Stream diatom biodiversity in islands and continents-A global perspective on effects of area, isolation and environment

Aurélien Jamoneau, Janne Soininen, Juliette Tison-Rosebery, Sébastien Boutry, William Budnick, Siwen He, Julien Marquié, Jenny Jyrkänkallio-Mikkola, Virpi Pajunen, Anette Teittinen, et al.

Stream diatom biodiversity in islands and continents—A global perspective on effects of area, isolation and environment

1 **Stream diatom biodiversity in islands and continents – a global**

2 **perspective on effects of area, isolation and environment**

3

4 **Short running title: Diatom in islands and continents**

5

6 Aurélien Jamoneau[1†*], Janne Soininen[2†], Juliette Tison-Rosebery[1], Sébastien Boutry[1],

7 William R. Budnick[3], Siwen He[2,4], Julien Marquié[1], Jenny Jyrkänkallio-Mikkola[2], Virpi

8 Pajunen[2], Anette Teittinen[2], Vilja Tupola[2], Beixin Wang[4], Jianjun Wang[5,6], Saúl Blanco[7],

9 Alex Borrini[8], Marco Cantonati[9], Adelaide Clode Valente[10], Cristina Delgado[11], Gerald

10 Dörflinger[12], Vítor Gonçalves[13,14], Bryan Kennedy[15], Helena Marques[13], Athina

11 Papatheodoulou[16], Pedro Miguel Raposeiro[13], Catarina Ritter[13], António Serafim[17], Bart

12 Van de Vijver[18,19] and Sophia I. Passy[20].

13

14 [1] INRAE, EABX, 33612 Cestas, France.

15 [2] Department of Geosciences and Geography, University of Helsinki, Helsinki, FIN-

16 00014, Finland.

17 [3] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI

18 48824, USA.

19 [4] Laboratory of Aquatic Insects and Stream Ecology, Department of Entomology,

20 Nanjing Agricultural University, Nanjing, China.

21 [5] State Key Laboratory of Lake Science and Environment, Nanjing Institute of

22 Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China.

23 [6] University of Chinese Academy of Sciences, Beijing 100049, China.

24 [7] Department of Biodiversity and Environment Management, University of León, Spain.

25 [8] ERSE soc. coop. s.t.p. (Ecological Research and Services for the Environment),

26 Viareggio, Italy.

27 [9] BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma

28 Mater Studiorum, University of Bologna, Bologna, Italy.

29 [10] Secretaria Regional de Ambiente, Recursos Naturais e Alterações Climáticas - Governo

30 Regional da Madeira, Funchal, Portugal.

31 [11] Department of Ecology and Animal Biology, Faculty of Science, University of Vigo,

32 36200 Vigo, Pontevedra, Spain.

33 [12] Water Development Department, 1646 Nicosia, Cyprus

34 [13] Centro de Investigação em Biodiversidade e Recursos Genéticos, Rede de Investigação

35 em Biodiversidade e Biologia Evolutiva – Laboratório Associado, 9500-321 Ponta

36 Delgada, Portugal.

37 [14] Faculty of Sciences and Technology, University of the Azores, 9500-321 Ponta

38 Delgada, Portugal.

39 [15] Environmental Protection Agency, Castlebar, Co. Mayo, Ireland F23KT91.

40 [16] I.A.CO Environmental & Water Consultants Ltd, Nicosia 2035, Cyprus.

41   [17] AS.AMBI, Consultoria Ambiental, 7005-480 Évora, Portugal.

42   [18] Meise Botanic Garden, Meise, Belgium.

43   [19] Department of Biology, ECOSPHERE, University of Antwerp, Wilrijk, Belgium

44   [20] Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA

45   [†] equal contribution

46   * Corresponding author: Aurélien Jamoneau

47   **Email: aurelien.jamoneau@inrae.fr**

48

67 **Abstract**

68 **Aim.** The species-area relationship (SAR) is one of the most distinctive biogeographic

69 patterns, but global comparisons of the SARs between island and mainland are lacking

70 for microbial taxa. Here, we explore whether the form of the SAR and the drivers of

71 species richness, including area, environmental heterogeneity, climate and

72 physicochemistry, differ between islands and similarly sized areas on mainland, referred

73 to as continental area equivalents (CAEs).

74 **Location.** Global.

75 **Major taxa studied.** Stream benthic diatoms.

76 **Methods.** We generated CAEs on six continental datasets and examined the SARs of

77 CAEs and islands (ISAR). Then, we compared CAEs and islands in terms of total

78 richness and richness of different ecological guilds. We tested the factors contributing to

79 richness in islands and CAEs with regressions. We used structural equation models to

80 determine the effects of area vs. environmental heterogeneity, climate and local

81 conditions on species richness.

82 **Results.** We found a non-significant ISAR, but a significant positive SAR in CAEs.

83 Richness in islands was related to productivity. Richness in CAEs was mainly dependent

84 on area and climate, but not directly on environmental heterogeneity. Species richness

85 within guilds exhibited inconsistent relationships with island isolation and area.

86 **Main conclusions.** Ecological and evolutionary processes shaping diatom island

87 biogeography do not depend on area at the worldwide scale probably due to the presence

88 of distinct species pool across islands. Conversely, area was an important driver of

89 diatom richness in continents, and this effect could be attributed to dispersal. Continents

90 had greater richness than islands, but this was a consequence of differences in

91 environmental conditions such as specific island climatic conditions. We stress the need

92 for more island data on benthic diatoms, particularly from archipelagos, to better

93 understand the biogeography of this most speciose group of algae.

94

95 **Keywords**

96 ecological guilds, freshwater diatoms, island biogeography, macroecology, species-area

97 relationship, streams

Main Text

Introduction

100 A fundamental ecological law that describes how the number of species increases with

101 area is the species-area relationship (SAR, Arrhenius, 1921). The SAR belongs to a few,

102 truly robust generalizations in ecology detected in a wide range of ecosystems and taxa

103 (Connor & McCoy, 1979; Rosenzweig, 1995; Drakare et al., 2006). Islands represent

104 perhaps the most straightforward study setting to explore the SAR because of their well-

105 defined area. Unlike most mainland habitat patches, islands are surrounded by an

106 inhospitable matrix for continental taxa, which cannot be colonized and, consequently,

107 cannot serve as a source of immigrants. This peculiar feature of islands inspired

108 MacArthur & Wilson to develop the theory of island biogeography (MacArthur &

109 Wilson, 1967), which has contributed enormously to modern biodiversity theory (Chase

110 & Leibold, 2003), metapopulation biology (Hanski & Gaggiotti, 2004), community

111 ecology (Mittelbach & McGill, 2019), landscape ecology (Farina, 2008) and biodiversity

112 conservation (Prugh et al., 2008).

113 Island biogeography investigates how species richness on islands varies spatially and

114 through time (Whittaker & Fernandez-Palacios, 2007). It postulates that larger and less

115 isolated islands host more species than small and remote islands because larger area

116 decreases extinction and proximity to mainland increases immigration. Larger islands

117 may also encompass more species because they provide a larger target for immigration,

118 higher habitat diversity (Lack, 1976) and have higher speciation rates (Whittaker &

119 Fernandez-Palacios, 2007). Lastly, since island age affects diversification and erosion, it

120   may also determine species richness, which tends to be the highest in islands of

121   intermediate age according to the general dynamic model (Whittaker et al., 2008, 2017).

122   In the light of this knowledge, Chase et al. (2019) recently presented a framework for the

123   ecological mechanisms underlying the island SAR (ISAR). They suggested that passive

124   sampling (i.e. larger islands passively sample more individuals and species from the

125   regional pool than smaller islands), disproportionate effects (e.g. different colonization

126   and extinction rates in larger vs. smaller islands) and habitat heterogeneity (greater in

127   large islands) would be the main drivers of ISAR. Nevertheless, the major patterns and

128   drivers of island vs. mainland SAR are still poorly understood, particularly for the species

129   rich microorganisms.

130   Given the importance of environmental heterogeneity and dispersal on the SAR (Chase et

131   al., 2019), functional groups varying in resource utilization and dispersal can have

132   different SARs (Lomolino & Weiser, 2001; Báldi, 2008; Schrader et al., 2020). For

133   example, the SAR slope was steeper for specialist than for generalist bird species

134   (Matthews et al., 2014) and functional traits related to dispersal explained the SAR

135   variation in plant communities (Schrader et al., 2020). Thus, evaluating the SAR of

136   different ecological guilds may improve the knowledge of the niche- vs. dispersal-related

137   processes behind the SAR patterns. As functional diversity may have a distinct

138   (Jamoneau et al., 2018; Schrader et al., 2020) and even stronger response to

139   environmental variation than species diversity (Krause et al., 2014; Abonyi et al., 2018),

140   the SAR for different functional groups may elucidate how community assembly

141   processes operate through space and time (Tilman et al., 1997).

