Correlations between broad-scale taxonomic and genetic differentiations suggest a dominant imprint of historical processes on beta diversities

Marine Robuchon, Boris Leroy, Céline Jézéquel, Bernard Hugueny

Correlations between broad-scale taxonomic and genetic differentiations suggest a dominant imprint of historical processes on beta diversities

Marine Robuchon, Boris Leroy, Céline Jézéquel, Bernard Hugueny

1   **Correlations between broad-scale taxonomic and genetic differentiations suggest a**

2   **dominant imprint of historical processes on beta diversities**

3   **Running title:** Broad-scale β-SGDCs in freshwater fishes

4   Marine Robuchon [1, 2 *], Boris Leroy [1], Céline Jézéquel [1, 3] and Bernard Hugueny [1, 3]

5   [1] Unité Biologie des organismes et écosystèmes aquatiques (BOREA), Muséum national

6   d'Histoire naturelle, Sorbonne Université, Université de Caen Normandie, Université des

7   Antilles, CNRS, IRD, CP 26, 57 rue Cuvier 75005 Paris, France

8   [2] Centre d'écologie et des sciences de la conservation (CESCO), Muséum national d'Histoire

9   naturelle, CNRS, Sorbonne Université, CP 135, 57 rue Cuvier 75005 Paris, France

10   [3] Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse

11   Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bat 4R1, 31062 Toulouse cedex 9,

12   France

13   * corresponding author: robuchon@mnhn.fr

14   **Acknowledgements**

Abstract

– **Aim:** Dispersal limitation, environmental selection and drift are known to influence both taxonomic similarity between communities and genetic similarity between populations. However, disentangling the relative roles of these processes on spatial patterns of differentiation - whether regarding taxonomic differentiation between communities or genetic differentiation between populations - is challenging. Investigating whether spatial patterns of taxonomic differentiation and genetic differentiation are correlated (β-SGDCs) is a promising approach to address this issue. Here, we investigated β-SGDCs over broad spatial scales and 22 freshwater fish species to elucidate the processes shaping taxonomic and genetic differentiations between drainage basins.

– **Location:** Global scope, data mainly from Europe and North America.

– **Taxon:** Actinopterygii and Petromyzontiformes (freshwater fishes).

– **Methods:** We used Mantel tests to investigate the raw correlation between taxonomic and genetic differentiations. We carried out multiple regressions to characterise the effects of geographic distance (proxy of dispersal limitation), environmental distance (proxy of environmental selection) and pairwise harmonic mean area between basins (proxy of drift) on taxonomic differentiation and genetic differentiation taken separately. We then analysed the correlation between the residuals of these two regressions with Mantel tests. Finally, we investigated whether the variation in β-SGDC between species was related to species traits.

– **Results**: We detected an overall positive β-SGDC, both for the raw and the residual correlations. This implies that, unmeasured effects, other than geographic distance, environmental distance and harmonic mean area, influence the β-SGDC observed. Values of β-SGDCs greatly varied between species, but this was not explained by any species traits.

48   –   **Main conclusions**: Independently of the effects of geographic distance, environmental

49   distance and harmonic mean area, the β-SGDC we observed suggests that historical

50   processes strongly shaped the patterns of taxonomic and genetic differentiations between

51   basins. Consequently, taxonomic differentiation may be an appropriate proxy to explain

52   the influence of historical processes on genetic differentiation.

53   **Keywords**

54   β-SGDC, dispersal limitation, distance decay, drift, environmental selection, freshwater

55   fishes, historical processes, macrogenetics, Mantel tests, multiple regression of distance

56   matrices

Introduction

59 Understanding the mechanisms that shape taxonomic differentiation between spatially

60 separated communities and those influencing genetic differentiation between spatially

61 separated populations are long-standing issues in ecology and evolution (Wright, 1943, 1951;

62 Whittaker, 1960, 1972; Preston, 1962; Mayr, 1963). More recently, these issues have been

63 united (Vellend, 2003; Vellend & Geber, 2005; Kahilainen, Puurtinen, & Kotiaho, 2014),

64 emphasising that drift, dispersal and selection influence both taxonomic differentiation

65 between communities and genetic differentiation between populations. Drift and selection

66 affect the species frequencies within communities and gene frequencies within populations,

67 which tend to increase differentiation. Conversely, dispersal of individuals between

68 populations and dispersal of species between communities will tend to decrease

69 differentiation.

70   Among the factors that can affect both taxonomic and genetic differentiation,

71 geographic distance is probably the most studied. The similarity in taxonomic composition

72 between communities tends to decrease with the geographic distance that separates them; a

73 pattern long-recognised in biogeography and formalised as the distance decay of similarity

74 (Nekola & White, 1999; Soininen, McDonald, & Hillebrand, 2007). Similarly, the genetic

75 similarity between populations tends to decrease with the geographic distance separating

76 them; a pattern long-recognised in population genetics and referred to as isolation-by-distance

77 (IBD; Wright, 1943; Rousset, 1997). Two main non-mutually exclusive explanations can

78 account for the distance decay of taxonomic and genetic similarity (see Soininen et al., 2007

79 and Orsini et al., 2013 for reviews). First, similarity decreases with geographic distance

80 because the dispersal of organisms (i.e. their movement from a place to another) is limited by

81 their intrinsic dispersal ability, which hence does not counteract anymore the differentiating

82 effect of drift. Secondly, similarity decays with geographic distance because of decreasing

83   similarity in spatially correlated environmental features; in such cases, the underlying

84   explanation is environmental selection of different species (community differentiation) or

85   genotypes (genetic differentiation) in distinct environments. However, differentiation between

86   sites does not always increase with geographic distance. This is the case when the migrants

87   are competitively excluded by already established communities (e.g. Almany, 2003) or

88   populations (e.g. Fraser et al., 2018) at high densities, therefore counteracting the

89   homogenising effect of dispersal whatever the intrinsic dispersal ability of the migrants. This

90   may also happen when the environmental variation between sites displays no – or negative –

91   spatial autocorrelation (e.g. Derry et al., 2009).

92   Although spatial patterns of taxonomic and genetic differentiations can be explained

93   by environmental selection independently of limited dispersal and vice-versa, disentangling

94   the relative strengths of these mechanisms is challenging, as communities and populations

95   that are geographically distant are also likely to inhabit different environments. Using modern

96   spatial statistical methods, some recent studies have nonetheless taken up the challenge. In

97   their study of global patterns of species turnover in terrestrial vertebrates, Qian & Ricklefs

98   (2012) have shown that both dispersal limitation and environmental selection have played

99   important roles in determining the patterns they observed. Similarly, in their reinterpretation

100 of 34 representative studies, Orsini *et al*. (2013) highlighted that patterns associated with

101 dispersal limitation were as common as those associated with local genetic adaptation in

102 structuring population genetic differentiation in the wild. However, the relative strength of

103 dispersal limitation and environmental selection on differentiation patterns seems to vary

104 according to the species and the spatial scale under consideration, both for taxonomic

105 (Astorga et al., 2012; Moritz et al., 2013) and genetic (Orsini et al., 2013) differentiations.

106 Comparative studies of multiple taxonomic groups suggest that the influence of dispersal

107 limitation on taxonomic differentiation is stronger for species with low dispersal ability than

108   for more effective dispersers, and this result holds true both at the broad-scale (e.g. Qian &

109   Ricklefs, 2012) and at the fine-scale (e.g. Astorga et al., 2012). Comparative studies of

110   multiple species have also been identified as a promising avenue to reveal the factors

111   influencing patterns of genetic differentiation (Wang et al., 2013). Yet, such studies are scarce

112   and mainly carried out at relatively small spatial scales (e.g. Wang et al., 2013; Fourtune, Paz-

113   Vinas, Loot, Prunier, & Blanchet, 2016) despite the increasing availability of broad-scale

114   datasets in population genetics, suggesting that it is now time to embrace macrogenetics

115   (Blanchet, Prunier, & De Kort, 2017).

116   Theories in population genetics and community ecology acknowledge that dispersal,

117   selection and drift together shape the differentiation between populations and between

118   communities (see e.g. Vellend & Orrock, 2009 for a review of processes in both disciplines).

119   However, the contribution of drift to taxonomic or genetic differentiation has been the subject

120   of few empirical investigations, in contrast with dispersal limitation and environmental

121   selection (Gilbert & Levine, 2017; Prunier, Dubut, Chikhi, & Blanchet, 2017). Genetic drift is

122   the evolutionary process of random fluctuations in allelic frequencies occurring naturally in

123   all populations due to their finite size, although it is stronger in small ones (Allendorf, 1986).

124   Similarly, ecological drift corresponds to the random fluctuations in species frequencies

125   occurring naturally in all communities due to their finite size and this too is stronger in small

126   ones. Drift is the result of random sampling during the processes of birth, death and

127   reproduction and ultimately leads to the loss of genetic diversity within populations, and

128   species diversity within communities. Therefore, drift increases the differentiation between

129   populations and between communities depending on their respective sizes: the smaller they

130   are, the more they will differentiate. Dispersal of organisms between populations and between

131   communities counteracts the differentiating effect of drift while environmental selection can

132   accelerate it. Attributing the observed spatial patterns of genetic or taxonomic differentiation

133   to dispersal limitation or to environmental selection without accounting for drift may thus be

134   misleading.

135   Most of the explanatory variables used in analyses to infer the role of dispersal

136   limitation and environmental selection on differentiation patterns are proxies representing

137   present-day conditions. Thus, using current explanatory variables to investigate dispersal

138   limitation and environmental selection on differentiation patterns is only relevant if the

139   contemporary geography and environment of the study area has not changed substantially

140   since these patterns were established (Wang et al., 2013). Otherwise, the historical processes

141   that may have shaped the observed differentiation patterns, including those explaining

142   colonisation histories, may be overlooked. In addition, if dispersal is not continuous (as

143   generally assumed) but rather intermittent because of landscape dynamics which have induced

144   the appearance and disappearance of dispersal barriers, then two geographically close

145   localities sharing a similar environment could, nevertheless, be dissimilar in terms of

146   community composition and allele frequencies provided they have been isolated for a

147   sufficiently long time (given their sizes) for drift to be effective. Nonetheless, examining

148   which part of differentiation patterns is not explained by current explanatory variables may

149   help to elucidate the influence of historical processes on those patterns.

150   In such complex contexts, understanding the rules that govern differentiation patterns

151   (both taxonomic and genetic) may be enhanced by comparing taxonomic dissimilarities

152   between communities and genetic dissimilarities between populations of a focal species

153   (Lamy, Laroche, David, Massol, & Jarne, 2017). The few empirical studies that have

154   examined these correlations, called beta species-genetic diversity correlations (β-SGDCs)

155   (Kahilainen et al., 2014), revealed a majority of positive correlations (reviewed by Lamy et

156   al., 2017). This suggests that the processes affecting taxonomic dissimilarity affect genetic

157   dissimilarity of the species under study in the same way (Baselga et al., 2013; Baselga,

158   Gómez-Rodríguez, & Vogler, 2015). However, because the relative strength of processes

159   shaping patterns of taxonomic and genetic differentiation is expected to vary according to the

160   species and the spatial scale under consideration, so is the strength of the putative resulting β-

161   SGDCs. Among the rare studies of β-SGDCs over multiple species, some have highlighted

162   that positive β-SGDCs were stronger for focal species with low dispersive abilities (e.g.

163   Papadoulou et al., 2011) while others did not find strong differences between species (e.g.

164   Fourtune et al., 2016), and all were carried out at relatively small spatial scales. These

165   contrasting results call for more studies of β-SGDCs for multiple species to gain a better

166   understanding of whether and how variations in species traits lead to distinct patterns of β-

167   SGDCs. In addition, β-SGDCs for multiple species have not been examined yet at broad

168   spatial scales.

