodiversity and Biogeography.
514   Princeton University Press, Princeton, NJ, pp. 392.

515   27. Rosindell, J., Wong, Y. & Etienne, R.S. (2008). A coalescence approach to spatial
516   neutral ecology. *Ecol. Inform.* 3(3), 259-271.

517   28. Rosindell, J. & Phillimore, A.B. (2011). A unified model of island biogeography
518   sheds light on the zone of radiation. *Ecol. Lett.* 14, 552-560.

519   29. Riginos, C., Douglas, C.E., Jin, Y., Shanahan, D.F. & Treml, E.A. (2011). Effects of
520   geography and life history traits on genetic differentiation in benthic marine fishes.
521   *Ecography.* 34, 566-575.

522   30. Robertson, D. R. & Allen, G. (2016). Fishes: East Pacific. An Identification Guide to
523   the Shore-Fish Fauna of the Tropical Eastern Pacific. (Copyright Smithsonian
524   Institution, Left Coast R&C, Santa Cruz, California).
525   http://biogeodb.stri.si.edu/bioinformatics/sftep/taxon.php, accessed in 2016.

526   31. Leis, J.M., Caselle, J.E., Bradbury, I.R., Kristiansen, T., Llopiz, J.K., Miller, M.J.,
527   O'Connor, M.I., Paris, C.B., Shanks, A.L., Sogard, S.M., Swearer, S.E., Treml, E.A.,
528   Vetter, R.D. & Warner, R.R. (2013). Does fish larvae dispersal differ between high
529   and low latitudes? *Proc. R. Soc. B.* 280: 20130327.

530   32. Wootton, R. J. (1992). Fish Ecology. Chapman & Hall, New York, pp. 212.

531   33. Munday, P. L. & Jones, G. P. (1998). The ecological implications of small body size
532   among coral-reef fishes. *Oceanography and Marine Biology: An Annual Review.* 36,
533   373-411.

534   34. Leis, J.M. (2006). Are larvae of demersal fishes plankton or nekton. *Advances in*
535   *Marine Biology.* 51, 57-141.

536   35. Floeter, S. R., Ferreira, C. E. L., Dominici-Arosemena, A. & Zalmon, I. R. (2004).
537   Latitudinal gradients in Atlantic reef fish communities: trophic structure and spatial
538   use patterns. *J. Fish. Biol.* 64, 1680-1699.

539   36. Magnusson, A. (2015) gmt: Interface between GMT map-making software and R. R
540   package version 1.2-0. http://CRAN.R-project.org/package=gmt. Accessed April
541   2015.

542   37. Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M.P.H. (2009). Approximate
543   Bayesian computation scheme for parameter inference and model selection in
544   dynamical systems. *J. Royal Soc. Interface.* 6, 187–202.

545   38. Holt, R.D. (2003). On the evolutionary ecology of species' ranges. *Evol. Ecol. Res.* 5,
546   159-178.

547   39. Janzen, T., Haegeman, B. & Etienne, R.S. (2015). A sampling formula for ecological
548   communities with multiple dispersal syndromes. *JEB*. 374, 94-106.

549   40. Rapoport, E.H. (1982). Aerography: Geographical Strategies of Species. Pergamon
550   Press, New York, pp. 269.

551   41. Rosindell, J., Cornell, S. J., Hubbell, S.P. & Etienne, R.S. (2010). Protracted
552   speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13(6), 716-727.

553   42. Stevens, G.C. (1989). The latitudinal gradient in geographical range: how so many
554   species coexist in the tropics. *Am. Nat.*133,240-56.

555   43. Stobutzki, I. C. & Bellwood, D. R. (1997). Sustained swimming abilities of the late
556   pelagic stages of coral reef fishes. *MEPS*. 149, 35-41.

557

558

559

560

561   Figures

562   Fig. 1. (a) Distribution of range sizes for different dispersal guilds of reef fishes in the

563   Tropical Eastern Pacific (TEP). A guild is defined as a group of species that share the same

564   spawning mode (pelagic and non-pelagic spawners) and adult mobility (low, medium and

565   high). Range size is shown in relative terms, where a range of 100% is the largest range

566   recorded for a species in the TEP. We used coastline distance as the range size metric, which

567   is the distance between the most distant points along the coastline. Individuals from oceanic

568   islands are excluded to be consistent to the one-dimensional nature of the model. Both sides

569   of the California Gulf coastline were shrunk into a single one. The distribution of ranges is

570   shown as cumulative distribution curves, which show the proportion of species (y axis) that

571   attain ranges larger than a given size (x axis). (b) Map showing the sampling intensity along

572   the coastline in the TEP: the number of occurrences recorded at each coastline point spaced

573   by 100 km.