142   The ISAR has been tested with larger-bodied organisms, including terrestrial arthropods

143   (Simberloff & Wilson, 1969) and reptiles (Algar & Losos, 2011), birds (Kalmar &

144   Currie, 2006, 2007), vascular plants (Kreft et al., 2008), and fish (Sandin et al., 2008).

145   However, ISAR patterns are still poorly understood for microorganisms. Earlier

146   microbial field studies that used microcosms (Smith et al., 2005), lakes (Reche et al.,

147   2005), trees (Bell et al., 2005; Peay et al., 2007) or spring ecosystems (Teittinen &

148   Soininen, 2015) as surrogates of islands, reported significantly positive ISARs in almost

149   all systems (but see Teittinen & Soininen, 2015). However, investigations on

150   microorganismal diversity in real islands at a global scale are, to our knowledge, still

151   missing.

152   Rosenzweig (1995) hypothesized that islands should have lower local and regional

153   species richness than similarly sized continental regions due to isolation (lower mass- and

154   rescue effect), but steeper SAR slopes. This is because area tends to be a more critical

155   factor for biota on islands than on continents due to its stronger effects on extinction and

156   colonization (Kreft et al., 2008). However, in an extensive meta-analysis, Drakare et al.

157   (2006) did not find evidence for steeper SARs on islands (ISARs) than on mainland

158   across multiple species groups. The SAR patterns are typically explored within

159   archipelagos due to the presence of a common species pool, allowing assessment of the

160   pure area effect. However, there are also more general models for the SAR at the global

161   scale, searching for broader influences on the SAR (Kalmar & Currie, 2006; Kreft et al.,

162   2008; Triantis et al., 2015), including differences in evolutionary history (Rosenzweig,

163   1995).

164

165 Here, we adopted a similar perspective and investigated freshwater diatom SAR at a

166 worldwide scale, given that diatoms have large distributions (Finlay, 2002) and are

167 strongly controlled by environmental conditions (Soininen et al., 2016). We compared

168 SARs, total species richness, and species richness drivers between islands and

169 corresponding areas on five continents, referred to as continental area equivalents

170 (CAEs). For this comparison, we devised a novel method based on island-mainland pairs

171 (Fig. 1), assuming that terrestrial area is a good surrogate for area of freshwater habitat

172 (see Appendix S1 in Supporting Information). The CAEs corresponded to the sampling

173 area of 18 islands. We then examined (1) if SAR slopes differed between islands (ISAR)

174 and continents, (2) whether islands showed overall lower diatom species richness than

175 CAEs, (3) if species richness of island was related to environment, spatial isolation or

176 island age, and (4) whether habitat diversity, passive sampling or disproportionate effects

177 explained the SAR. We investigated these research questions separately for total diatom

178 species richness and species richness of ecological guilds, differing in dispersal capacity

179 and tolerance to nutrient limitation and disturbance (Passy, 2007, 2016), all expected to

180 influence the SAR (Matthews et al., 2014; Schrader et al., 2020).

181 **Materials and Methods**

182 *Biological and environmental datasets*

183 In total, we included 18 island datasets (Corsica, Cyprus, Guadeloupe, Iceland, Ireland,

184 Kauai, La Réunion, Martinique, Madeira, Majorca, Mayotte, New Caledonia, North New

185 Zealand, Oahu, Possession, São Miguel, Sardinia and South New Zealand) and six

186 continental datasets (China, Finland, France, French Guiana, Kenya and USA) in our

187 study (see Appendix S2). Diatoms were sampled from hard substrates (typically stones)

188 or macrophytes, generally during the low flow period (see Appendix S2 for details).

189 Although diatoms in some datasets were collected over several years, we did not expect a

190 substantial effect of interannual variation in our study, because we were interested in

191 regional diversity patterns and included environmental variables to account for this

192 potential variation.

193 Diatoms were cleaned with acid or hydrogen peroxide. A total of 400-700 diatom valves

194 were counted for each sampling site, which is sufficient for reliable estimates of total

195 diversity (Heino & Soininen, 2005). As the number of counted valves varied somewhat

196 among the samples, we studied if this would affect our richness estimates. We estimated

197 species richness with 300 valves and tested the correlation with the observed species

198 richness. We observed a very strong relationship between the estimated and the observed

199 species richness (R^2_{aj} = 0.98). Also, valve counts did not differ significantly between

200 islands and continents (Cliff test difference for large dataset, delta=-0.15). We thus

201 believe that the number of counted valves has only marginal impact on our richness

202 results.

203 Diatoms were generally identified up to species level, except in some rare case where

204 some of the valves were identified only to genus level (representing less than 5% of the

205 entire dataset). Homogenization of the taxonomy among regions was performed using the

206 OMNIDIA database (Lecointe et al., 1993, updated in November 2020). To ensure that

207 we have a proper estimate of the diversity, we i) evaluated the proportion of observed

208 species compared to the size of the species-pool in each region using basic Chao equation

209 (Chao, 1987) and calculated a 'corrected' species richness according to this ratio (i.e. the

210 observed species richness was increased relative to the proportion of missing species

211   estimated from the species pool) and ii) calculated a genus-based richness assuming that

212   genus level identification varies much less among diatomists than species identification.

213   We then ran analyses with observed species richness, corrected species richness and

214   genus richness (see Data analyses section). In total, our datasets comprised 1967 taxa,

215   further classified into four ecological guilds: low profile (species of short stature), high

216   profile (species of tall stature, typically filamentous, colonial or branched), motile

217   (species moving freely in the biofilm) and planktonic species (species not innate to the

218   benthos but originating from planktonic sedimentation) (Passy, 2007; Rimet & Bouchez,

219   2012; Soininen et al., 2016). Contrary to motile and high-profile species, low-profile

220   species are tolerant to nutrient limitation and disturbance and exhibit wider distributions

221   (Passy, 2016), suggesting potentially higher dispersal capabilities (Heino & Soininen,

222   2006). Planktonic species may indicate important features of the sites such as low current

223   velocity and large rivers.

224   Physico-chemical data of each sampling site included pH, conductivity (μS.cm^{-1}), total

225   phosphorus (mg.l^{-1}) and water temperature (°C), with the exception of Finland and

226   Possession island (with no water temperature data) and Ireland, Kenya and New Zealand

227   (with no total phosphorus data). Physico-chemical data were collected up to two months

228   before the diatom sampling. Climate data were obtained from WorldClim database at 0.5

229   minutes resolution (Hijmans et al., 2005), including annual precipitation (mm),

230   seasonality in precipitation (%), annual temperature (°C), and temperature seasonality

231   (standard deviation of monthly mean temperatures). For each sampling site, we also

232   extracted elevation from the Global Multi-resolution Terrain Elevation Data 2010

233   (Danielson & Gesch, 2011) and computed terrain slope as a proxy for current velocity.

234   For islands, we determined age of formation from the literature (see Appendix S2) and

235   isolation using the isolation index of Dahl (Dahl, 1991, Gillespie et al., 2008). This index

236   (equation 1) is based on the sum of square root distances to the nearest equivalent or

237   larger island (d_i), the nearest island group or archipelago (d_a), and the nearest continent

238   (d_c).

239   *Isolation index = $\sqrt{(d_i)} + \sqrt{(d_a)} + \sqrt{(d_c)}$ (1)*

240   *Creation of continental area equivalents (CAEs)*

241   For a reliable comparison of species-area relationships between islands and continents,

242   which are vastly different in size, we generated CAEs, comparable in size to the islands

243   by taking subsets of the continental data (see Algar & Losos, 2011) for a related

244   approach). The method used to create these CAEs (Fig. 1) was as follows.

245   We first computed the geographical centroid of each island and calculated $D_{c\text{-}i}$, a vector

246   representing the Euclidean distance between the centroid and each island sample site i.

247   Second, for each continent, we calculated D_{jj} the Euclidean distance matrix between

248   sample sites j. All Euclidean distances were calculated from geographical coordinates

249   expressed in a projected geographical system adapted for each region (see Appendix S2).

250   Third, we treated all continental sites as candidate CAE centroids and calculated $D_{v\text{-}j}$

251   representing the Euclidean distance between the candidate CAE centroid v and all other j

252   continental sites. We then computed a matrix $DD_{v\text{-}j,c\text{-}j}$ (equation 2), which represented the

253   absolute difference between i) the distance between a candidate CAE centroid and all

254   other sites in the focal continent ($D_{v\text{-}j}$, i.e. raw of the matrix D_{jj}) and ii) the distance

255   between the island centroid and all other sites in the focal island ($D_{c\text{-}i}$).