169   Freshwater fishes are a relevant model to start investigating this question for at least

170   three reasons. First, as β-SGDCs for multiple freshwater fish species have been conducted at

171   small scales within a single drainage basin (Fourtune et al., 2016), computing β-SGDCs for

172   multiple freshwater fish species at broader scales will permit the comparison of β-SGDC

173   variations across scales for the same biological model. Secondly, for strictly freshwater fishes,

174   at present, almost no exchange of individuals occurs between drainage basins (hereafter

175   "basins") because they are isolated from each other by land and/or sea (although dispersal

176   may occasionally occur between adjacent basins). Therefore, if basins are isolated from each

177   other independently of the geographic distance between them, we can expect dispersal

178   limitation to play a minor role on the broad-scale between-basin patterns of taxonomic and

179   genetic differentiations. On the contrary, if basins have been isolated from each other for a

180   long time, we can expect drift to play a significant role on the broad-scale between-basin

181   patterns of taxonomic and genetic differentiations, as dispersal no longer counteracts the

182   effect of drift. This is particularly true if population sizes and community sizes are small: the

183   larger they are, the longer it will take to observe differentiation. Finally, while several recent

184   studies have tried to disentangle the role of dispersal limitation, environmental selection and

185   historical processes on patterns of taxonomic differentiation between basins (Leprieur et al.,

186   2009, 2011; Dias et al., 2014), the processes shaping genetic differentiation at such broad

187   scales have rarely been examined together. Studying β-SGDCs may therefore reveal whether

188   the processes influencing between-basin taxonomic differentiation, affect between-basin

189   genetic differentiation in the same way.

190   In this meta-analysis of 22 freshwater fish species, our main goal was to elucidate the

191   processes underlying β-SGDCs at broad spatial scales. In particular, we wanted to test

192   whether β-SGDCs can be explained by the parallel action of measurable processes (dispersal

193   limitation, environmental selection and drift) on taxonomic and genetic differentiation, or are

194   due to other, non-measurable processes. To that end, we first computed measures of genetic

195   differentiation between basins for each fish species. Then, for each species, we investigated i)

196   the raw correlation between taxonomic and genetic differentiations, ii) the effects of

197   geographic distance (as a proxy of dispersal limitation), environmental distance (as a proxy of

198   environmental selection) and pairwise harmonic mean area between basins (as a proxy of

199   drift) on taxonomic differentiation and genetic differentiation taken separately and iii) the

200   correlation between taxonomic and genetic differentiations independent of the effects of

201   geographic distance, environmental distance and harmonic mean area (Fig. 1). Such

202   independent correlation integrates the effects of unmeasured processes, including the

203   historical processes of dispersal, environmental selection and drift related to the historical

204   connectivity of basins. This allowed us to characterise the overall (i.e. mean over the species)

205   raw β-SGDC, the overall effects associated with each predictor for each response variable,

206   and the overall independent β-SGDC. Finally, we tested whether the variation in β-SGDCs

207   between species could be attributed to variations in species traits.

208

209   **Materials and methods**

210   *Genetic differentiation and geographic distribution of genetic data*

211   As we wanted to investigate genetic differentiation at the broad scale, we restricted our

212   analyses of spatial genetic differentiation to species exhibiting a wide distribution. We

213   downloaded sequence records of the mitochondrial gene cytochrome c oxidase I (COI) from

214   BOLD (www.boldsystems.org) in May 2017 for the following widespread taxa: Cyprinidae,

215   Salmonidae, *Lampetra*, *Petromyzon*, *Cobitis*, *Misgurnus*, *Barbatula*, *Silurus*, *Esox*, *Lota lota*,

216   *Gymnocephalus* and *Sander*. We first cleaned and aligned sequences by taxon and then re-

217   aligned them together to keep the same fragment of 501 nucleotides for the analyses.

218   Sequence records without sufficient geographic information to assign them to a drainage

219   basin (hereafter "basin") or without a species name were discarded from our dataset. We

220   defined a basin as the drainage area upstream from its mouth at the sea. According to this

221   definition, a basin is isolated from other basins by barriers (sea or land) which are impassable

222   for strictly freshwater fishes. The validity of species names was assessed using FishBase

223   (Froese & Pauly, 2017), resulting in a total of 6637 sequence records with a valid species

224   names belonging to 716 distinct species distributed in 218 basins. We checked the

225   environment (i.e. "freshwater", "brackish", "saltwater") of each species using FishBase

226   (Froese & Pauly, 2017), the status (i.e. "native", "exotic" or "unknown") of each species in

227   each basin using the global database on freshwater fish species occurrence (Tedesco et al.,

228   2017). We conserved only native records of strictly freshwater species, i.e. 4894 records,

229   corresponding to a total of 509 species in 176 basins.

230   We calculated between-basin genetic differentiation (by pooling individuals in the

231   drainage basins) for the polymorphic species present in at least four basins with a minimum of

232   three individuals per basin. To calculate such genetic differentiation, we converted single

233   nucleotide polymorphisms (SNPs) of the sequence data into a table of individual genotypes

234   and then calculated pairwise genetic distances for each pair of basins using G''_{ST} (Meirmans

235   & Hedrick, 2011), the corrected version of Hedrick's G'_{ST} (Hedrick, 2005). G''_{ST} is

236   independent from within population diversity and does not underestimate genetic distance

237   when the number of populations is small (Meirmans & Hedrick, 2011). As G''_{ST} likely

238   removes the effect of drift, we also calculated the pairwise genetic distances for each pair of

239   basins using F_{ST} (Nei, 1973) for comparison purpose.

240   All these different selection filters drastically reduced the number of species we

241   included in analyses: we finally focused our analyses of between-basin genetic differentiation

242   on a total of 22 species in 38 basins (for more details about these basins, see Appendix S1 in

243   Supporting Information). For instance, from all the species of *Lampetra* and *Petromyzon* we

244   searched in BOLD, we conserved only *L. planeri* for the analyses. The number of species

245   with analysed genetic data per basin varies from one (for 20 basins) to seventeen (for one

246   basin, the basin "Odra"; Fig. 2; Appendix S1). The number of basins per species varies from

247   four (our minimum threshold) to eleven (for the species *Esox lucius*) covering on average 76

248   % (standard deviation (sd) = 18 %) of their native range, with species exhibiting on average

249   11.0 individuals per basin (sd = 11.0; Appendix S2). Most records are located in Europe and

250   North-America (Fig. 2).

251   *Taxonomic differentiation*

252   We calculated pairwise taxonomic differentiation between each pair of basins based on the list

253   of native species per basin extracted from the global database on freshwater fish species

254   occurrence (Tedesco et al., 2017). We used Simpson dissimilarity (β_{sim}) - the turnover

255   component of Sørensen dissimilarity (Sørensen, 1948) - which is independent from total

256   taxonomic richness (see Baselga 2012 for review).

257   *Geographic distance, environmental distance and pairwise harmonic mean area*

258   We computed between-basin geographic distance by calculating the distance between basin

259   centroids. The average distance between basins was 5 295 km (sd = 3 920 km). Measurements

260   of geographic distance were then centred and scaled before further analyses.

261   To characterise the environment of the basins, we used variables related to the

262   contemporary climate known to influence freshwater fish alpha diversity patterns at the global

263   scale (e.g. Dias et al. 2014; Oberdorff et al. 2011; Tedesco et al. 2012; Tisseuil et al. 2013) i.e.

264   mean annual temperature and precipitation (http://www.worldclim.org/), mean annual surface

265   runoff (http://www.grdc.sr.unh.edu/) and mean annual actual and potential evapotranspiration

266   (http://csi.cgiar.org/Aridity/). The mean values of all the variables were computed for each

267   basin. To compute pairwise environmental distances between each pair of basins, we

268   performed a principal component analysis (PCA) on these environmental variables

269   (previously centred and scaled) and calculated the Euclidean distance between basins based

270   on the first three PCA axes.

271   Finally, we calculated the pairwise harmonic mean area between each pair of basins.

272   Indeed, the harmonic mean population size between pairs allows the role of drift on genetic

273   differentiation (Serrouya et al., 2012) to be taken into account and can be computed using

274   environmental proxies for local carrying capacities (Prunier et al., 2017). In our case, we used

275   the surface area of the basin as a raw environmental proxy for local carrying capacity. The

276   basins studied had an average surface area of 257 589 km^2 (sd = 584 803 km^2). Measurements

277   of harmonic mean area were centred and scaled before further analyses.

278   *Statistical analyses*

279   First, we investigated the raw correlation between taxonomic differentiation and genetic

280   differentiation for each species individually using a Mantel correlation test with 999

281   permutations. To test the null hypothesis that there is no relationship between taxonomic

282   differentiation and genetic differentiation using the information provided by all the species

283   examined, we considered that this hypothesis has been tested independently *n* times, with *n*

284   being the number of species included in the analyses. To combine the results from several

285   independent tests bearing upon the same overall hypothesis, we used a modification of the

286   Fisher's combined probability test**.** Under the null hypothesis that taxonomic and genetic

287   differentiations are unrelated, observed p-values associated with the Mantel's test of each

288   species are expected to be distributed according to a uniform distribution in [0, 1]. We

289   therefore compared the observed mean p-value over the species to a null distribution of mean

290   p-values obtained by sampling randomly 10 000 times one p-value by species in a uniform

291   distribution in [0, 1]. We calculated the p-value associated with the combined test as the

292   frequency at which null mean p-values were below the observed mean p-value. This

293   procedure avoids some problems identified by the use of a chi-2 test (e.g. Whitlock, 2005)

294   which was the method initially proposed by Fisher.

295   To investigate the putative roles of dispersal limitation, environmental selection and

296   drift on taxonomic and genetic differentiations, for each species we built multiple regressions

297   of geographic distance (D_{geo}), environmental distance (D_{env}) and harmonic mean area (D_{area})

298   on taxonomic differentiation (TD) and genetic differentiation (GD) taken separately (Fig. 1)

299   as follows: $TD = \alpha_0 + \alpha_1 {*} Dgeo + \alpha_2 {*} D_{env} + \alpha_3 {*} D_{area}$ and $GD = \beta_0 + \beta_1 {*} Dgeo + \beta_2 {*} D_{env} +$

300    $\beta_3 {*} D_{area}$. Our approach is very similar to multiple regressions on distance matrices (MRM;

301   Lichstein 2007) except in the way we calculated p-values associated with partial regression

302   coefficients. As in MRM, we permuted the rows and associated columns of the response

303   distance matrix simultaneously. We repeated this operation 1999 times while holding the

304   explanatory distance matrices constant to generate null distributions for partial regression

305   coefficients. Then, in MRM, the p-value associated with a partial regression coefficient is

306   calculated with a two-sided permutation test using the pseudo-t of Legendre et al. (1994), i.e.

307   it tests whether the t-statistic associated with each regression coefficient is lower or higher

308   than expected under a null distribution of pseudo-t obtained by permutation. In our approach,

309   we hypothesised that geographic distance and environmental distance have a positive effect

310   on both taxonomic and genetic differentiations (dispersal limitation and environmental

311   selection enhance differentiation and are positively related to geographic and environmental

312   distances respectively) while harmonic mean area has a negative effect on taxonomic and

313   genetic differentiations (drift enhances differentiation and is negatively related to the

314   harmonic mean area). To increase the statistical power of the tests, we therefore carried out

315   one-sided tests for calculating the p-values associated with each partial regression coefficient.

316   We tested the overall (i.e. over the species) effects of geographic distance, environmental

317   distance and harmonic mean area on taxonomic and genetic differentiations by performing a

318   combined probability test as described for the raw correlation between taxonomic and genetic

319   differentiations, but this time considering probabilities associated with partial regression

320   coefficients.

321   To examine whether taxonomic and genetic differentiations were still correlated once

322   the effects of geographic distance and environmental distance had been taken into account, we

323   then performed a Mantel correlation test between the residuals of the multiple regression on

324   taxonomic differentiation and the residuals of the multiple regression on genetic

325   differentiation with 999 permutations for each species. We tested this overall (i.e. over the

326   species) residual correlation using a combined probability test as described for the raw

327   correlation between taxonomic and genetic differentiations.