574

575   Fig. 2. Effect of (a) the mean dispersal distance X_{mean}, (b) the shape parameter of the dispersal

576   kernel α, (c) the sampling proportion, (d) speciation μ, and (e) the time to speciation τ

577   (protractedness) on the distribution of range sizes. Lines show the average value of 100

578   replicates and the shadows represent the 95% CI. For all simulations, the lattice size was

579   100,000 individuals. We use one fixed parameter setting, for which only the variable of

580   interest varied: $s = 100\%$, $\alpha = 3.0$, $X_{\mathrm{mean}} = 0.02$, $\mu = 0.0005$, $\tau = 10$.

581

582   Fig. 3. Range size distributions of the best model fitted to each dispersal guild, shown as an

583   inverse cumulative distribution curve. Mean of 5 replicates and 95% CI are shown. Dashed

584   lines represent the empirical data and coloured bands represent the distribution of values in

585   the best fitting model for that guild. Estimated X_{mean} (mean of >90.000 estimates) are shown

586   per each dispersal guild.

587

588   Fig. 4. Empirical range size distributions (as inverse cumulative distribution curves) for each

589   dispersal guild. The data shown separately for species that are endemics to the TEP and for
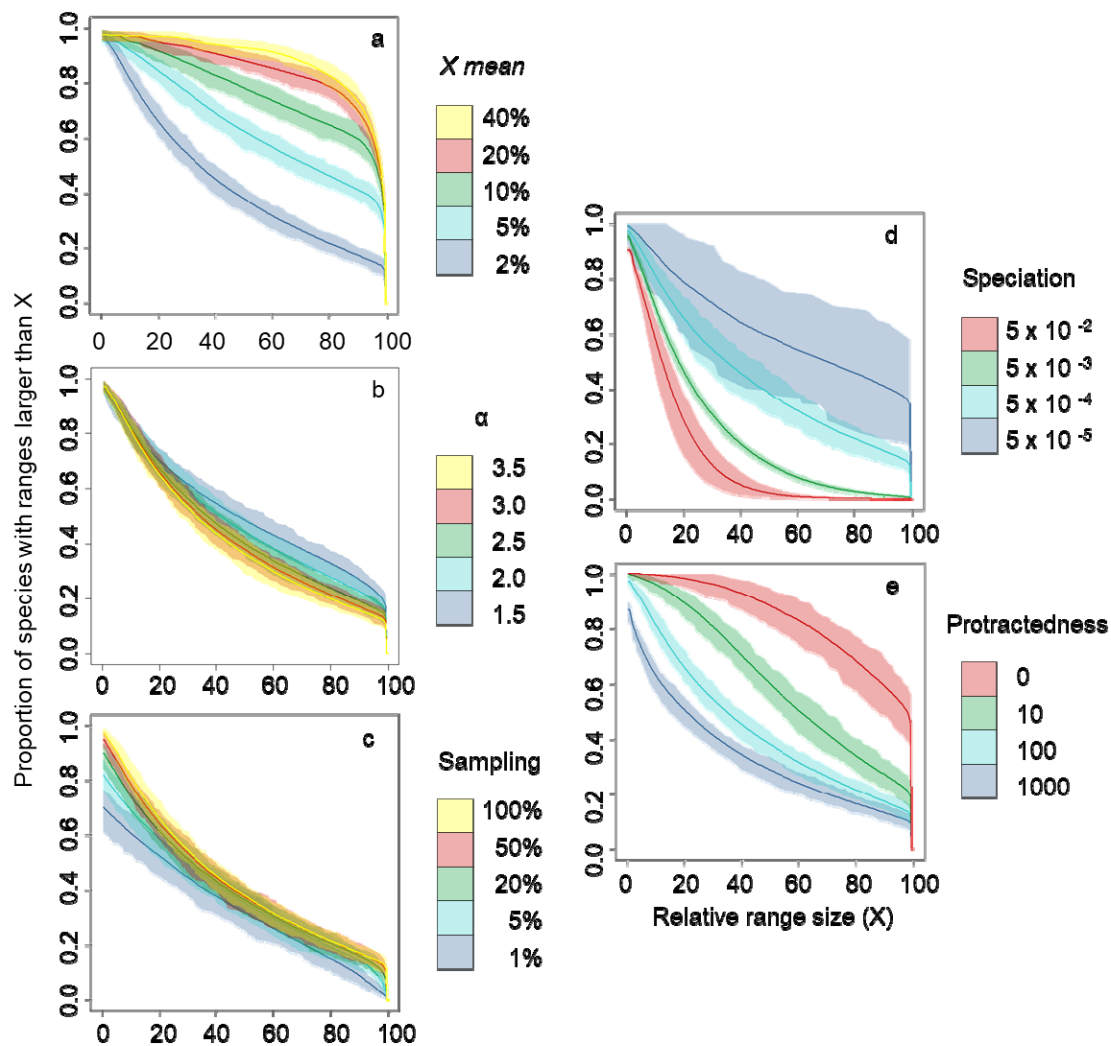
590   TEP non-endemic species.
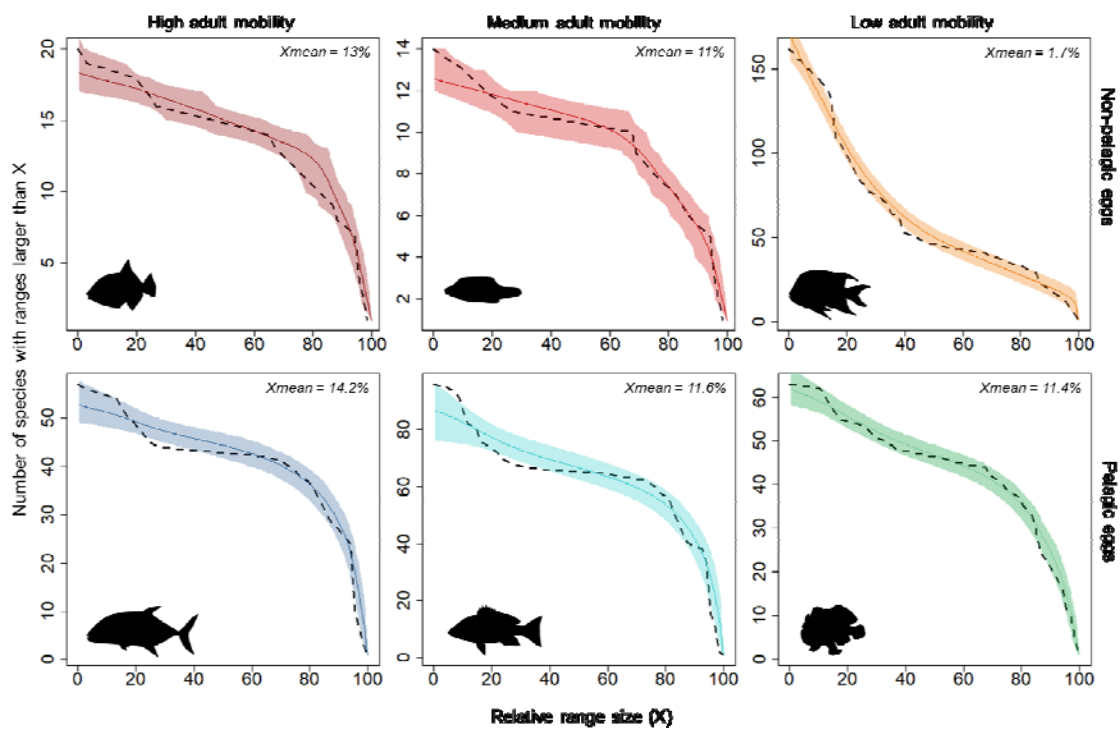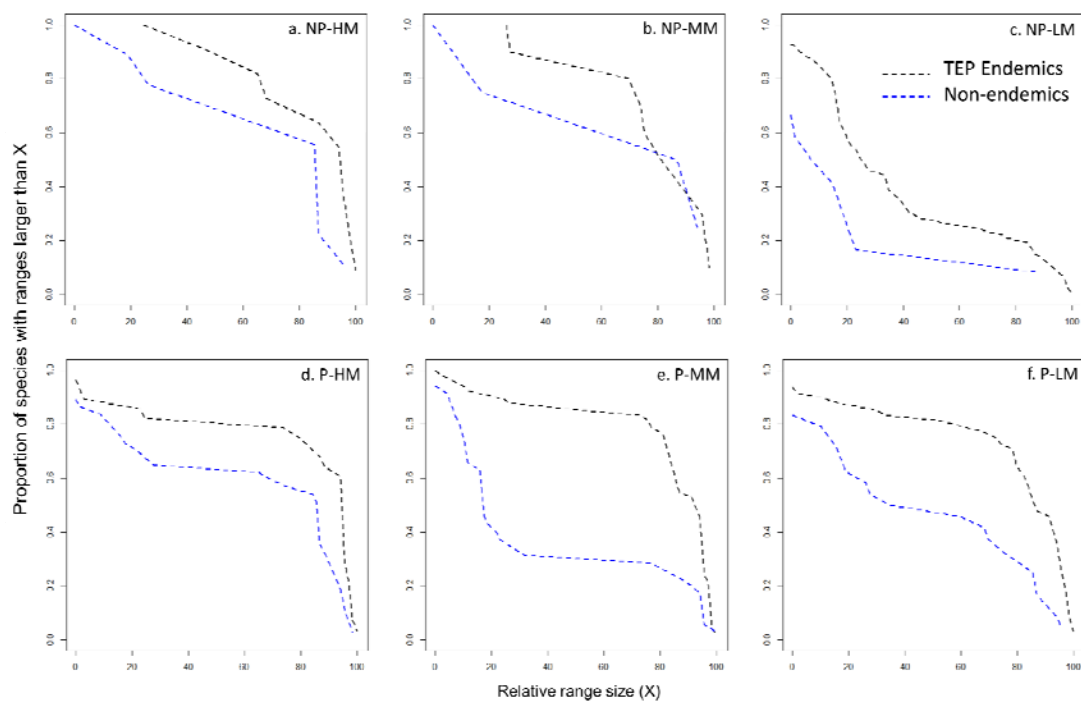
591