256     $DD_{v\text{-}j,c\text{-}i} = |D_{v\text{-}j} - D_{c\text{-}i}|$ (2)

257   Note that the minimum value of $DD_{v\text{-}j,c\text{-}i}$ is theoretically 0, indicating that the distance

258   between a centroid and an island site i is identical to the distance between a CAE centroid

259   and a continent site j. Thus, smaller $DD_{v\text{-}j,c\text{-}i}$ equates to similar distances between an island

260   centroid and island sites and the distances between a candidate CAE centroid and

261   continental sites. We then assigned for each centroid-island site distance a unique

262   corresponding CAE centroid-continent site distance ($\Delta_{c\text{-}i,v\text{-}j}$, i.e. the minimum value of the

263   column of $DD_{v\text{-}j,c\text{-}i}$, equation 3).

264     $\Delta_{c\text{-}i,v\text{-}j} = min(DD_{.,c\text{-}i})$ (3)

265   Then, we considered that the CAE centroid could be considered as the centroid of a CAE

266   only if at least $N=15$ of the selected CAE centroid-continent site distances $\Delta_{c\text{-}i,v\text{-}j}$ were

267   below a threshold value (θ) set to 5 km. Thus, theoretically, the number of sites in each

268   CAE could vary between 15 and the total number of sites in each island. Note that

269   because Kauai and Possession islands have less than 15 sites, N was set to 10 for the

270   creation of their CAEs. Finally, to avoid pseudoreplication within the sites of CAEs, we

271   selected for each continent-island pair only CAE separated by a distance of at least twice

272   the mean distance between centroids and their corresponding sites. Due to this procedure,

273   the size of CAEs could be, in some rare case, much smaller than the corresponding island

274   size.

275   The CAEs, corresponding to the sampling area of an island, were successfully created in

276   all continents (see appendix S3). Kenya was an exception because we were unable to

277   create CAEs corresponding to Corsica, Iceland, Ireland, New Caledonia, Sardinia, North

278 and South New Zealand, which were larger in size. Also, following our methodology, it

279 was not possible to create CAEs corresponding to the island of Mayotte in USA and

280 Finland, the island of São Miguel in Finland and USA and Possession Island in China and

281 USA because their continental sites were more spread out than the island sites.

282 Consequently, the total number of continent-island pairs for creating CAEs was 96.

283 *Randomization procedure for calculation of species richness and other environmental*

284 *variables*

285 For each continent-island pair, we randomly selected 15 sites within the CAEs and 15

286 sites within each island (10 sites for Kauai and Possession and their respective CAEs; 20

287 iterations) to achieve comparable sampling effort for islands and CAEs. For each random

288 subset, we calculated species richness as the total number of species observed among the

289 15 sites, and area from the convex hull around these 15 sites. We found that the areas in

290 islands estimated using convex hulls were good surrogates for whole island areas (see

291 appendix S4). We also calculated median values for each environmental variable for the

292 15 sites in the CAEs and islands and computed their environmental heterogeneity as the

293 multivariate dispersion of all environmental variables using the average distance of all

294 samples to the sample centroid in the multivariate space with the *betadisper* function in

295 the *vegan* package. Environmental variables used in the analyses and computation of

296 environmental heterogeneity were selected because they are known to be important for

297 stream diatom distributions (Soininen, 2007; Soininen et al., 2016). For the computation

298 of environmental heterogeneity in the Kenya and Ireland dataset, we respectively used

299 total nitrogen and orthophosphate concentrations instead due to the lack of total

300   phosphorus data (none of the nutrient concentration was used in the computation of

301   heterogeneity for New Zealand and Possession islands due to missing data).

302   *Data analyses*

303   We conducted separate analyses for CAEs and islands to examine the relationship

304   between species richness and area (SAR). We used linear mixed models (LMMs) for

305   CAEs to account for continental influences that may underlie differences in species pools

306   and the potential lack of independence among CAEs, given that multiple CAEs were

307   created within a continent (i.e. continents were included as random factors). We

308   performed traditional linear models for islands. We tested SAR with three commonly

309   used models (DeMalach et al., 2019), including power (Arrhenius model), logarithmic

310   (Gleason model) and Michaelis-Menten, and selected the best model based on the lowest

311   Akaike Information Criterion. We also tested relationships between area and the

312   'corrected' species richness (according to the size of the species pool) and genus richness

313   to ensure that the sampling effort or the taxonomic resolution did not influence our

314   results.

315   To test for passive sampling, we estimated species richness from rarefaction curves based

316   on species occurrence. For each CAE and island, we pulled at random 15 sites and

317   randomly selected 130 species occurrences without replacement, thus ensuring that the

318   maximum occurrence of each species did not exceed 15. Species richness was then

319   estimated from the 130 occurrences and used to generate the SAR, which was fit with

320   mixed models for CAEs and traditional linear models for islands. According to Chase et

321   al. (2019), failure to detect SAR using this estimation of species richness would suggest

322   that SAR is caused by passive sampling only. However, the reverse is not true, and

323 significant SAR observed with this estimation of species richness does not necessarily

324 prove the absence of passive sampling (Chase et al., 2019).

325 Then, to test for the effect of area on species richness after controlling for environmental

326 variation, we first computed global LMMs for total and guild species richness and eight

327 environmental variables (pH, conductivity, elevation, annual temperature, annual

328 precipitation, temperature seasonality, precipitation seasonality and environmental

329 heterogeneity). Models were constructed using the median values of species richness as

330 the response variable and median environmental variables obtained from the subsampling

331 procedure as explanatory variables (N = 851, i.e. one value for each 833 CAEs and each

332 18 island). Prior to analyses, explanatory variables were log-transformed to improve

333 normality when necessary and standardized, but we did not treat for multicollinearity

334 here, as this does not affect the fit of the model. Second, residuals from these regressions

335 were regressed against log-transformed area with LMMs for CAEs and simple linear

336 models for islands.

337 The number of islands in our study is comparable to the number of islands in many other

338 SAR studies (see data used in Matthews et al., 2019) but admittedly not very high (N=18)

339 for a study at the worldwide scale (Kalmar & Currie, 2006). Therefore, we performed a

340 sensitivity test with our continental datasets to determine the number of CAEs required

341 for observing a significant SAR, acknowledging that the number of islands and CAEs

342 necessary to detect a SAR may be different. We used the median values of the species

343 richness and area obtained from the randomization procedure for each continent-island

344 pair (N=96), and randomly sampled (1000 times) K continent-island pairs. Each time we

345 fit the SAR with the best SAR model (logarithmic) and extracted the probability (P) of

346   observing a significant SAR, as well as the median values of model coefficients. We

347   varied K from 11 to 96, i.e. the total number of continent-island pairs available in our

348   dataset. We then identified the minimum number of 'islands' needed to observe a

349   significant SAR with our data ($P>95\%$). We performed these analyses with both

350   traditional linear models and LMMs (e.g. assuming a common species pool).

351   We compared species richness of islands and species richness of their corresponding

352   CAEs with Cliff's non-parametric effect size statistic (Romano et al., 2006; Tecchio et

353   al., 2016), due to the large number of data points resulting from the randomization

354   procedure (i.e., decreasing variance around the mean). We also used Cliff's tests to

355   compare the species richness of each ecological guild between CAEs and islands.

356   To compare species richness of CAEs and islands after removing the effect of

357   environment, we computed LMMs as above but also included all the values of random

358   subsampling (x20, N=17020). We therefore used a nested design in the random factors of

359   the models, so that subsampling values are nested within each continent/island. Residual

360   richness values were then extracted from the models and compared between CAEs and

361   islands with Cliff's tests.

362   We used linear mixed models for CAEs and traditional linear models for islands to

363   examine the relationship between species richness, environmental heterogeneity, the

364   median of all environmental variables and the median values of latitude and longitude.

365   Environmental explanatory variables were log-transformed to improve normality when

366   necessary and we run separate regression models with each environmental factor and

367   species richness to avoid multicollinearity. We also tested for non-linear relationships

368   separately with all environmental variables with the same procedure.

369   Finally, to disentangle the possible drivers of the SAR for continents, we implemented

370   piecewise structural equation modeling (SEM, Lefcheck, 2016) using linear mixed

371   models with continental dataset as a random factor. We could not implement such models

372   for islands due to an insufficient number of data points. We assumed an *a priori* model

373   (Fig. 2) predicting species richness as directly influenced by area, environmental

374   heterogeneity (as defined above), local environmental conditions and climatic conditions.

375   We used conductivity and elevation as predictors of local conditions, temperature

376   seasonality and annual precipitations as predictors of climate, as they were significant

377   predictors of species richness in global LMMs and exhibited low collinearity in pairwise

378   correlations tests (see Appendix S5). We assumed that the effect of area on species

379   richness could also be indirect through environmental heterogeneity, according to the

380   habitat diversity hypothesis (Lack, 1976). Finally, we also assumed that temperature

381   seasonality and precipitation are directly influenced by elevation. We included a

382   correlation between temperature seasonality and precipitation as well as between

383   conductivity and precipitation (see Appendix S5). We used the Fisher's *C* statistic to test

384   the consistency of the theoretical model with the data. All analyses were run for total

385   richness and separately for richness of each ecological guild.