328   Finally, we examined the heterogeneity between species regarding the residual

329   correlation between taxonomic and genetic differentiations and tested whether some species

330   traits could explain such heterogeneity. The species traits initially considered were body

331   length, longevity in the wild, vulnerability, used in aquaculture, use as bait, use in the

332   aquarium, habitat, migratory behaviour and dispersal ability – which was calculated using a

333   formula giving dispersal distance as a function of body size and caudal fin aspect ratio

334   (Radinger & Wolter, 2014). However, as migratory behaviour was 'potamodromous' for 16

335   out of the 22 species, 'non-migratory' for one species and the information was not available

336   for the 5 species left, we excluded this non-informative trait from our analyses. Species traits

337   were directly extracted from FishBase (Froese & Pauly, 2017). All the analyses and the

338   corresponding figures were realised with R 3.5.1 (R Core Team 2018) using the packages

339   'ade4' (Dray & Dufour, 2007), 'ape' (Paradis, Claude, & Strimmer, 2004), 'betapart'

340   (Baselga, Orme, Villeger, De Bortoli, & Leprieur, 2018), 'Biostrings' (Pagès, Aboyoun,

341   Gentleman, & DebRoy, 2018), 'broom' (Robinson & Hayes, 2018), 'ggplot2' (Wickham,

342   2016), 'ggthemes' (Arnold, 2018), 'hierfstat' (Goudet & Jombart, 2015), 'maptools' (Bivand

343   & Lewin-Koh, 2018), 'mmod' (Winter, 2012), 'RColorBrewer' (Neuwirth, 2014), 'rgdal'

344   (Bivand, Keitt, & Rowlingson, 2018), 'rgeos' (Bivand & Rundel, 2018), 'reshape2'

345   (Wickham, 2007), 'seriation' (Hahsler, Buchta, & Hornik, 2018) and 'vegan' (Oksanen et al.,

346   2018).

347

348   **Results**

349   We detected an overall positive correlation between taxonomic and genetic differentiations,

350   significantly different from 0. This result holds true both for the raw correlation between

351   taxonomic and genetic differentiations ($r_{mean \pm sd} = 0.433 \pm 0.420$; Fig. 3a; Appendices S3, S4)

352   and for the residual correlation, i.e. the correlation between taxonomic and genetic

353   differentiations independent of the effects of geographic distance, environmental distance and

354   harmonic mean area ($r_{mean \pm sd}$ = 0.379 ± 0.641; Fig. 3b; Appendices S3, S4).

355   There was almost no overall effect of geographic distance on taxonomic differentiation

356   ($\alpha_{1\ mean \pm sd}$ = 0.032 ± 1.051) or on genetic differentiation ($\beta_{1\ mean \pm sd}$ = -0.011 ± 1.060), and

357   these overall effects did not differ significantly from 0 (Fig. 4; Appendices S5, S6, S7). In

358   contrast, we detected an overall positive effect of environmental distance on both taxonomic

359   differentiation ($\alpha_{2\ mean \pm sd}$ = 0.405 ± 0.956) and genetic differentiation ($\beta_{2\ mean \pm sd}$ = 0.229 ±

360   0.865), although it was only significantly different from 0 for taxonomic differentiation (Fig.

361   4; Appendices S5, S6, S7). There was an overall negative effect of harmonic mean area on

362   both taxonomic differentiation ($\alpha_{3\ mean \pm sd}$ = -0.178 ± 0.617) and genetic differentiation ($\beta_{3\ mean}$

363    $_{\pm sd}$ = -0.092 ± 0.713), but it was only significantly different from 0 for taxonomic

364   differentiation (Fig. 4; Appendices S5, S6, S7). Using F_{ST} as a measure of genetic

365   differentiation did not change the trend we observed: there was an overall negative but not

366   significant effect of harmonic area on F_{ST} (Appendices S6, S7).

367   The results of these multiple regressions as well as the residual correlations between

368   taxonomic and genetic differentiations are summarised in Fig. 5. Overall, they suggest that the

369   effect of geographic distance, environmental distance and harmonic mean area only explain a

370   small part of the correlation between taxonomic and genetic differentiations.

371   Despite this strong overall independent correlation between taxonomic differentiation

372   of freshwater fish communities and genetic differentiation of freshwater fish populations, the

373   correlation between taxonomic and genetic differentiations varied greatly between species

374   (Fig. 3b; Appendix S3). For some species, such as the common roach *Rutilus rutilus* (Fig. 6a,

375   b), there was a sound correlation between taxonomic differentiation and genetic

376 differentiation, whether we considered the raw or residual correlation. In other species, such

377 as the common chub *Squalius cephalus*, taxonomic differentiation and genetic differentiation

378 were not correlated (Fig. 6c, d). In this species, genetic differentiation was better predicted by

379 geographic distance and harmonic mean area (Appendix S5). However, none of the species

380 traits examined explained this heterogeneity of responses between species (Appendix S8).

381

382 **Discussion**

383 Our analyses indicated that taxonomic and genetic differentiations are strongly correlated

384 overall, and independently of the parallel effects of geographic distance, environmental

385 distance and harmonic mean area on taxonomic and genetic differentiations. This suggests

386 that the parallel influence of dispersal limitation, environmental selection and drift on

387 taxonomic and genetic differentiations only partially explains the overall β-SGDC we

388 observed. This independent, unexplained correlation between taxonomic and genetic

389 differentiations thus probably results from processes that we did not explicitly consider in our

390 study. The most probable explanation is that the independent β-SGDC we observed is due to

391 the parallel action of historical processes, which occurred when the basins were still

392 connected, on taxonomic and genetic differentiations. These historical processes include all

393 the different processes that explain the colonisation history of basins (Orsini et al., 2013): i)

394 the dispersal of individuals between previously connected basins in interaction with

395 environmental and/or biotic selection on their dispersal route and ii) environmental and/or

396 biotic selection in interaction with drift within their place of arrival and/or origin that may

397 have led to extinction. This explanation is supported by previous studies on the influence of

398 historical processes on broad-scale taxonomic differentiation between freshwater fish

399 communities on the one hand and on broad-scale genetic differentiation between freshwater

400   fish populations on the other hand. Two recent studies on global freshwater fish biodiversity

401   have shown that present-day patterns of taxonomic dissimilarity across basins are well

402   explained by the historical connectivity of basins (Dias et al., 2014) and also historical

403   climatic oscillations that took place during the Quaternary and the Holocene (Leprieur et al.,

404   2011). For fish species, it is well known that historical dispersal influences genetic structure

405   across basins (Carvalho, 1993), especially for zones that were subject to intense post-glacial

406   recolonisation such as our study area. Multispecies comparisons at large spatial scales in

407   Europe (Seifertová, Bryja, Vyskočilová, Martínková, & Šimková, 2012) and in North

408   America (April, Hanner, Mayden, & Bernatchez, 2013) have shown that the climatic

409   fluctuations that caused glacial cycles during the Pleistocene had a generalised effect on the

410   patterns of fish genetic divergence. In addition, three other, non-exclusive hypotheses may

411   explain the independent β-SGDC we observed. First, the proxies we used to account for

412   dispersal limitation, environmental selection and drift may not fully represent these processes.

413   For instance, we may have overlooked some environmental variables that are important in

414   driving environmental selection. Alternatively, we may have poorly estimated the role of drift,

415   either because the area of the basin we used to account for drift may be too approximate to

416   accurately represent the carrying capacities of populations and communities or because we

417   did not consider that the influence of drift within a basin increases with its age of isolation.

418   Secondly, the sequence data we used to calculate genetic differentiation may not detect

419   dispersal limitation and/or environmental selection properly, either because they did not cover

420   the entire native ranges of species (see possible consequences of such sampling bias in e.g.

421   Meirmans, 2015), or because the mitochondrial gene COI is not directly under environmental

422   selection. Finally, the proposed frameworks to interpret SGDCs (Vellend & Geber, 2005;

423   Lamy et al., 2017) suggest that correlations between species diversity and genetic diversity

424   may also result from a causal action of species diversity on genetic diversity or vice-versa.

425   For instance, if the focal species (i.e. the species for which we studied genetic diversity) is a

426   facilitator for the other component species of the community, we would expect their

427   populations sizes to co-vary positively, resulting in a positive α-SGDC; the converse is

428   expected if the focal species is a competitor. If the nature of the interaction between the focal

429   species and the other species of the community is generally the same within each basin, such

430   causal effect is not expected to influence the β-SGDC we observed. However, if the focal

431   species is facilitator in one basin and competitor in another, we would expect the population

432   sizes of the other component species of the community to increase in the first case and

433   decrease in the second case. This would result in more ecological drift within the community

434   where the focal species is competitor, and therefore more taxonomic differentiation between

435   basins than would be expected without this causal effect. In the end, such causal effect would

436   tend to reduce β-SGDC because it would increase taxonomic differentiation between basins

437   without influencing genetic differentiation.

438   Moreover, our results contribute new insights into the study of β-SGDCs. With the

439   study of 22 new β-SGDCs, we increased the number of β-SGDCs (43) previously reported in

440   the literature by more than 50 % (see Lamy et al., 2017 for the most recent review). We found

441   an average raw correlation between taxonomic and genetic differentiations of 0.433, and an

442   independent correlation (i.e. independent from the parallel effects of geographic distance,

443   environmental distance and harmonic mean area on taxonomic and genetic differentiations) of

444   0.379. Both of these values are superior to the average value of 0.221 reported by Lamy *et al.*

445   (2017). They are also around three times superior to the average value of 0.139 found by

446   Fourtune *et al.* (2016) in their study of four freshwater fish species at a smaller scale, within

447   one drainage basin. This suggests that the strong correlations we found are probably more

448   related to the size and/or the isolation of the habitats we studied (i.e. large drainage basins

449   isolated from each other) rather than our biological model. Previous syntheses have reported

450   that α-SGDCs (i.e. correlations between genetic diversity within populations and species

451   diversity within communities) are stronger and more often positive in island-like habitats

452   (Vellend & Geber, 2005; Vellend et al., 2014), in agreement with the predictions of

453   theoretical models (Laroche et al., 2015). As drainage basins are typically island-like habitats

454   (Sepkoski & Rex, 1974; Hugueny, 1989), our findings suggest that this explanation may hold

455   true for β-SGDCs as well. The observed differences also suggest that for a system without

456   present-day dispersal the β-SGDC is stronger than in systems with present-day dispersal.

457   Actually, the opposite could have been expected (i.e. weaker β-SGDCs for systems without

458   present-day dispersal than for systems with present-day dispersal) if the genetic similarity

459   between two previously connected populations had been erased by genetic drift occurring

460   after their isolation without changes in species composition (i.e. no species extinction).