592   Fig. 1.
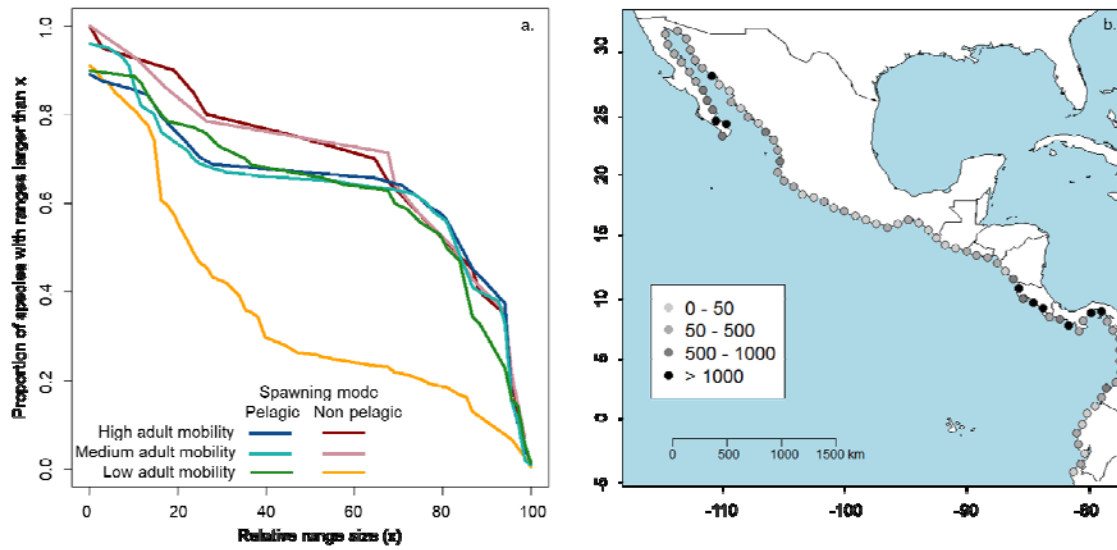
593

594   Fig. 2.

595

596

597 Fig. 3.

598

599   Fig. 4.

600