386   All analyses were conducted with R (R Core Team, 2019) using packages 'vegan'

387   (Oksanen et al., 2019), 'spatstat' (Baddeley et al., 2015), 'raster' (Hijmans, 2019), 'sf'

388   (Pebesma, 2018), 'lmerTest' (Kuznetsova et al., 2017), 'lme4' (Bates et al., 2015),

389   'effsize' (Torchiano, 2020) and 'piecewiseSEM' (Lefcheck, 2016).

390   **Results**

SAR patterns

We found a significant positive SAR for total species richness in CAEs, but not in islands (Fig. 3a). The best model describing the SAR in CAEs was the logarithmic model (see Appendix S6). The observed R^2 values were relatively low compared to values usually observed for islands but comparable to those found in continental areas (Kreft et al., 2008). Similar results emerged with rarefied richness (see Appendix S7), 'corrected' species richness given the size of the species pool (see Appendix S8), genus richness (see Appendix S9) and also after removing the effect of environmental variation (see Appendix S10).

The sensitivity analysis revealed that a minimum of 52 continent-islands pairs is needed to observe a significant SAR with our data. This number dropped to 16 when using mixed models with continent (a surrogate for the species pool) as a random effect (see Appendix S11).

About half of the 1967 identified species belonged to the motile guild (see Appendix S12), followed by the high profile and low-profile guilds. Planktonic species and species with variable guilds represented a minor part of the communities. Species richness within all guilds was significantly and positively related to area in CAEs (Fig. 3b-e) and this relationship persisted for all but the high-profile guild after controlling for the environment (see Appendix S10).

Comparison of species richness of islands and CAEs

Overall, species richness was significantly lower in the islands than in the respective CAEs for more than 50% of all continent-island pairs ($N = 96$) (Fig. 5a). Similar results

20

413   emerged for the guilds, especially for the planktonic guild, where over 70% of the

414   comparisons had significantly higher species richness in CAEs. The only exception was

415   the low-profile guild whose species richness tended to be higher in islands (ca. 60%).

416   Importantly, however, when environmental variation was accounted for, the species

417   richness differences between CAEs and islands disappeared in more than 80% of cases

418   (Fig. 5b).

419   *Ecological variables driving species richness*

420   In islands, we found significant relationships between species richness and isolation for

421   total, low profile species richness (U-shaped pattern) and planktonic guild (negative

422   linear pattern and a weak non-linear pattern) (Fig. 4, see Appendix S13). There was no

423   relationship between richness and age of island for the total community or any of the

424   ecological guilds (see Appendix S13). Apart from isolation, total species richness in

425   islands was significantly related only to phosphorus concentration. Species richness of

426   guilds was also significantly related to some other environmental variables depending on

427   the guild considered (see Appendix S13).

428   Total species richness in CAEs was significantly related to environmental heterogeneity,

429   pH, conductivity, phosphorus concentration, all climatic variables and longitude (see

430   Appendix S13). The piecewise SEM models (Lefcheck, 2016) disentangled the effects of

431   the influencing factors and demonstrated that diatom species richness in CAEs was

432   related to area, habitat heterogeneity, physicochemistry, elevation, and climate. The data

433   fitted well the *a priori* model (Fig. 6) for total species richness and species richness of all

434   ecological guilds. The marginal R^2 (variance explained by the fixed effects only) for total

435   species richness was 0.71 and varied between 0.14 (for low-profile species) to 0.59 (for

436 motile species). In the SEMs, area explained species richness independently, without any

437 indirect effect through environmental heterogeneity, except for the motile species

438 richness where the effect of environmental heterogeneity was negative. Indeed, total

439 species richness and richness of the motile guild were mainly driven by area and

440 temperature seasonality (Fig. 6a, d). Low-profile species richness was only explained by

441 area and precipitation (Fig. 6b). High-profile species richness was mainly explained by

442 area, elevation and conductivity, while climate had no direct effect (Fig. 6c). Finally,

443 planktonic species richness was solely determined by elevation and was thus the only

444 group without a significant relationship with area.

445 **Discussion**

446 Here, we conducted the first comparative analysis of island vs. mainland species-area

447 relationship for microbes, providing insight into the roles of area, environmental

448 heterogeneity, isolation and island age on species richness patterns. We showed for

449 freshwater diatoms that: (i) there was a significant SAR in continents but not in islands

450 (except for high profile), (ii) regional species richness was higher in continents than in

451 islands, but this difference was explained entirely by environmental conditions (iii) the

452 effect of isolation varied among diatom guilds and (iv) area and median environmental

453 conditions but not environmental heterogeneity were significant predictors of diatom

454 richness. Next, we will discuss the main findings in more detail and highlight our major

455 conclusions about total community and guild richness.

456 *Drivers of species richness in islands*

457 The lack of a significant SAR in islands may be due to low sample size (N = 18) or may

458 represent a real biogeographical pattern. Sensitivity analyses performed for CAEs

459   revealed that 16 islands are needed to detect a SAR given a common species pool. This

460   result is consistent with numerous studies on other organisms, reporting ISAR for a

461   relatively small number of islands within archipelagos (Matthews et al., 2019). However,

462   at a global scale, a much higher number of islands (N=52, Appendix S11) may be

463   required for detection of diatom ISAR.

464   The absence of diatom ISAR may have evolutionary and ecological causes. First, diatoms

465   may have distinct species pools across the globe (Soininen et al., 2016) and differences in

466   island area may not be sufficient to predict richness on islands that differ greatly in

467   species pool. As the size of the species pool influences the shape of the SAR (Catano et

468   al., 2021), future analyses on archipelagos will be essential for determining whether

469   ISAR exists for diatoms (but see Jüttner et al., 2018). Second, environmental

470   heterogeneity, which increased with island size (Fig. 3, and see Appendix S5) and is

471   recognized as an important driver of SAR (Lack, 1976; Chase et al., 2019), had no direct

472   impact on island species richness. Third, island richness was related only to total

473   phosphorus, suggesting that productivity is a key factor explaining island diatom species

474   richness at this scale. Note however, that due to data availability, only phosphorus

475   concentration was considered as a resource factor for explaining species richness. The

476   consideration of other nutrient resources, known to influence diatom diversity (e.g.

477   nitrogen, iron, Passy, 2007, Soininen, 2007), may improve the understanding of diatoms

478   species richness in islands.

479   We found that isolation might have some effect on species richness in islands. Two of the

480   most isolated islands (Oahu and New Caledonia) actually showed high species richness,

481   resulting in a U-shaped relationship between species richness and isolation for total and

482   low-profile species richness. Oahu and New Caledonia still had the highest species

483   richness when the latter is corrected by species pool but the U-shaped relationship is only

484   marginally significant (p = 0.09, see Appendix S8). Greater speciation in the most

485   isolated islands, which have many endemic species, e.g. New Caledonia has been dubbed

486   "Galapagos of diatoms" (Moser et al., 1998), may explain their higher richness

487   considering that endemic and total species richness are typically correlated (Kallimanis et

488   al., 2010). We could, however, not exclude the fact that some other unmeasured

489   environmental factors, particularly related to islands conditions, may also be responsible

490   for this pattern. Finally, our finding further suggests that the biogeographical drivers of

491   diatom richness on real islands are trait dependent.

492   Following Rosenzweig (1995), we hypothesized that islands would harbour lower species

493   richness than continents due to diminished dispersal and rescue effects. While species

494   richness was indeed lower in islands compared to continents, this difference disappeared

495   when we accounted for environmental differences. Thus, annual precipitation, higher in

496   islands than continents, was associated with lower species richness (see Appendix S10),

497   likely because of its positive effect on current velocity, and subsequently, shear stress

498   (Heino & Soininen, 2007).

499   *Drivers of species richness in continents*

500   We tested whether SARs in continents could result from passive sampling and

501   environmental heterogeneity, which are major drivers of the SAR (Lack, 1976;

502   Rosenzweig, 1995; Stein et al., 2014; Chase et al., 2019). Surprisingly, species richness

503   in continents was not directly explained by environmental heterogeneity in the SEM.