461   However, our results suggest that drift did not influence significantly the patterns of genetic

462   differentiation we observed. As this result holds true both for G''_{ST} and F_{ST}, this is probably

463   because we mainly focused on large drainage basins and consequently large population sizes.
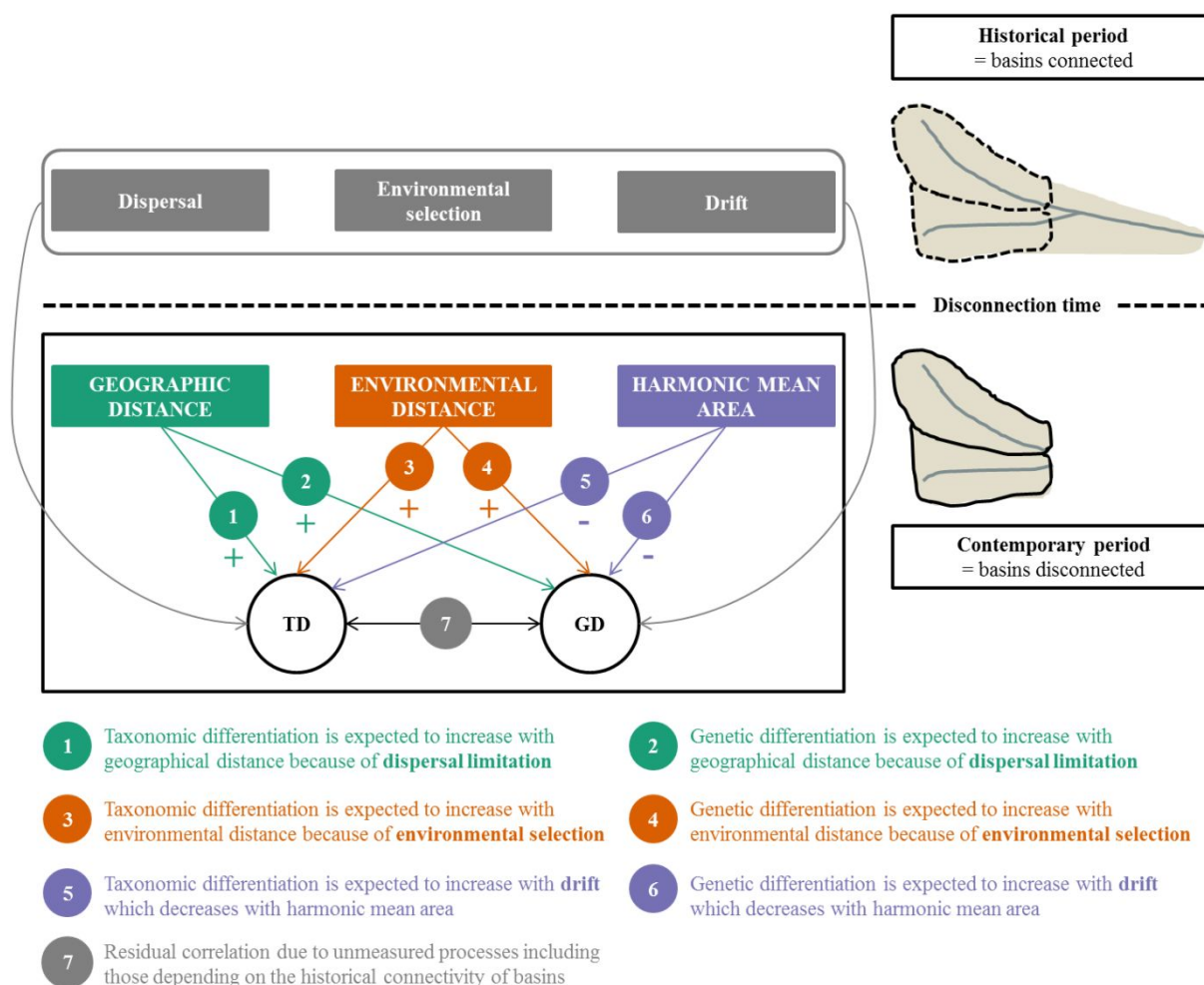
464   Our results also indicate that the sign and magnitude of β-SGDCs varied greatly

465   between the different species we examined. We provided two examples. On the one hand, we

466   showed that the β-SGDC was strong for *Rutilus rutilus*, suggesting that historical processes

467   largely shaped the patterns of genetic differentiation observed in this species. This is

468   consistent with a previous study indicating that *R. rutilus* is composed of two historically

469   isolated, independently evolving sets of populations (Larmuseau et al., 2009). On the other

470   hand, our results indicate that genetic differentiation in *Squalius cephalus* was not correlated

471   to taxonomic differentiation, but was well explained by geographic distance. This suggests

472   that dispersal limitation has a more important effect than any other processes on broad-scale

473   genetic differentiation in *S. cephalus*. This might seem surprising, since previous analyses of

474   the mitochondrial gene cytochrome b in that species indicated four main lineages originating

475   from multiple glacial refugia (Durand, Persat, & Bouvet, 1999; Seifertová et al., 2012).

476   However, the use of a combination of both mitochondrial and nuclear DNA, Seifertová et al.

477   (2012) also suggested that genetic differentiation in *S. cephalus* resulted from a combination

478   of different factors, i.e. post-glacial colonisation from different refugia or recent evolutionary

479   processes such as drift or dispersal limitation. Our results based on the analysis of the

480   mitochondrial gene COI strongly support the dispersal limitation hypothesis although, as we

481   did not design the sampling, but analysed existing data, we might not have captured the whole

482   picture. As *R. rutilus* and *S. cephalus* have similar dispersal abilities, this difference of β-

483   SGDC between the two species does not seem to be related to their difference in dispersal

484   abilities. Beyond these two species and this single trait, our analyses did not permit us to

485   uncover any trait that could explain the variability of β-SGDC over the 22 species we

486   examined. This is surprising because life-history traits of species are known to shape spatial

487   patterns of genetic differentiation (e.g. Duminil et al., 2007; Kelly & Palumbi, 2010), and

488   therefore β-SGDC. This unexpected finding may be due to different reasons. First, the

489   variability of β-SGDC may be due to species traits that we did not consider. Secondly, the β-

490   SGDC at the specific scale was sometimes examined for a small number of basins, which may

491   result in wide variations in the estimated correlation coefficients. Therefore, single species

492   outcomes should be interpreted with great caution.

493   To conclude, this study of broad-scale β-SGDC for multiple freshwater fish species

494   suggests that historical processes greatly contributed to the shape of present-day patterns of

495   taxonomic and genetic differentiations, independently of contemporary processes. This

496   implies that taxonomic differentiation may be an appropriate proxy to explain the role of

497   historical processes on patterns of genetic differentiation, which remains a challenging issue

498   in landscape genetics (Dyer, Nason, & Garrick, 2010). We believe that further studies of

499   broad-scale β-SGDCs comparing organisms whose dispersal is limited within habitat patches

500   with organisms that are able to disperse across habitat patches, will help to investigate the

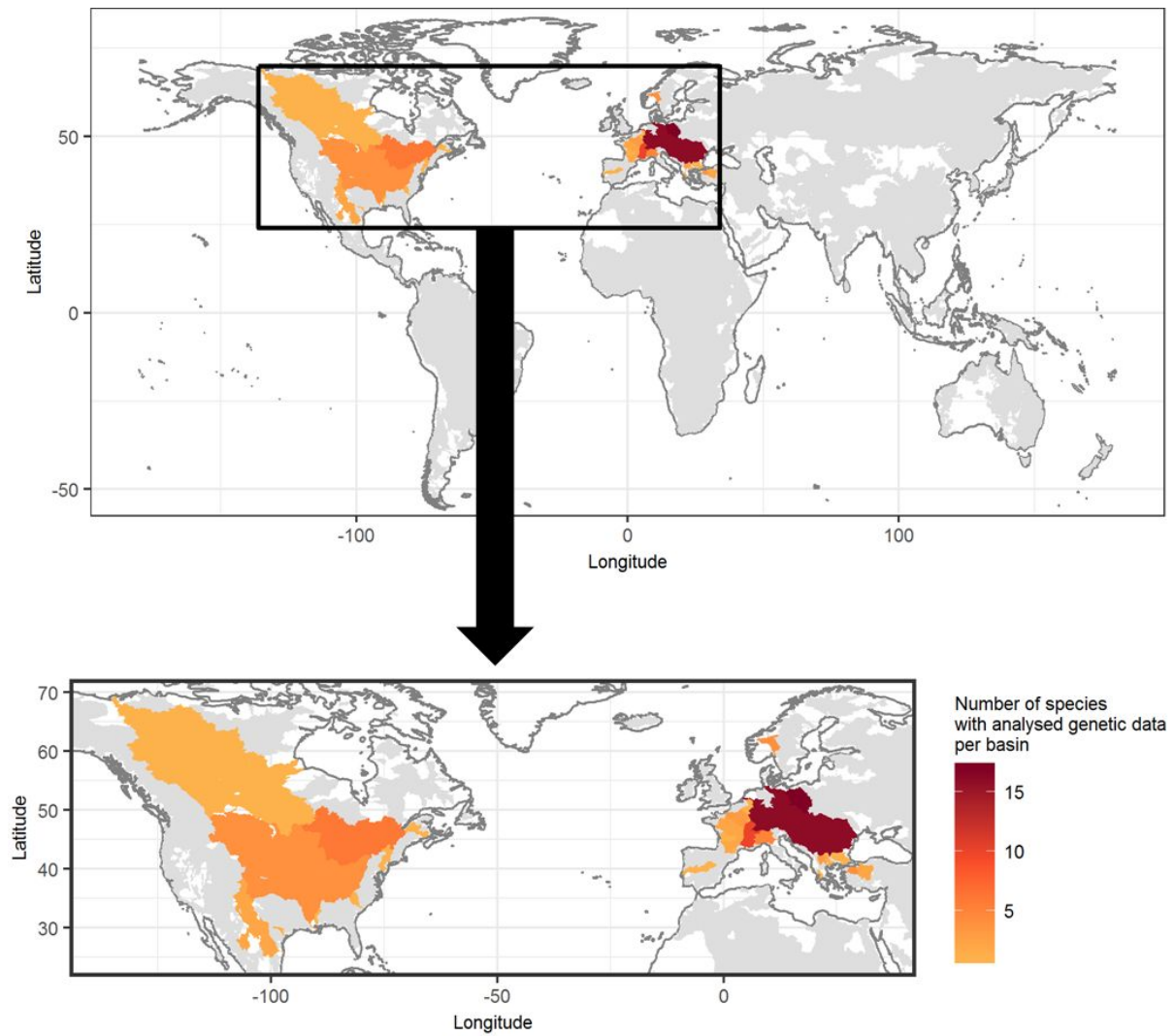501   extent to which our findings can be generalised.

503 **Figures**

504

505 **Figure 1.** Illustration of the different processes expected to act in parallel on between-basin

506 taxonomic differentiation of freshwater fish communities (TD) and on between-basin genetic

507 differentiation of freshwater fish populations (GD) at different periods and the proxies used in

508 this study to infer those processes. The black frame encompasses the causal diagram depicting

509 the relationships that we actually tested in this study, i.e. the effects of geographic distance

510 (proxy of contemporary dispersal limitation), environmental distance (proxy of contemporary

511 environmental selection) and pairwise harmonic mean area between basins (proxy of

512 contemporary drift) on TD and GD. The double arrow between TD and GD represent the

513 residual correlation, i.e. the correlation between TD and GD which is not explained by the

514 parallel influence of contemporary dispersal limitation, contemporary environmental selection

515   and contemporary drift on TD and GD. The numbers on the arrows and the signs below them

516   correspond to our theoretical expectations regarding the different effects which are detailed in

517   the text below the black frame. The grey arrows represent the putative effects of historical

518   dispersal limitation, historical environmental selection and historical drift on TD and GD that

519   we did not directly test in our statistical approach. The grey polygons on the right side

520   represent the evolution of the basin configuration through time from one unique basin to two

521   distinct and disconnected basins, with almost no contemporary dispersal of individuals

522   between them (dispersal may occasionally occur between adjacent basins).

524

Figure 2. Spatial distribution of the number of species with analysed genetic data per basin

studied in our analyses at the global scale. Top: the rectangle indicates the location of the

enlarged area. Bottom: enlarged area including Europe and North America. In total, our study

encompasses 22 species with analysed genetic data in 38 basins.

530

Figure 3. Violin plots showing the distribution and the probability density of (a) raw

correlations coefficients and (b) residual correlation coefficients between taxonomic

differentiation and genetic differentiation for the 22 fish species examined at global scale.

Crossbars indicate mean ± standard deviation.

536

537   **Figure 4.** Violin plots showing the distribution and the probability density of the effect of

538   geographic distance, environmental distance and harmonic mean area (as standardised

539   regression coefficients) for the 22 fish species examined at global scale on (a) taxonomic

540   differentiation and (b) genetic differentiation. Crossbars indicate mean ± standard deviation.

Figure 5. Causal diagram depicting the results of multiple regressions of geographic distance, environmental distance and harmonic mean area on TD and GD (taken separately). Arrows represent putative causal effects through the standardised regression coefficients (positive values: solid arrows; negative values: dotted arrows; arrow width is proportional to coefficient value). The R^2 represent the part of variance explained by the multiple regressions on TD (left) and GD (right). U_{TD} and U_{GD} represent unspecified factors influencing TD and GD respectively. The double arrow between SD and GD corresponds to the residual correlation. All numerical values represent means for the 22 fish species examined at global scale.

551

Figure 6. Observed (dots) and predicted (dashed line) raw values (a, c) and residual values from the multiple regressions (b, d) of genetic differentiation as a function of taxonomic differentiation for the two <u>fish</u> species (a, b) *Rutilus rutilus* and (c, d) *Squalius cephalus*.