504   Although area was strongly related to environmental heterogeneity (but poorly related to

505   other environmental variables, see Appendix S5), none of the SEM models showed a

506   direct effect of habitat heterogeneity on either total or guild species richness, except for

507   motile species. For the latter, the direct effect of habitat heterogeneity was negative,

508   contrary to the results observed in univariate regressions (see Appendix 13) due to the

509   strong collinearity between area and heterogeneity. Given that we still observed a

510   significant SAR with species richness estimated from the rarefaction curves, passive

511   sampling cannot be completely ruled out (Chase et al., 2019). However, the impact of

512   area on richness in continents might also be due to disproportionate effects, including

513   dispersal, extinction and speciation. While extinction and speciation have been less

514   studied in diatoms, dispersal and mass effects (whereby species maintain their presence

515   in unfavorable conditions via immigration, Shmida & Wilson, 1985) were shown to have

516   a notable influence on regional to subcontinental diatom communities (Soininen, 2007;

517   Jamoneau et al., 2018; Leboucher et al., 2020). For continental diatoms, larger areas may

518   thus increase the probability of immigration from the surrounding landscape, particularly

519   for species with high dispersal capabilities (mass-effect species), thereby increasing

520   CAE's diversity.

521   Environmental factors, such as nutrients, climate and elevation, were also important

522   predictors of total and guild species richness. Total species richness decreased with

523   temperature seasonality, as did the species richness of motile species, which represented

524   ca. 50% of the whole community (see Appendix S12). As motile species are generally

525   warm-water species (Pound et al., 2021) and high seasonality occurs in colder areas, it is

526   possible that motile guild richness was limited by unfavorable temperatures. Species

527 richness of the high profile and planktonic guilds was the lowest at high elevation. For

528 high-profile species, high elevation is stressful due to increased current velocity and

529 probability for dislodgement. For planktonic species, high elevations do not provide

530 sufficient habitat, given that these species require large water bodies. Species richness of

531 the low-profile guild is positively influenced by annual precipitation probably because

532 this guild is tolerant to physical disturbance (Passy, 2007), which should increase its

533 richness in the community.

534 **Conclusions**

535 We examined diatom ISARs and compared them with the SARs of similarly sized

536 continental area equivalent across five continents. Contrary to most previous studies, we

537 did not find significant ISAR for total species richness but detected significant

538 relationships of richness with total phosphorus. These results imply that diatom richness

539 in islands is not related to area but is controlled by productivity. However, the lack of

540 ISAR may be due to distinct species pool across islands in our study. Species richness

541 was typically higher in continental areas than in similarly sized islands, most probably

542 due to differences in climate and related environmental conditions, such as current

543 velocity. The significant SAR for continents may originate from disproportionate effects,

544 such as mass effect, but not from environmental heterogeneity. Isolation influenced the

545 richness of the whole community and some diatom guilds in islands. These finding

546 indicate that there are important differences in richness responses to island properties

547 among ecological guilds and between the community level and the functional level.

548 Finally, the proposed new method for species-area comparisons between islands and

549 continental area equivalents will advance research on biogeography of islands vs.

550 mainland. We advocate obtaining global diatom data, particularly from archipelagos to

551 better understand the drivers of island species diversity.

Figures legend
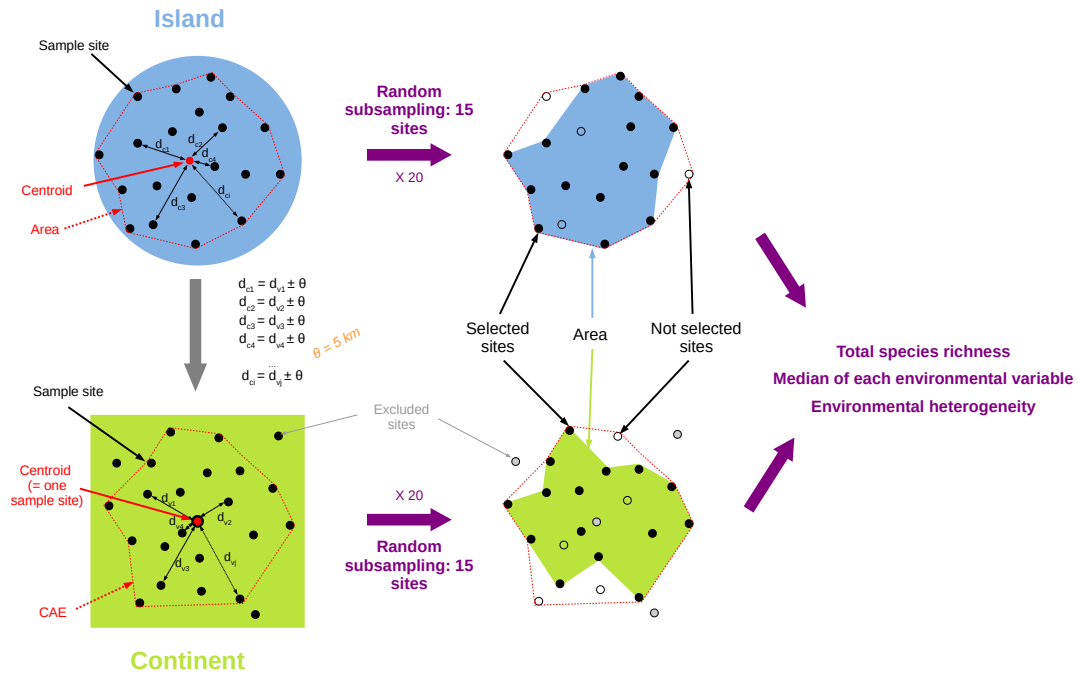
Figure 1: Descriptive diagram of the methods. Diagram describing the methodological

process used for creation of continental area equivalents (CAE) and subsampling of both

islands and CAEs.
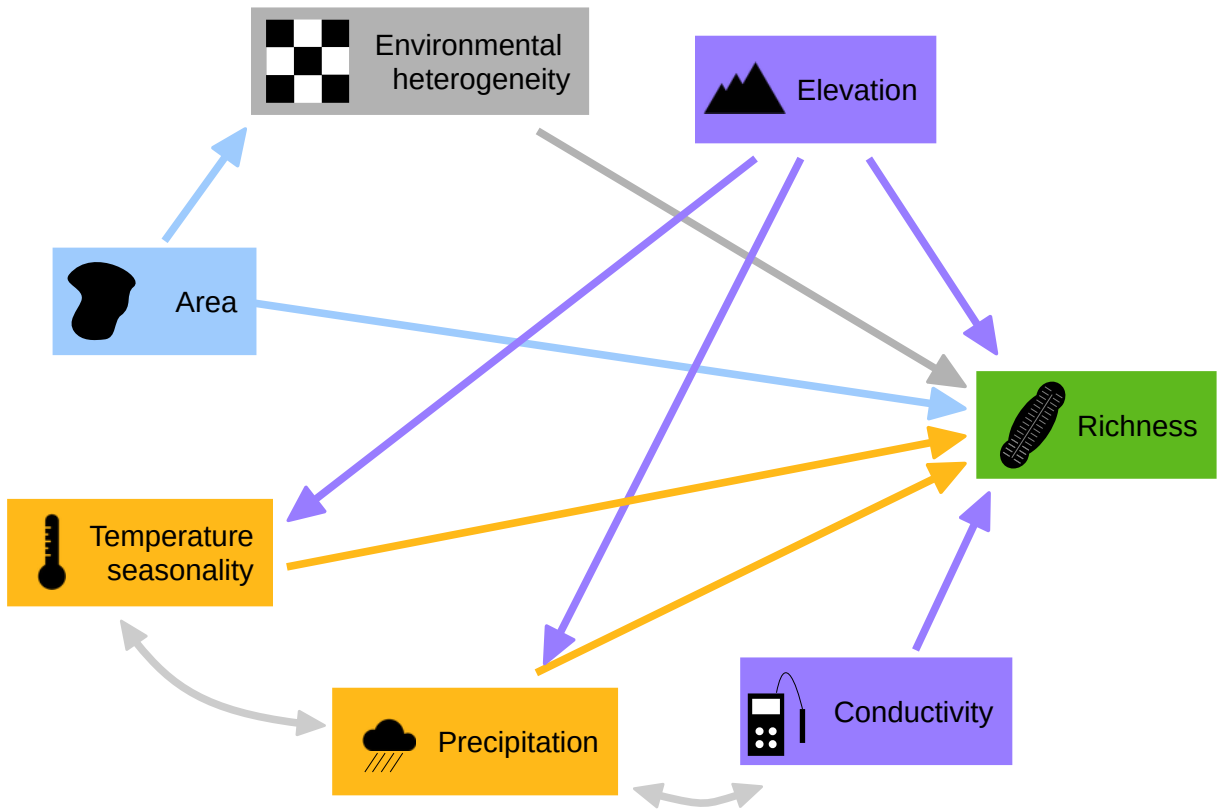
Figure 2. A priori model explaining diatom species richness. Species richness is modeled as a function of area, environmental heterogeneity, local environmental conditions (purple) and climate (orange).

(a) – All taxa

Environmental heterogeneity

27 55 120

(b) – Low profile

Total species richness 100 150 200

(c) – High profile

(d) – Motile

(e) – Planktonic

Area (km²)
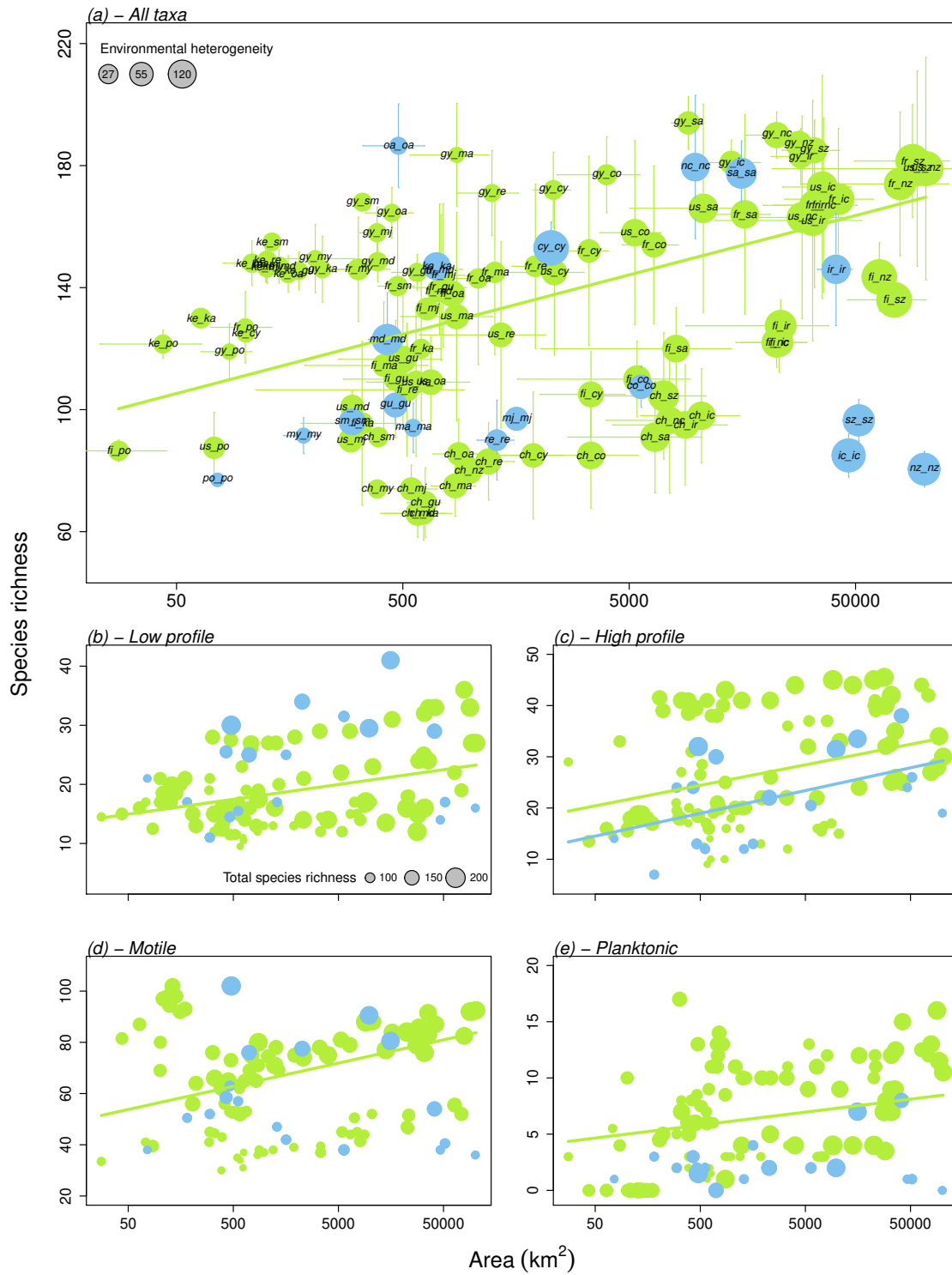
Species richness

562 **Figure 3. Species-area relationships for continents and islands.** Species-area

563 relationships for continental area equivalents (CAE) (in green, N=96) and islands (in

564 blue, N=18) for total species richness (a), and richness of low profile (b), high profile (c),

565 motile (d) and (e) planktonic species. Green regression lines represent significant linear

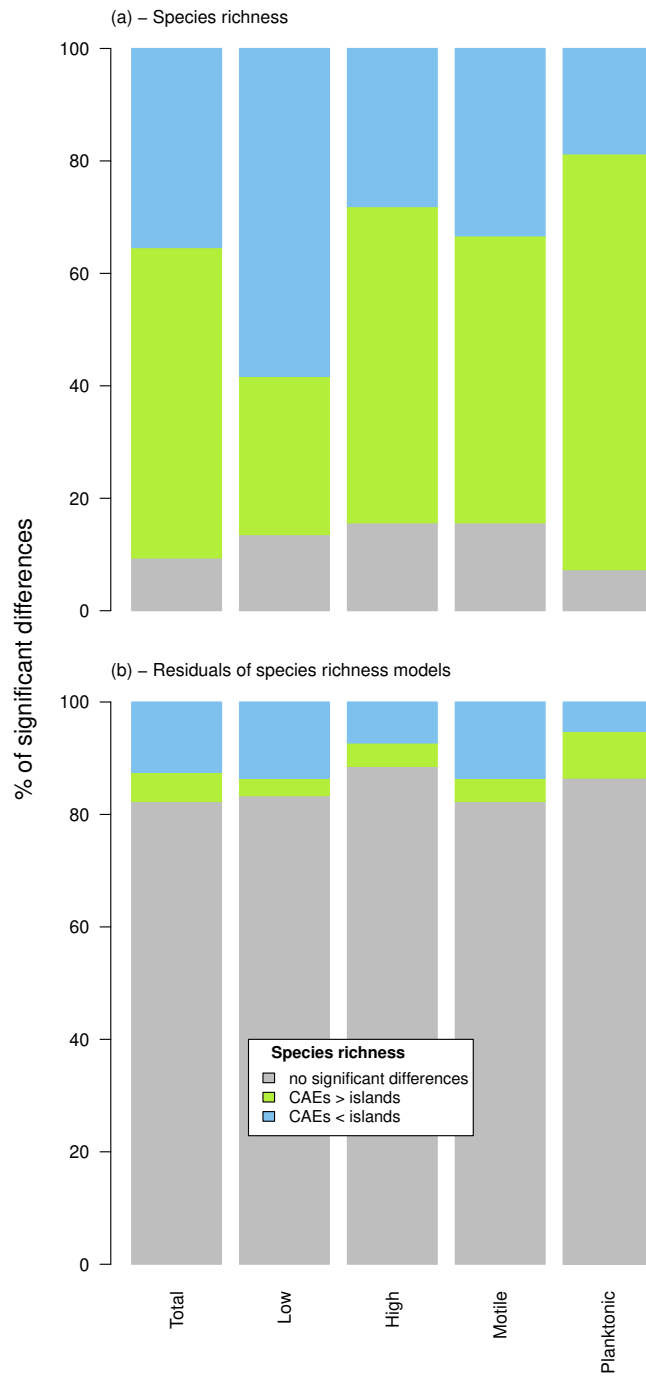566 fits in mixed models for CAEs: richness = 8.44x + 72.24, $R^2_m = 0.22$ for total species

567 richness, 1.09x + 10.69, $R^2_m = 0.13$ for low-profile species, 1.74x + 13.61, $R^2_m = 0.11$ for

568 high-profile species, 3.94x + 38.36, $R^2_m = 0.11$ for motile species and 0.50x + 2.69, $R^2_m =$

569 0.05 for planktonic species, where x = log(area). The blue regression line represents

570 significant linear fit for high-profile species of islands: 1.93x + 6.95, $R^2_{aj} = 0.23$. Dot

571 sizes are proportional to environmental heterogeneity (in log) for all taxa (a) and

572 proportional to total species richness for functional groups (b-e). Error bars represent

573 standard deviation estimated from the subsampling procedure. Text in dots indicate the

574 dataset used for computing species richness and area. For example, 'fr_my' indicates the

575 position of Mayotte CAE in France. Continental datasets are indicated by 'fr' for France,

576 'us' for US, 'fi' for Finland, 'ch' for China, 'ke' for Kenya and 'gy' for French Guiana

577 and islands indicated by 'ic' for Iceland, 'co' for Corsica, 'gu' for Guadeloupe, 'ma' for