556 **References**

557 Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity.

558 *Zoobiology*, 5(2), 181–190.

559 Almany, G. R. (2003). Priority effects in coral reef fish communities. *Ecology*, 84(7), 1920–

560 1935.

561 April, J., Hanner, R. H., Mayden, R. L., & Bernatchez, L. (2013). Metabolic rate and climatic

562 fluctuations shape continental wide pattern of genetic divergence and biodiversity in

563 fishes. *PLoS ONE*, 8(7), e70296.

564 Arnold, J.B. (2018). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package

565 version 4.0.1. Retrieved from https://CRAN.R-project.org/package=ggthemes.

566 Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012).

567 Distance decay of similarity in freshwater communities: do macro-and microorganisms

568 follow the same rules?. *Global Ecology and Biogeography*, 21(3), 365–375.

569 Baselga, A. (2012). The relationship between species replacement, dissimilarity derived from

570 nestedness, and nestedness. *Global Ecology and Biogeography*, 21(12), 1223–1232.

571 Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., &

572 Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal

573 continuum of biodiversity at species and genetic levels. *Nature Communications*,

574 4(1892).

575 Baselga, A., Gómez-Rodríguez, C., & Vogler, A. P. (2015). Multi-hierarchical macroecology

576 at species and genetic levels to discern neutral and non-neutral processes. *Global

577 Ecology and Biogeography*, 24(8), 873–882.

578   Baselga, A., Orme, D., Villeger, S., De Bortoli, J., & Leprieur, F. (2018). betapart:

579   Partitioning Beta Diversity into Turnover and Nestedness Components. R package

580   version 1.5.0. Retrieved from https://CRAN.R-project.org/package=betapart.

581   Bivand, R., & Lewin-Koh, N. (2018). maptools: Tools for Reading and Handling Spatial

582   Objects. R package version 0.9-3. Retrieved from https://CRAN.R-

583   project.org/package=maptools.

584   Bivand, R., Keitt, T., & Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data

585   Abstraction Library. R package version 1.3-4. Retrieved from https://CRAN.R-

586   project.org/package=rgdal.

587   Bivand, R., & Rundel, C. (2018). rgeos: Interface to Geometry Engine - Open Source

588   ('GEOS'). R package version 0.3-28. Retrieved from https://CRAN.R-

589   project.org/package=rgeos.

590   Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to go bigger: emerging patterns in

591   macrogenetics. *Trends in Genetics*, 33(9), 579–580.

592   Carvalho, G. R. (1993). Evolutionary aspects of fish distribution: genetic variability and

593   adaptation. *Journal of Fish Biology*, 43(Supplement A), 53–73.

594   Derry, A. M., Arnott, S. E., Shead, J. A., Hebert, P. D., & Boag, P. T. (2009). Ecological

595   linkages between community and genetic diversity in zooplankton among boreal shield

596   lakes. *Ecology*, 90(8), 2275–2286.

597   Dias, M. S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J. F., … Tedesco,

598   P. A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity.

599   *Ecology Letters*, 17(9), 1130–1140.

600   Dray, S., & Dufour, A. (2007). The ade4 package: Implementing the Duality Diagram for

601   Ecologists. *Journal of Statistical Software*, 22(4), 1–20.

602   Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J.

603   (2007). Can population genetic structure be predicted from life-history traits?. *The*

604   *American Naturalist*, 169(5), 662–672.

605   Durand, J. D., Persat, H., & Bouvet, Y. (1999). Phylogeography and postglacial dispersion of

606   the chub (*Leuciscus cephalus*) in Europe. *Molecular Ecology*, 8(6), 989–997.

607   Dyer, R. J., Nason, J. D., & Garrick, R. C. (2010). Landscape modelling of gene flow:

608   improved power using conditional genetic distance derived from the topology of

609   population networks. *Molecular Ecology*, 19(17), 3746–3759.

610   Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J. G., & Blanchet, S. (2016). Lessons from the

611   fish: a multi-species analysis reveals common processes underlying similar species-

612   genetic diversity correlations. *Freshwater Biology*, 61(11), 1830–1845.

613   Fraser, C. I., Davies, I. D., Bryant, D., Waters, J. M. (2018). How disturbance and dispersal

614   influence intraspecific structure. *Journal of Ecology*, 106,1298–1306.

615   Froese, R., & Pauly, D. (2017). FishBase. World Wide Web electronic publication.

616   www.fishbase.org, version (06/2017).

617   Gilbert, B., & Levine, J. M. (2017). Ecological drift and the distribution of species diversity.

618   *Proceedings of the Royal Society B: Biological Sciences,* ~~*R. Soc. B,*~~ 284(1855),

619   20170507.

620   Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics.

621   R package version 0.04-22. Retrieved from https://CRAN.R-

622   project.org/package=hierfstat

623   Hahsler, M., Buchta, C., & Hornik, K. (2018). seriation: Infrastructure for Ordering Objects

624   Using Seriation. R package version 1.2-3.Retrieved from https://CRAN.R-

625   project.org/package=seriation.

626   Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8),

627   1633–1638.

628   Hugueny, B. (1989). West African rivers as biogeographic islands: species richness of fish

629   communities. *Oecologia*, 79, 236–243.

630   Kahilainen, A., Puurtinen, M., & Kotiaho, J. S. (2014). Conservation implications of species-

631   genetic diversity correlations. *Global Ecology and Conservation*, 2, 315–323.

632   Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern

633   Pacific rocky intertidal community. *PLoS ONE*~~PLoS One~~, 5(1), e8594.

634   Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species-

635   genetic diversity correlations to the understanding of community assembly rules. *Oikos*,

636   126(6), 759–771.

637   Larmuseau, M. H. D., Freyhof, J., Volckaert, F. A. M., & Van Houdt, J. K. J. (2009).

638   Matrilinear phylogeography and demographical patterns of *Rutilus rutilus*: implications

639   for taxonomy and conservation. *Journal of Fish Biology*, 75(2), 332–353.

640   Laroche, F., Jarne, P., Lamy, T., David, P., & Massol, F. (2014). A neutral theory for

641   interpreting correlations between species and genetic diversity in communities. *The*

642   *American Naturalist*, 185(1), 59–69.

643   Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior:

644   a permutational regression approach. *Evolution*, 48(5), 1487–1499.

645   Leprieur, F., Olden, J. D., Lek, S., & Brosse, S. (2009). Contrasting patterns and mechanisms

646   of spatial turnover for native and exotic freshwater fish in Europe. *Journal of*

647   *Biogeography*, 36(10), 1899–1912.

648   Leprieur, F., Tedesco, P. A., Hugueny, B., Beauchard, O., Dürr, H. H., Brosse, S., &

649   Oberdorff, T. (2011). Partitioning global patterns of freshwater fish beta diversity reveals

650   contrasting signatures of past climate changes. *Ecology Letters*, 14(4), 325–334.

651   Lichstein, J. W. (2007). Multiple regression on distance matrices: a multivariate spatial

652   analysis tool. *Plant Ecology*, 188(2), 117–131.

653   Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

654   Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid

655   them. *Molecular Ecology*, 24, 3223–3231.

656   Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F_{ST} and related

657   measures. *Molecular Ecology Resources*, 11(1), 5–18.

658   Moritz, C., Meynard, C. N., Devictor, V., Guizien, K., Labrune, C., Guarini, J. M., &

659   Mouquet, N. (2013). Disentangling the role of connectivity, environmental filtering, and

660   spatial structure on metacommunity dynamics. *Oikos*, 122(10), 1401–1410.

661   Oberdorff, T., Tedesco, P. A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr,

662   H. H. (2011). Global and regional patterns in riverine fish species richness: a review.

663   *International Journal of Ecology*, 201(967631).

664   Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ...Wagner, H.

665   (2018). vegan: Community Ecology Package. R package version 2.5-2. Retrieved from

666   https://CRAN.R-project.org/package=vegan.

667   Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of

668   population genetic differentiation in the wild: isolation by dispersal limitation, isolation

669   by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999.

670   Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the*

671   *National Academy of Sciences*, 70(12), 3321–3323.

672   Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and

673   ecology. *Journal of Biogeography*, 26(4), 867–878.

674   Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.

675   Retrieved from https://CRAN.R-project.org/package=RColorBrewer.

676   Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2018). Biostrings: Efficient

677   manipulation of biological strings. R package version 2.48.0.

678   Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis,

679   A., & Vogler, A. P. (2011). Testing the species–genetic diversity correlation in the

680   Aegean Archipelago: toward a haplotype-based macroecology? *The American*

681   *Naturalist*, 178(2), 241–255.

682   Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution

683   in R language. *Bioinformatics*, 20, 289–290.

684   Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part II. *Ecology*,

685   43(3), 410–432.

686   Prunier, J. G., Dubut, V., Chikhi, L., & Blanchet, S. (2017). Contribution of spatial

687   heterogeneity in effective population sizes to the variance in pairwise measures of

688   genetic differentiation. *Methods in Ecology and Evolution*, 8(12), 1866–1877.

689   Qian, H., & Ricklefs, R. E. (2012). Disentangling the effects of geographic distance and

690   environmental dissimilarity on global patterns of species turnover. *Global ecology and*

691   *Biogeography*, 21(3), 341–351.

692   R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna,

693   Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-

694   project.org/.

695   Radinger, J., & Wolter, C. (2014). Patterns and predictors of fish dispersal in rivers. *Fish and*

696   *fisheries*, 15(3), 456–473.

697   Robinson, D., & Hayes, A. (2018). broom: Convert Statistical Analysis Objects into Tidy

698   Tibbles. R package version 0.5.0. Retrieved from https://CRAN.R-

699   project.org/package=broom.

700   Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge, UK: Cambridge

701   University Press.

702   Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under

703   isolation by distance. *Genetics*, 145(4), 1219-–1228.

704   Seifertová, M., Bryja, J., Vyskočilová, M., Martínková, N., & Šimková, A. (2012). Multiple

705   Pleistocene refugia and post-glacial colonization in the European chub (*Squalius*

706   *cephalus*) revealed by combined use of nuclear and mitochondrial markers. *Journal of*

707   *Biogeography*, 39(6), 1024–1040.

708   Sepkoski, J. J., & Rex, M. A. (1974). Distribution of freshwater mussels: coastal rivers as

709   biogeographic islands. *Systematic Zoology*, 23(2), 165–188.

710   Serrouya, R., Paetkau, D., McLellan, B. N., Boutin, S., Campbell, M., & Jenkins, D. A.

711   (2012). Population size and major valleys explain microsatellite variation better than

712   taxonomic units for caribou in western Canada. *Molecular Ecology*, 21(11), 2588–2601.

713   Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in

714   ecological communities. *Ecography*, 30(1), 3–12.

715   Sørensen, T. J. (1948). A method of establishing groups of equal amplitude in plant sociology

716   based on similarity of species content, and its application to analyses of the vegetation on

717   Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, 5, 1–

718   34.

719   Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., … Oberdorff,

720   T. (2017). A global database on freshwater fish species occurrence in drainage basins.

721   *Scientific Data*, 4(170141).

722   Tedesco, P. A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H. H., Beauchard, O., …

723   Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global

724   Ecology and Biogeography*, 21(10), 977–987.

725   Tisseuil, C., Cornu, J. F., Beauchard, O., Brosse, S., Darwall, W., Holland, R., … Oberdorff,

726   T. (2013). Global diversity patterns and cross-taxa convergence in freshwater systems.

727   *Journal of Animal Ecology*, 82(2), 365–376.

728   Vellend, M. (2003). Island biogeography of genes and species. *The American Naturalist*,

729   162(3), 358–365.

730   Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic

731   diversity. *Ecology Letters*, 8(7), 767–781.

732　Vellend, M., & Orrock, J. L. (2009). Ecological and genetic models of diversity. *The theory of*

733　　　*Island biogeography revisited*. Princeton Univ. Press, Princeton, NJ, 439–462.

734　Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., & Garant, D. (2014).

735　　　Drawing ecological inferences from coincident patterns of population- and community-

736　　　level biodiversity. *Molecular Ecology*, 23(12), 2890–2901.

737　Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and

738　　　geography in spatial genetic divergence. *Ecology Letters*, 16(2), 175–182.