578 Martinique, 're' for La Réunion, 'my' for Mayotte, 'nz' for North New Zealand, 'sz' for

579 South New Zealand, 'nc' for New Caledonia, 'ka' for Kauai, 'oa' for Oahu, 'po' for

580 Possession, 'cy' for Cyprus, 'ir' for Ireland, 'md' for Madeira, 'mj' for Majorca, 'sm' for

581 São Miguel and 'sa' for Sardinia.
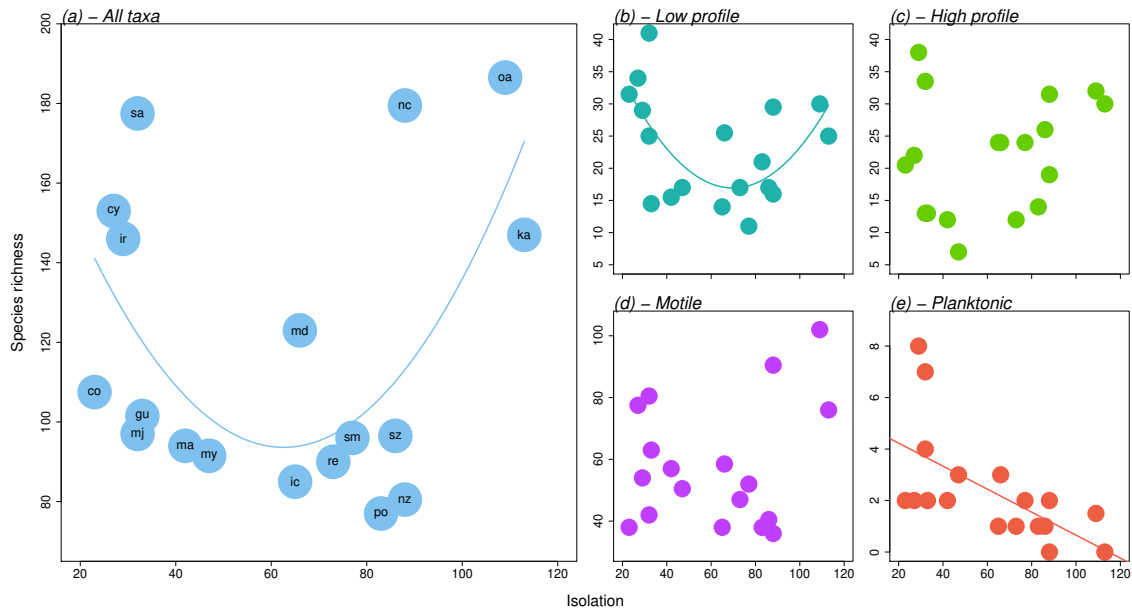
Figure 4. Relationships between island species richness and isolation. Relationship between total species richness (a) and species richness of each ecological guild (b-e) with island isolation for islands ($N = 18$). Significant linear and quadratic relationships (p<0.05) are shown by regression fits (only the fit with the lower AIC is shown if both are significant, see Appendix S13): $0.03x^2 - 3.79x + 212.32$, $R^2_{aj} = 0.26$ for total richness, $0.01x^2 - 0.96x + 50.44$, $R^2_{aj} = 0.32$ for low-profile and $-0.04x + 5.12$, $R^2_{aj} = 0.34$ for planktonic species. For island names, see Fig. 1. Isolation is based on index defined by Dahl (Dahl, 1991).

(a) – Species richness

(b) – Residuals of species richness models

Species richness
- no significant differences
- CAEs > islands
- CAEs < islands

% of significant differences

Total Low High Motile Planktonic

592 **Figure 5. Comparison of species richness between continental area equivalents**

593 **(CAEs) and islands.** Percentage of significant and non-significant tests ($N = 96$

594 continent-island pairs) between CAEs and islands for species richness (a) and species

614 richness residuals (b). Tests were performed for total and guild species richness. Species

615 richness residuals were estimated from linear mixed models with species richness as the

616 dependent variable, and pH, conductivity, elevation, mean annual temperature and

617 precipitation, temperature and precipitation seasonality and environmental heterogeneity

618 as explanatory variables and continent as a random factor. Comparisons of values (i.e.,

619 species richness or residuals of species richness) were performed with Cliff's test,

620 whereby tests with delta >0.33 indicated significant differences (Romano et al., 2006).

621
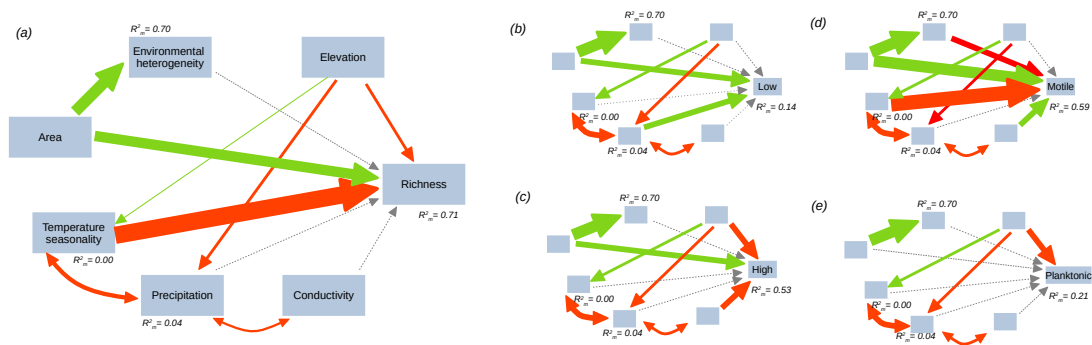
Figure 6. Structural equation models explaining species richness in continental area equivalents (CAEs). Structural equation models for total species richness (a), low profile (b), high profile (c), motile (d) and planktonic (e) species richness in continents (N = 96 CAEs). Green and red arrows represent significant positive and negative relationships, respectively, whereas gray-dashed arrows represent non-significant relationships. Arrow widths are proportional to the standardized regression coefficients and R^2_m values represent marginal R^2 from a linear mixed model. All models fitted well the *a priori* model, i.e. the model including all shown causal relationships (Fisher's C = 14.99, df = 14, p = 0.38 for all models).

632 **Data Availability Statement**

633 Data are available under the following link: https://doi.org/10.57745/ZPBSLT

634

635 **References**

Abonyi, A., Horváth, Z., & Ptacnik, R. (2018). Functional richness outperforms
taxonomic richness in predicting ecosystem functioning in natural phytoplankton
communities. *Freshwater Biology*, *63*(2), 178–186.

Algar, A. C., & Losos, J. B. (2011). Evolutionary assembly of island faunas reverses the
classic island–mainland richness difference in Anolis lizards. *Journal of
Biogeography*, *38*(6), 1125–1137.

Arrhenius, O. (1921). Species and area. *The Journal of Ecology*, 95–99.

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and
Applications with R*. Chapman and Hall/CRC Press.

Báldi, A. (2008). Habitat heterogeneity overrides the species–area relationship. *Journal
of Biogeography*, *35*(4), 675–681.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H.,
Dai, B., & Grothendieck, G. (2015). *lme4: Linear Mixed-Effects Models using
"Eigen" and S4* (1.1-8) [Computer software].

Bell, T., Ager, D., Song, J.-I., Newman, J. A., Thompson, I. P., Lilley, A. K., & Gast, C.
J. van der. (2005). Larger Islands House More Bacterial Taxa. *Science*,
308(5730), 1884–1884.

Catano, C. P., Grman, E., Behrens, E., & Brudvig, L. A. (2021). Species pool size alters
species–area relationships during experimental community assembly. *Ecology*,
102(1), e03231.

Chao, A. (1987). Estimating the Population Size for Capture-Recapture Data with
Unequal Catchability. *Biometrics*, *43*(4), 783–791.

Chase, J. M., Gooriah, L., & May, F. (2019). A framework for disentangling ecological
mechanisms underlying the island species–area. *Frontiers of Biogeography*.

Chase, J. M., & Leibold, M. A. (2003). *Ecological Niches: Linking Classical and
Contemporary Approaches*. University of Chicago Press.

Connor, E. F., & McCoy, E. D. (1979). The Statistics and Biology of the Species-Area
Relationship. *The American Naturalist*, *113*(6), 791–833.

Dahl, A. L. (1991). *Island Directory* (UNEP Regional Seas Directories and
Bibliographies No. 35; p. 573). UNEP.

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data
2010 (GMTED2010)* (U.S. Geo - Logical Survey Open-File Report No. 2011–
1073; p. 26).

DeMalach, N., Saiz, H., Zaady, E., & Maestre, F. T. (2019). Plant species–area
relationships are determined by evenness, cover and aggregation in drylands
worldwide. *Global Ecology and Biogeography*, *28*(3), 290–299.

Drakare, S., Lennon, J. J., & Hillebrand, H. (2006). The imprint of the geographical,
evolutionary and ecological context on species–area relationships. *Ecology
Letters*, *9*(2), 215–227.

Farina, A. (2008). *Principles and Methods in Landscape Ecology: Towards a Science of
the Landscape*. Springer.

Finlay, B. J. (2002). Global Dispersal of Free-Living Microbial Eukaryote Species.
Science, *296*(5570), 1061–1063.