739　Whitlock, M. C. (2005). Combining probability from independent tests: the weighted Z-

740　　　method is superior to Fisher's approach. *Journal of Evolutionary Biology*, 18(5), 1368–

741　　　1373.

742　Whittaker, R. H. (1960). Vegetation of the Siskiyou mountains, Oregon and California.

743　　　*Ecological monographs*, 30(3), 279–338.

744　Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 213-251.

745　Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical*

746　　　*Software*, 21(12), 1-20.

747　Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New-York, NY: Springer-

748　　　Verlag.

749　Winter, D. J. (2012). mmod: an R library for the calculation of population differentiation

750　　　statistics. *Molecular Ecology Resources*, 12(6), 1158-1160.

751　Wright, S. (1943). Isolation by distance. *Genetics*, 28(March), 114–138.

752　Wright, S. (1951). The genetical structure of natural populations. *Annals of Eugenics*, 15(4),

753　　　323–354.

755 **Biosketch**

756 The authors are particularly interested in documenting and explaining spatial patterns of

757 biodiversity at large spatial scales. They try to disentangle the interplay between historical and

758 contemporary factors (including anthropogenic pressures) in shaping patterns of alpha and

759 beta diversities in aquatic populations and communities and value this knowledge to guide

760 conservation strategies of aquatic ecosystems.

761 **Supporting information**

762 Supporting information for this article may be found online.

763 **Data availability**

764 The data and the script~~Upon acceptance of this manuscript, all data and scripts~~ used for the

765 analyses are available from the Dryad Digital Repository:

766 https://doi.org/10.5061/dryad.8vb6160.~~will be uploaded in a Dryad repository.~~

1   **Correlations between broad-scale taxonomic and genetic differentiations suggest a**

2   **dominant imprint of historical processes on beta diversities**

3   **Running title:** Broad-scale β-SGDCs in freshwater fishes

4   Marine Robuchon [1, 2 *], Boris Leroy [1], Céline Jézéquel [1, 3] and Bernard Hugueny [1, 3]

5   [1] Unité Biologie des organismes et écosystèmes aquatiques (BOREA), Muséum national

6   d'Histoire naturelle, Sorbonne Université, Université de Caen Normandie, Université des

7   Antilles, CNRS, IRD, CP 26, 57 rue Cuvier 75005 Paris, France

8   [2] Centre d'écologie et des sciences de la conservation (CESCO), Muséum national d'Histoire

9   naturelle, CNRS, Sorbonne Université, CP 135, 57 rue Cuvier 75005 Paris, France

10 [3] Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse

11 Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bat 4R1, 31062 Toulouse cedex 9,

12 France

13 * corresponding author: robuchon@mnhn.fr

14 **Acknowledgements**

untypeset proof

Abstract

23

24 – **Aim:** Dispersal limitation, environmental selection and drift are known to influence both

25 taxonomic similarity between communities and genetic similarity between populations.

26 However, disentangling the relative roles of these processes on spatial patterns of

27 differentiation - whether regarding taxonomic differentiation between communities or

28 genetic differentiation between populations - is challenging. Investigating whether spatial

29 patterns of taxonomic differentiation and genetic differentiation are correlated (β-SGDCs)

30 is a promising approach to address this issue. Here, we investigated β-SGDCs over broad

31 spatial scales and 22 freshwater fish species to elucidate the processes shaping taxonomic

32 and genetic differentiations between drainage basins.

33 – **Location:** Global scope, data mainly from Europe and North America.

34 – **Taxon:** Actinopterygii and Petromyzontiformes (freshwater fishes).

35 – **Methods:** We used Mantel tests to investigate the raw correlation between taxonomic and

36 genetic differentiations. We carried out multiple regressions to characterise the effects of

37 geographic distance (proxy of dispersal limitation), environmental distance (proxy of

38 environmental selection) and pairwise harmonic mean area between basins (proxy of drift)

39 on taxonomic differentiation and genetic differentiation taken separately. We then

40 analysed the correlation between the residuals of these two regressions with Mantel tests.

41 Finally, we investigated whether the variation in β-SGDC between species was related to

42 species traits.

43 – **Results**: We detected an overall positive β-SGDC, both for the raw and the residual

44 correlations. This implies that, unmeasured effects, other than geographic distance,

45 environmental distance and harmonic mean area, influence the β-SGDC observed. Values

46 of β-SGDCs greatly varied between species, but this was not explained by any species

47 traits.

48   –   **Main conclusions**: Independently of the effects of geographic distance, environmental

49   distance and harmonic mean area, the β-SGDC we observed suggests that historical

50   processes strongly shaped the patterns of taxonomic and genetic differentiations between

51   basins. Consequently, taxonomic differentiation may be an appropriate proxy to explain

52   the influence of historical processes on genetic differentiation.

53   **Keywords**

54   β-SGDC, dispersal limitation, distance decay, drift, environmental selection, freshwater

55   fishes, historical processes, macrogenetics, Mantel tests, multiple regression of distance

56   matrices

Introduction

59   Understanding the mechanisms that shape taxonomic differentiation between spatially

60   separated communities and those influencing genetic differentiation between spatially

61   separated populations are long-standing issues in ecology and evolution (Wright, 1943, 1951;

62   Whittaker, 1960, 1972; Preston, 1962; Mayr, 1963). More recently, these issues have been

63   united (Vellend, 2003; Vellend & Geber, 2005; Kahilainen, Puurtinen, & Kotiaho, 2014),

64   emphasising that drift, dispersal and selection influence both taxonomic differentiation

65   between communities and genetic differentiation between populations. Drift and selection

66   affect the species frequencies within communities and gene frequencies within populations,

67   which tend to increase differentiation. Conversely, dispersal of individuals between

68   populations and dispersal of species between communities will tend to decrease

69   differentiation.

70   Among the factors that can affect both taxonomic and genetic differentiation,

71   geographic distance is probably the most studied. The similarity in taxonomic composition

72   between communities tends to decrease with the geographic distance that separates them; a

73   pattern long-recognised in biogeography and formalised as the distance decay of similarity

74   (Nekola & White, 1999; Soininen, McDonald, & Hillebrand, 2007). Similarly, the genetic

75   similarity between populations tends to decrease with the geographic distance separating

76   them; a pattern long-recognised in population genetics and referred to as isolation-by-distance

77   (IBD; Wright, 1943; Rousset, 1997). Two main non-mutually exclusive explanations can

78   account for the distance decay of taxonomic and genetic similarity (see Soininen et al., 2007

79   and Orsini et al., 2013 for reviews). First, similarity decreases with geographic distance

80   because the dispersal of organisms (i.e. their movement from a place to another) is limited by

81   their intrinsic dispersal ability, which hence does not counteract anymore the differentiating

82   effect of drift. Secondly, similarity decays with geographic distance because of decreasing

83   similarity in spatially correlated environmental features; in such cases, the underlying

84   explanation is environmental selection of different species (community differentiation) or

85   genotypes (genetic differentiation) in distinct environments. However, differentiation between

86   sites does not always increase with geographic distance. This is the case when the migrants

87   are competitively excluded by already established communities (e.g. Almany, 2003) or

88   populations (e.g. Fraser et al., 2018) at high densities, therefore counteracting the

89   homogenising effect of dispersal whatever the intrinsic dispersal ability of the migrants. This

90   may also happen when the environmental variation between sites displays no – or negative –

91   spatial autocorrelation (e.g. Derry et al., 2009).

92   Although spatial patterns of taxonomic and genetic differentiations can be explained

93   by environmental selection independently of limited dispersal and vice-versa, disentangling

94   the relative strengths of these mechanisms is challenging, as communities and populations

95   that are geographically distant are also likely to inhabit different environments. Using modern

96   spatial statistical methods, some recent studies have nonetheless taken up the challenge. In

97   their study of global patterns of species turnover in terrestrial vertebrates, Qian & Ricklefs

98   (2012) have shown that both dispersal limitation and environmental selection have played

99   important roles in determining the patterns they observed. Similarly, in their reinterpretation

100   of 34 representative studies, Orsini *et al*. (2013) highlighted that patterns associated with

101   dispersal limitation were as common as those associated with local genetic adaptation in

102   structuring population genetic differentiation in the wild. However, the relative strength of

103   dispersal limitation and environmental selection on differentiation patterns seems to vary

104   according to the species and the spatial scale under consideration, both for taxonomic

105   (Astorga et al., 2012; Moritz et al., 2013) and genetic (Orsini et al., 2013) differentiations.

106   Comparative studies of multiple taxonomic groups suggest that the influence of dispersal

107   limitation on taxonomic differentiation is stronger for species with low dispersal ability than

108   for more effective dispersers, and this result holds true both at the broad-scale (e.g. Qian &

109   Ricklefs, 2012) and at the fine-scale (e.g. Astorga et al., 2012). Comparative studies of

110   multiple species have also been identified as a promising avenue to reveal the factors

111   influencing patterns of genetic differentiation (Wang et al., 2013). Yet, such studies are scarce

112   and mainly carried out at relatively small spatial scales (e.g. Wang et al., 2013; Fourtune, Paz-

113   Vinas, Loot, Prunier, & Blanchet, 2016) despite the increasing availability of broad-scale

114   datasets in population genetics, suggesting that it is now time to embrace macrogenetics

115   (Blanchet, Prunier, & De Kort, 2017).

116   Theories in population genetics and community ecology acknowledge that dispersal,

117   selection and drift together shape the differentiation between populations and between

118   communities (see e.g. Vellend & Orrock, 2009 for a review of processes in both disciplines).

119   However, the contribution of drift to taxonomic or genetic differentiation has been the subject

120   of few empirical investigations, in contrast with dispersal limitation and environmental

121   selection (Gilbert & Levine, 2017; Prunier, Dubut, Chikhi, & Blanchet, 2017). Genetic drift is

122   the evolutionary process of random fluctuations in allelic frequencies occurring naturally in

123   all populations due to their finite size, although it is stronger in small ones (Allendorf, 1986).

124   Similarly, ecological drift corresponds to the random fluctuations in species frequencies

125   occurring naturally in all communities due to their finite size and this too is stronger in small

126   ones. Drift is the result of random sampling during the processes of birth, death and

127   reproduction and ultimately leads to the loss of genetic diversity within populations, and

128   species diversity within communities. Therefore, drift increases the differentiation between

129   populations and between communities depending on their respective sizes: the smaller they

130   are, the more they will differentiate. Dispersal of organisms between populations and between

131   communities counteracts the differentiating effect of drift while environmental selection can

132   accelerate it. Attributing the observed spatial patterns of genetic or taxonomic differentiation

133   to dispersal limitation or to environmental selection without accounting for drift may thus be

134   misleading.

135   Most of the explanatory variables used in analyses to infer the role of dispersal

136   limitation and environmental selection on differentiation patterns are proxies representing

137   present-day conditions. Thus, using current explanatory variables to investigate dispersal

138   limitation and environmental selection on differentiation patterns is only relevant if the

139   contemporary geography and environment of the study area has not changed substantially

140   since these patterns were established (Wang et al., 2013). Otherwise, the historical processes

141   that may have shaped the observed differentiation patterns, including those explaining

142   colonisation histories, may be overlooked. In addition, if dispersal is not continuous (as

143   generally assumed) but rather intermittent because of landscape dynamics which have induced

144   the appearance and disappearance of dispersal barriers, then two geographically close

145   localities sharing a similar environment could, nevertheless, be dissimilar in terms of

146   community composition and allele frequencies provided they have been isolated for a

147   sufficiently long time (given their sizes) for drift to be effective. Nonetheless, examining

148   which part of differentiation patterns is not explained by current explanatory variables may

149   help to elucidate the influence of historical processes on those patterns.