Gillespie, R. G., Claridge, E. M., & Roderick, G. K. (2008). Biodiversity dynamics in
isolated island communities: Interaction between natural and human-mediated
processes. *Molecular Ecology,* 17(1), 45–57.

Hanski, I. A., & Gaggiotti, O. E. (2004). *Ecology, Genetics and Evolution of Metapopulations*. Elsevier Academic Press.

Heino, J., & Soininen, J. (2005). Assembly rules and community models for unicellular organisms: Patterns in diatoms of boreal streams. *Freshwater Biology*, *50*(4), 567–577.

Heino, J., & Soininen, J. (2006). Regional occupancy in unicellular eukaryotes: A reflection of niche breadth, habitat availability or size-related dispersal capacity? *Freshwater Biology*, *51*(4), 672–685.

Heino, J., & Soininen, J. (2007). Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biological Conservation*, *137*(1), 78–89.

Hijmans, R. J. (2019). *raster: Geographic Data Analysis and Modeling*.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978.

Jamoneau, A., Passy, S. I., Soininen, J., Leboucher, T., & Tison-Rosebery, J. (2018). Beta diversity of diatom species and ecological guilds: Response to environmental and spatial mechanisms along the stream watercourse. *Freshwater Biology*, *63*(1), 62–73.

Jüttner, I., Vijver, B. V. de, Williams, D. M., Lange-Bertalot, H., & Ector, L. (2018). The genus Eunotia (Bacillariophyta) in the Falkland Islands and species-area relationships in sub-Antarctic islands. *Diatom Research*, *33*(4), 413–452.

Kallimanis, A. S., Bergmeier, E., Panitsa, M., Georghiou, K., Delipetrou, P., & Dimopoulos, P. (2010). Biogeographical determinants for total and endemic species richness in a continental archipelago. *Biodiversity and Conservation*, *19*(5), 1225–1235.

Kalmar, A., & Currie, D. J. (2006). A global model of island biogeography. *Global Ecology and Biogeography*, *15*(1), 72–81.

Kalmar, A., & Currie, D. J. (2007). A Unified Model of Avian Species Richness on Islands and Continents. *Ecology*, *88*(5), 1309–1321.

Krause, S., Le Roux, X., Niklaus, P. A., Van Bodegom, P. M., Lennon, J. T., Bertilsson, S., Grossart, H.-P., Philippot, L., & Bodelier, P. L. E. (2014). Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology*, *5*.

Kreft, H., Jetz, W., Mutke, J., Kier, G., & Barthlott, W. (2008). Global diversity of island floras from a macroecological perspective. *Ecology Letters*, *11*(2), 116–127.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26.

Lack, D. (1976). *Island biology: Illustrated by the land birds of Jamaica*. Blackwell Scientific Publications, Oxford, UK.

Leboucher, T., Tison-Rosebery, J., Budnick, William. R., Jamoneau, A., Vyverman, W., Soininen, J., Boutry, S., & Passy, S. I. (2020). A new metacommunity approach for detecting species influenced by mass effect. *Journal of Applied Ecology*.

Lecointe, C., Coste, M., & Prygiel, J. (1993). "Omnidia": Software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia*, *269–270*(1), 509–513.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579.

Lomolino, M. V., & Weiser, M. D. (2001). Towards a More General Species-Area Relationship: Diversity on All Islands, Great and Small. *Journal of Biogeography*, *28*(4), 431–445. JSTOR.

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

Matthews, T. J., Cottee-Jones, H. E., & Whittaker, R. J. (2014). Habitat fragmentation and the species–area relationship: A focus on total species richness obscures the impact of habitat loss on habitat specialists. *Diversity and Distributions*, *20*(10), 1136–1146.

Matthews, T. J., Rigal, F., Triantis, K. A., & Whittaker, R. J. (2019). A global model of island species–area relationships. *Proceedings of the National Academy of Sciences*, *116*(25), 12337–12342.

Mittelbach, G. G., & McGill, B. J. (2019). *Community Ecology*. Oxford University Press.

Moser, G., Lange-Bertalot, H., & Metzeltin, D. (1998). *Insel der Endemiten. Geobotanisches Phänomen Neukaledonien—Island of Endemics. New Caledonia a geobotanical phenomenon* (J. Cramer). Schweizerbart'sche Verlagsbuchhandlung.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package*.

Passy, S. I. (2007). Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany*, *86*(2), 171–178.

Passy, S. I. (2016). Abundance Inequality in Freshwater Communities Has an Ecological Origin. *The American Naturalist*, *187*(4), 502–516.

Peay, K. G., Bruns, T. D., Kennedy, P. G., Bergemann, S. E., & Garbelotto, M. (2007). A strong species–area relationship for eukaryotic soil microbes: Island size matters for ectomycorrhizal fungi. *Ecology Letters*, *10*(6), 470–480.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446.

Pound, K. L., Larson, C. A., & Passy, S. I. (2021). Current distributions and future climate-driven changes in diatoms, insects and fish in U.S. streams. *Global Ecology and Biogeography*, *30*, 63–78.

Prugh, L. R., Hodges, K. E., Sinclair, A. R., & Brashares, J. S. (2008). Effect of habitat area and isolation on fragmented animal populations. *Proceedings of the National Academy of Sciences*, *105*(52), 20770–20775.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Reche, I., Pulido-Villena, E., Morales-Baquero, R., & Casamayor, E. O. (2005). Does Ecosystem Size Determine Aquatic Bacterial Richness? *Ecology*, *86*(7), 1715–1722.

Rimet, F., & Bouchez, A. (2012). Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. *Knowledge and Management of Aquatic Ecosystems*, *406*, 01.

Romano, J., Kromrey, J. D., Coraggio, J., & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys. *Annual Meeting of the Florida Association of Institutional Research*, 1–33.

Rosenzweig, M. L. (1995). *Species diversity in space and times*. Cambridge: Cambridge University Press.

Sandin, S. A., Vermeij, M. J. A., & Hurlbert, A. H. (2008). Island biogeography of Caribbean coral reef fish. *Global Ecology and Biogeography*, *17*(6), 770–777.

Schrader, J., König, C., Triantis, K. A., Trigas, P., Kreft, H., & Weigelt, P. (2020). Species–area relationships on small islands differ among plant growth forms. *Global Ecology and Biogeography*, *29*(5), 814–829.

Shmida, A., & Wilson, M. V. (1985). Biological determinants of species diversity. *Journal of Biogeography*, *12*(1), 1–20.

Simberloff, D. S., & Wilson, E. O. (1969). Experimental Zoogeography of Islands: The Colonization of Empty Islands. *Ecology*, *50*(2), 278–296.

Smith, V. H., Foster, B. L., Grover, J. P., Holt, R. D., Leibold, M. A., & deNoyelles, F. (2005). Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proceedings of the National Academy of Sciences*, *102*(12), 4393–4396.

Soininen, J. (2007). Environmental and Spatial Control of Freshwater Diatoms—A Review. *Diatom Research*, *22*(2), 473–490.

Soininen, J., Jamoneau, A., Rosebery, J., & Passy, S. I. (2016). Global patterns and drivers of species and trait composition in diatoms. *Global Ecology and Biogeography*, *25*(8), 940–950.

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, *17*(7), 866–880.

Tecchio, S., Chaalali, A., Raoux, A., Tous Rius, A., Lequesne, J., Girardin, V., Lassalle, G., Cachera, M., Riou, P., Lobry, J., Dauvin, J.-C., & Niquil, N. (2016). Evaluating ecosystem-level anthropogenic impacts in a stressed transitional environment: The case of the Seine estuary. *Ecological Indicators*, *61, Part 2*, 833–845.

Teittinen, A., & Soininen, J. (2015). Testing the theory of island biogeography for microorganisms   patterns for spring diatoms. *Aquatic Microbial Ecology*, *75*(3), 239–250.

Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., & Siemann, E. (1997). The Influence of Functional Diversity and Composition on Ecosystem Processes. *Science*, *277*(5330), 1300–1302.

Torchiano, M. (2020). *effsize: Efficient Effect Size Computation*.

Triantis, K. A., Economo, E. P., Guilhaumon, F., & Ricklefs, R. E. (2015). Diversity regulation at macro-scales: Species richness on oceanic archipelagos. *Global Ecology and Biogeography*, *24*(5), 594–605.

Whittaker, R. J., & Fernandez-Palacios, J. M. (2007). *Island Biogeography: Ecology, Evolution, and Conservation*. OUP Oxford.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature's laboratories. *Science*, *357*(6354).

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, *35*(6), 977–994.

636

637 **Biosketch**

638 Aurélien Jamoneau is a researcher in community ecology at INRAE institute of Cestas-

639 Gazinet in France. He is interested in aquatic community assemblages at multiple spatial

640 and temporal scales. Janne Soininen in a professor in spatial environmental research at

641 the University of Helsinki. He is interested in large-scale community ecology and

642 especially in the distribution of small aquatic organisms.