150   In such complex contexts, understanding the rules that govern differentiation patterns

151   (both taxonomic and genetic) may be enhanced by comparing taxonomic dissimilarities

152   between communities and genetic dissimilarities between populations of a focal species

153   (Lamy, Laroche, David, Massol, & Jarne, 2017). The few empirical studies that have

154   examined these correlations, called beta species-genetic diversity correlations (β-SGDCs)

155   (Kahilainen et al., 2014), revealed a majority of positive correlations (reviewed by Lamy et

156   al., 2017). This suggests that the processes affecting taxonomic dissimilarity affect genetic

157   dissimilarity of the species under study in the same way (Baselga et al., 2013; Baselga,

158   Gómez-Rodríguez, & Vogler, 2015). However, because the relative strength of processes

159   shaping patterns of taxonomic and genetic differentiation is expected to vary according to the

160   species and the spatial scale under consideration, so is the strength of the putative resulting β-

161   SGDCs. Among the rare studies of β-SGDCs over multiple species, some have highlighted

162   that positive β-SGDCs were stronger for focal species with low dispersive abilities (e.g.

163   Papadoulou et al., 2011) while others did not find strong differences between species (e.g.

164   Fourtune et al., 2016), and all were carried out at relatively small spatial scales. These

165   contrasting results call for more studies of β-SGDCs for multiple species to gain a better

166   understanding of whether and how variations in species traits lead to distinct patterns of β-

167   SGDCs. In addition, β-SGDCs for multiple species have not been examined yet at broad

168   spatial scales.

169   Freshwater fishes are a relevant model to start investigating this question for at least

170   three reasons. First, as β-SGDCs for multiple freshwater fish species have been conducted at

171   small scales within a single drainage basin (Fourtune et al., 2016), computing β-SGDCs for

172   multiple freshwater fish species at broader scales will permit the comparison of β-SGDC

173   variations across scales for the same biological model. Secondly, for strictly freshwater fishes,

174   at present, almost no exchange of individuals occurs between drainage basins (hereafter

175   "basins") because they are isolated from each other by land and/or sea (although dispersal

176   may occasionally occur between adjacent basins). Therefore, if basins are isolated from each

177   other independently of the geographic distance between them, we can expect dispersal

178   limitation to play a minor role on the broad-scale between-basin patterns of taxonomic and

179   genetic differentiations. On the contrary, if basins have been isolated from each other for a

180   long time, we can expect drift to play a significant role on the broad-scale between-basin

181   patterns of taxonomic and genetic differentiations, as dispersal no longer counteracts the

182   effect of drift. This is particularly true if population sizes and community sizes are small: the

183   larger they are, the longer it will take to observe differentiation. Finally, while several recent

184   studies have tried to disentangle the role of dispersal limitation, environmental selection and

185   historical processes on patterns of taxonomic differentiation between basins (Leprieur et al.,

186   2009, 2011; Dias et al., 2014), the processes shaping genetic differentiation at such broad

187   scales have rarely been examined together. Studying β-SGDCs may therefore reveal whether

188   the processes influencing between-basin taxonomic differentiation, affect between-basin

189   genetic differentiation in the same way.

190   In this meta-analysis of 22 freshwater fish species, our main goal was to elucidate the

191   processes underlying β-SGDCs at broad spatial scales. In particular, we wanted to test

192   whether β-SGDCs can be explained by the parallel action of measurable processes (dispersal

193   limitation, environmental selection and drift) on taxonomic and genetic differentiation, or are

194   due to other, non-measurable processes. To that end, we first computed measures of genetic

195   differentiation between basins for each fish species. Then, for each species, we investigated i)

196   the raw correlation between taxonomic and genetic differentiations, ii) the effects of

197   geographic distance (as a proxy of dispersal limitation), environmental distance (as a proxy of

198   environmental selection) and pairwise harmonic mean area between basins (as a proxy of

199   drift) on taxonomic differentiation and genetic differentiation taken separately and iii) the

200   correlation between taxonomic and genetic differentiations independent of the effects of

201   geographic distance, environmental distance and harmonic mean area (Fig. 1). Such

202   independent correlation integrates the effects of unmeasured processes, including the

203   historical processes of dispersal, environmental selection and drift related to the historical

204   connectivity of basins. This allowed us to characterise the overall (i.e. mean over the species)

205   raw β-SGDC, the overall effects associated with each predictor for each response variable,

206   and the overall independent β-SGDC. Finally, we tested whether the variation in β-SGDCs

207   between species could be attributed to variations in species traits.

208

209   **Materials and methods**

210   *Genetic differentiation and geographic distribution of genetic data*

211   As we wanted to investigate genetic differentiation at the broad scale, we restricted our

212   analyses of spatial genetic differentiation to species exhibiting a wide distribution. We

213   downloaded sequence records of the mitochondrial gene cytochrome c oxidase I (COI) from

214   BOLD (www.boldsystems.org) in May 2017 for the following widespread taxa: Cyprinidae,

215   Salmonidae, *Lampetra*, *Petromyzon*, *Cobitis*, *Misgurnus*, *Barbatula*, *Silurus*, *Esox*, *Lota lota*,

216   *Gymnocephalus* and *Sander*. We first cleaned and aligned sequences by taxon and then re-

217   aligned them together to keep the same fragment of 501 nucleotides for the analyses.

218   Sequence records without sufficient geographic information to assign them to a drainage

219   basin (hereafter "basin") or without a species name were discarded from our dataset. We

220   defined a basin as the drainage area upstream from its mouth at the sea. According to this

221   definition, a basin is isolated from other basins by barriers (sea or land) which are impassable

222   for strictly freshwater fishes. The validity of species names was assessed using FishBase

223   (Froese & Pauly, 2017), resulting in a total of 6637 sequence records with a valid species

224   names belonging to 716 distinct species distributed in 218 basins. We checked the

225   environment (i.e. "freshwater", "brackish", "saltwater") of each species using FishBase

226   (Froese & Pauly, 2017), the status (i.e. "native", "exotic" or "unknown") of each species in

227   each basin using the global database on freshwater fish species occurrence (Tedesco et al.,

228   2017). We conserved only native records of strictly freshwater species, i.e. 4894 records,

229   corresponding to a total of 509 species in 176 basins.

230   We calculated between-basin genetic differentiation (by pooling individuals in the

231   drainage basins) for the polymorphic species present in at least four basins with a minimum of

232   three individuals per basin. To calculate such genetic differentiation, we converted single

233   nucleotide polymorphisms (SNPs) of the sequence data into a table of individual genotypes

234   and then calculated pairwise genetic distances for each pair of basins using G''_{ST} (Meirmans

235   & Hedrick, 2011), the corrected version of Hedrick's G'_{ST} (Hedrick, 2005). G''_{ST} is

236   independent from within population diversity and does not underestimate genetic distance

237   when the number of populations is small (Meirmans & Hedrick, 2011). As G''_{ST} likely

238   removes the effect of drift, we also calculated the pairwise genetic distances for each pair of

239   basins using F_{ST} (Nei, 1973) for comparison purpose.

240   All these different selection filters drastically reduced the number of species we

241   included in analyses: we finally focused our analyses of between-basin genetic differentiation

242   on a total of 22 species in 38 basins (for more details about these basins, see Appendix S1 in

243   Supporting Information). For instance, from all the species of *Lampetra* and *Petromyzon* we

244   searched in BOLD, we conserved only *L. planeri* for the analyses. The number of species

245   with analysed genetic data per basin varies from one (for 20 basins) to seventeen (for one

246   basin, the basin "Odra"; Fig. 2; Appendix S1). The number of basins per species varies from

247   four (our minimum threshold) to eleven (for the species *Esox lucius*) covering on average 76

248   % (standard deviation (sd) = 18 %) of their native range, with species exhibiting on average

249   11.0 individuals per basin (sd = 11.0; Appendix S2). Most records are located in Europe and

250   North-America (Fig. 2).

251   *Taxonomic differentiation*

252   We calculated pairwise taxonomic differentiation between each pair of basins based on the list

253   of native species per basin extracted from the global database on freshwater fish species

254   occurrence (Tedesco et al., 2017). We used Simpson dissimilarity (β_{sim}) - the turnover

255   component of Sørensen dissimilarity (Sørensen, 1948) - which is independent from total

256   taxonomic richness (see Baselga 2012 for review).

257   *Geographic distance, environmental distance and pairwise harmonic mean area*

258   We computed between-basin geographic distance by calculating the distance between basin

259   centroids. The average distance between basins was 5 295 km (sd = 3 920 km). Measurements

260   of geographic distance were then centred and scaled before further analyses.

261   To characterise the environment of the basins, we used variables related to the

262   contemporary climate known to influence freshwater fish alpha diversity patterns at the global

263   scale (e.g. Dias et al. 2014; Oberdorff et al. 2011; Tedesco et al. 2012; Tisseuil et al. 2013) i.e.

264   mean annual temperature and precipitation (http://www.worldclim.org/), mean annual surface

265   runoff (http://www.grdc.sr.unh.edu/) and mean annual actual and potential evapotranspiration

266   (http://csi.cgiar.org/Aridity/). The mean values of all the variables were computed for each

267   basin. To compute pairwise environmental distances between each pair of basins, we

268   performed a principal component analysis (PCA) on these environmental variables

269   (previously centred and scaled) and calculated the Euclidean distance between basins based

270   on the first three PCA axes.

271   Finally, we calculated the pairwise harmonic mean area between each pair of basins.

272   Indeed, the harmonic mean population size between pairs allows the role of drift on genetic

273   differentiation (Serrouya et al., 2012) to be taken into account and can be computed using

274   environmental proxies for local carrying capacities (Prunier et al., 2017). In our case, we used

275   the surface area of the basin as a raw environmental proxy for local carrying capacity. The

276   basins studied had an average surface area of 257 589 km^2 (sd = 584 803 km^2). Measurements

277   of harmonic mean area were centred and scaled before further analyses.

278   *Statistical analyses*

279   First, we investigated the raw correlation between taxonomic differentiation and genetic

280   differentiation for each species individually using a Mantel correlation test with 999

281   permutations. To test the null hypothesis that there is no relationship between taxonomic

282   differentiation and genetic differentiation using the information provided by all the species

283   examined, we considered that this hypothesis has been tested independently *n* times, with *n*

284   being the number of species included in the analyses. To combine the results from several

285   independent tests bearing upon the same overall hypothesis, we used a modification of the

286   Fisher's combined probability test**.** Under the null hypothesis that taxonomic and genetic

287   differentiations are unrelated, observed p-values associated with the Mantel's test of each

288   species are expected to be distributed according to a uniform distribution in [0, 1]. We

289   therefore compared the observed mean p-value over the species to a null distribution of mean

290   p-values obtained by sampling randomly 10 000 times one p-value by species in a uniform

291   distribution in [0, 1]. We calculated the p-value associated with the combined test as the

292   frequency at which null mean p-values were below the observed mean p-value. This

293   procedure avoids some problems identified by the use of a chi-2 test (e.g. Whitlock, 2005)

294   which was the method initially proposed by Fisher.

295   To investigate the putative roles of dispersal limitation, environmental selection and

296   drift on taxonomic and genetic differentiations, for each species we built multiple regressions

297   of geographic distance (D_{geo}), environmental distance (D_{env}) and harmonic mean area (D_{area})

298   on taxonomic differentiation (TD) and genetic differentiation (GD) taken separately (Fig. 1)

299   as follows: $TD = \alpha_0 + \alpha_1*Dgeo + \alpha_2*D_{env} + \alpha_3*D_{area}$ and $GD = \beta_0 + \beta_1*Dgeo + \beta_2*D_{env} +$

300   β_3*D_{area}. Our approach is very similar to multiple regressions on distance matrices (MRM;

301   Lichstein 2007) except in the way we calculated p-values associated with partial regression

302   coefficients. As in MRM, we permuted the rows and associated columns of the response

303   distance matrix simultaneously. We repeated this operation 1999 times while holding the

304   explanatory distance matrices constant to generate null distributions for partial regression

305   coefficients. Then, in MRM, the p-value associated with a partial regression coefficient is

306   calculated with a two-sided permutation test using the pseudo-t of Legendre et al. (1994), i.e.

307   it tests whether the t-statistic associated with each regression coefficient is lower or higher

308   than expected under a null distribution of pseudo-t obtained by permutation. In our approach,

309   we hypothesised that geographic distance and environmental distance have a positive effect

310   on both taxonomic and genetic differentiations (dispersal limitation and environmental

311   selection enhance differentiation and are positively related to geographic and environmental

312   distances respectively) while harmonic mean area has a negative effect on taxonomic and

313   genetic differentiations (drift enhances differentiation and is negatively related to the

314   harmonic mean area). To increase the statistical power of the tests, we therefore carried out

315   one-sided tests for calculating the p-values associated with each partial regression coefficient.

316   We tested the overall (i.e. over the species) effects of geographic distance, environmental

317   distance and harmonic mean area on taxonomic and genetic differentiations by performing a

318   combined probability test as described for the raw correlation between taxonomic and genetic

319   differentiations, but this time considering probabilities associated with partial regression

320   coefficients.

321   To examine whether taxonomic and genetic differentiations were still correlated once

322   the effects of geographic distance and environmental distance had been taken into account, we

323   then performed a Mantel correlation test between the residuals of the multiple regression on

324   taxonomic differentiation and the residuals of the multiple regression on genetic

325   differentiation with 999 permutations for each species. We tested this overall (i.e. over the

326   species) residual correlation using a combined probability test as described for the raw

327   correlation between taxonomic and genetic differentiations.

328   Finally, we examined the heterogeneity between species regarding the residual

329   correlation between taxonomic and genetic differentiations and tested whether some species

330   traits could explain such heterogeneity. The species traits initially considered were body

331   length, longevity in the wild, vulnerability, used in aquaculture, use as bait, use in the

332   aquarium, habitat, migratory behaviour and dispersal ability – which was calculated using a

333   formula giving dispersal distance as a function of body size and caudal fin aspect ratio

334   (Radinger & Wolter, 2014). However, as migratory behaviour was 'potamodromous' for 16

335   out of the 22 species, 'non-migratory' for one species and the information was not available

336   for the 5 species left, we excluded this non-informative trait from our analyses. Species traits

337   were directly extracted from FishBase (Froese & Pauly, 2017). All the analyses and the

338   corresponding figures were realised with R 3.5.1 (R Core Team 2018) using the packages

339   'ade4' (Dray & Dufour, 2007), 'ape' (Paradis, Claude, & Strimmer, 2004), 'betapart'

340   (Baselga, Orme, Villeger, De Bortoli, & Leprieur, 2018), 'Biostrings' (Pagès, Aboyoun,

341   Gentleman, & DebRoy, 2018), 'broom' (Robinson & Hayes, 2018), 'ggplot2' (Wickham,

342   2016), 'ggthemes' (Arnold, 2018), 'hierfstat' (Goudet & Jombart, 2015), 'maptools' (Bivand

343   & Lewin-Koh, 2018), 'mmod' (Winter, 2012), 'RColorBrewer' (Neuwirth, 2014), 'rgdal'

344   (Bivand, Keitt, & Rowlingson, 2018), 'rgeos' (Bivand & Rundel, 2018), 'reshape2'

345   (Wickham, 2007), 'seriation' (Hahsler, Buchta, & Hornik, 2018) and 'vegan' (Oksanen et al.,

346   2018).

347

348   **Results**

349   We detected an overall positive correlation between taxonomic and genetic differentiations,

350   significantly different from 0. This result holds true both for the raw correlation between

351   taxonomic and genetic differentiations ($r_{mean \pm sd}$ = 0.433 ± 0.420; Fig. 3a; Appendices S3, S4)

352 and for the residual correlation, i.e. the correlation between taxonomic and genetic

353 differentiations independent of the effects of geographic distance, environmental distance and

354 harmonic mean area ($r_{\text{mean} \pm \text{sd}}$ = 0.379 ± 0.641; Fig. 3b; Appendices S3, S4).

355 There was almost no overall effect of geographic distance on taxonomic differentiation

356 ($\alpha_{1\ \text{mean} \pm \text{sd}}$ = 0.032 ± 1.051) or on genetic differentiation ($\beta_{1\ \text{mean} \pm \text{sd}}$ = -0.011 ± 1.060), and

357 these overall effects did not differ significantly from 0 (Fig. 4; Appendices S5, S6, S7). In

358 contrast, we detected an overall positive effect of environmental distance on both taxonomic

359 differentiation ($\alpha_{2\ \text{mean} \pm \text{sd}}$ = 0.405 ± 0.956) and genetic differentiation ($\beta_{2\ \text{mean} \pm \text{sd}}$ = 0.229 ±

360 0.865), although it was only significantly different from 0 for taxonomic differentiation (Fig.

361 4; Appendices S5, S6, S7). There was an overall negative effect of harmonic mean area on

362 both taxonomic differentiation ($\alpha_{3\ \text{mean} \pm \text{sd}}$ = -0.178 ± 0.617) and genetic differentiation ($\beta_{3\ \text{mean}}$

363 $_{\pm \text{sd}}$ = -0.092 ± 0.713), but it was only significantly different from 0 for taxonomic

364 differentiation (Fig. 4; Appendices S5, S6, S7). Using F_{ST} as a measure of genetic

365 differentiation did not change the trend we observed: there was an overall negative but not

366 significant effect of harmonic area on F_{ST} (Appendices S6, S7).

367 The results of these multiple regressions as well as the residual correlations between

368 taxonomic and genetic differentiations are summarised in Fig. 5. Overall, they suggest that the

369 effect of geographic distance, environmental distance and harmonic mean area only explain a

370 small part of the correlation between taxonomic and genetic differentiations.

371 Despite this strong overall independent correlation between taxonomic differentiation

372 of freshwater fish communities and genetic differentiation of freshwater fish populations, the

373 correlation between taxonomic and genetic differentiations varied greatly between species

374 (Fig. 3b; Appendix S3). For some species, such as the common roach *Rutilus rutilus* (Fig. 6a,

375 b), there was a sound correlation between taxonomic differentiation and genetic

376   differentiation, whether we considered the raw or residual correlation. In other species, such

377   as the common chub *Squalius cephalus*, taxonomic differentiation and genetic differentiation

378   were not correlated (Fig. 6c, d). In this species, genetic differentiation was better predicted by

379   geographic distance and harmonic mean area (Appendix S5). However, none of the species

380   traits examined explained this heterogeneity of responses between species (Appendix S8).

381

382   **Discussion**

383   Our analyses indicated that taxonomic and genetic differentiations are strongly correlated

384   overall, and independently of the parallel effects of geographic distance, environmental

385   distance and harmonic mean area on taxonomic and genetic differentiations. This suggests

386   that the parallel influence of dispersal limitation, environmental selection and drift on

387   taxonomic and genetic differentiations only partially explains the overall β-SGDC we

388   observed. This independent, unexplained correlation between taxonomic and genetic

389   differentiations thus probably results from processes that we did not explicitly consider in our

390   study. The most probable explanation is that the independent β-SGDC we observed is due to

391   the parallel action of historical processes, which occurred when the basins were still

392   connected, on taxonomic and genetic differentiations. These historical processes include all

393   the different processes that explain the colonisation history of basins (Orsini et al., 2013): i)

394   the dispersal of individuals between previously connected basins in interaction with

395   environmental and/or biotic selection on their dispersal route and ii) environmental and/or

396   biotic selection in interaction with drift within their place of arrival and/or origin that may

397   have led to extinction. This explanation is supported by previous studies on the influence of

398   historical processes on broad-scale taxonomic differentiation between freshwater fish

399   communities on the one hand and on broad-scale genetic differentiation between freshwater

400   fish populations on the other hand. Two recent studies on global freshwater fish biodiversity

401   have shown that present-day patterns of taxonomic dissimilarity across basins are well

402   explained by the historical connectivity of basins (Dias et al., 2014) and also historical

403   climatic oscillations that took place during the Quaternary and the Holocene (Leprieur et al.,

404   2011). For fish species, it is well known that historical dispersal influences genetic structure

405   across basins (Carvalho, 1993), especially for zones that were subject to intense post-glacial

406   recolonisation such as our study area. Multispecies comparisons at large spatial scales in

407   Europe (Seifertová, Bryja, Vyskočilová, Martínková, & Šimková, 2012) and in North

408   America (April, Hanner, Mayden, & Bernatchez, 2013) have shown that the climatic

409   fluctuations that caused glacial cycles during the Pleistocene had a generalised effect on the

410   patterns of fish genetic divergence. In addition, three other, non-exclusive hypotheses may

411   explain the independent β-SGDC we observed. First, the proxies we used to account for

412   dispersal limitation, environmental selection and drift may not fully represent these processes.

413   For instance, we may have overlooked some environmental variables that are important in

414   driving environmental selection. Alternatively, we may have poorly estimated the role of drift,

415   either because the area of the basin we used to account for drift may be too approximate to

416   accurately represent the carrying capacities of populations and communities or because we

417   did not consider that the influence of drift within a basin increases with its age of isolation.

418   Secondly, the sequence data we used to calculate genetic differentiation may not detect

419   dispersal limitation and/or environmental selection properly, either because they did not cover

420   the entire native ranges of species (see possible consequences of such sampling bias in e.g.

421   Meirmans, 2015), or because the mitochondrial gene COI is not directly under environmental

422   selection. Finally, the proposed frameworks to interpret SGDCs (Vellend & Geber, 2005;

423   Lamy et al., 2017) suggest that correlations between species diversity and genetic diversity

424   may also result from a causal action of species diversity on genetic diversity or vice-versa.

425 For instance, if the focal species (i.e. the species for which we studied genetic diversity) is a

426 facilitator for the other component species of the community, we would expect their

427 populations sizes to co-vary positively, resulting in a positive α-SGDC; the converse is

428 expected if the focal species is a competitor. If the nature of the interaction between the focal

429 species and the other species of the community is generally the same within each basin, such

430 causal effect is not expected to influence the β-SGDC we observed. However, if the focal

431 species is facilitator in one basin and competitor in another, we would expect the population

432 sizes of the other component species of the community to increase in the first case and

433 decrease in the second case. This would result in more ecological drift within the community

434 where the focal species is competitor, and therefore more taxonomic differentiation between

435 basins than would be expected without this causal effect. In the end, such causal effect would

436 tend to reduce β-SGDC because it would increase taxonomic differentiation between basins

437 without influencing genetic differentiation.

438 Moreover, our results contribute new insights into the study of β-SGDCs. With the

439 study of 22 new β-SGDCs, we increased the number of β-SGDCs (43) previously reported in

440 the literature by more than 50 % (see Lamy et al., 2017 for the most recent review). We found

441 an average raw correlation between taxonomic and genetic differentiations of 0.433, and an

442 independent correlation (i.e. independent from the parallel effects of geographic distance,

443 environmental distance and harmonic mean area on taxonomic and genetic differentiations) of

444 0.379. Both of these values are superior to the average value of 0.221 reported by Lamy *et al*.

445 (2017). They are also around three times superior to the average value of 0.139 found by

446 Fourtune *et al*. (2016) in their study of four freshwater fish species at a smaller scale, within

447 one drainage basin. This suggests that the strong correlations we found are probably more

448 related to the size and/or the isolation of the habitats we studied (i.e. large drainage basins

449 isolated from each other) rather than our biological model. Previous syntheses have reported

450   that α-SGDCs (i.e. correlations between genetic diversity within populations and species

451   diversity within communities) are stronger and more often positive in island-like habitats

452   (Vellend & Geber, 2005; Vellend et al., 2014), in agreement with the predictions of

453   theoretical models (Laroche et al., 2015). As drainage basins are typically island-like habitats

454   (Sepkoski & Rex, 1974; Hugueny, 1989), our findings suggest that this explanation may hold

455   true for β-SGDCs as well. The observed differences also suggest that for a system without

456   present-day dispersal the β-SGDC is stronger than in systems with present-day dispersal.

457   Actually, the opposite could have been expected (i.e. weaker β-SGDCs for systems without

458   present-day dispersal than for systems with present-day dispersal) if the genetic similarity

459   between two previously connected populations had been erased by genetic drift occurring

460   after their isolation without changes in species composition (i.e. no species extinction).

461   However, our results suggest that drift did not influence significantly the patterns of genetic

462   differentiation we observed. As this result holds true both for G''_{ST} and F_{ST}, this is probably

463   because we mainly focused on large drainage basins and consequently large population sizes.
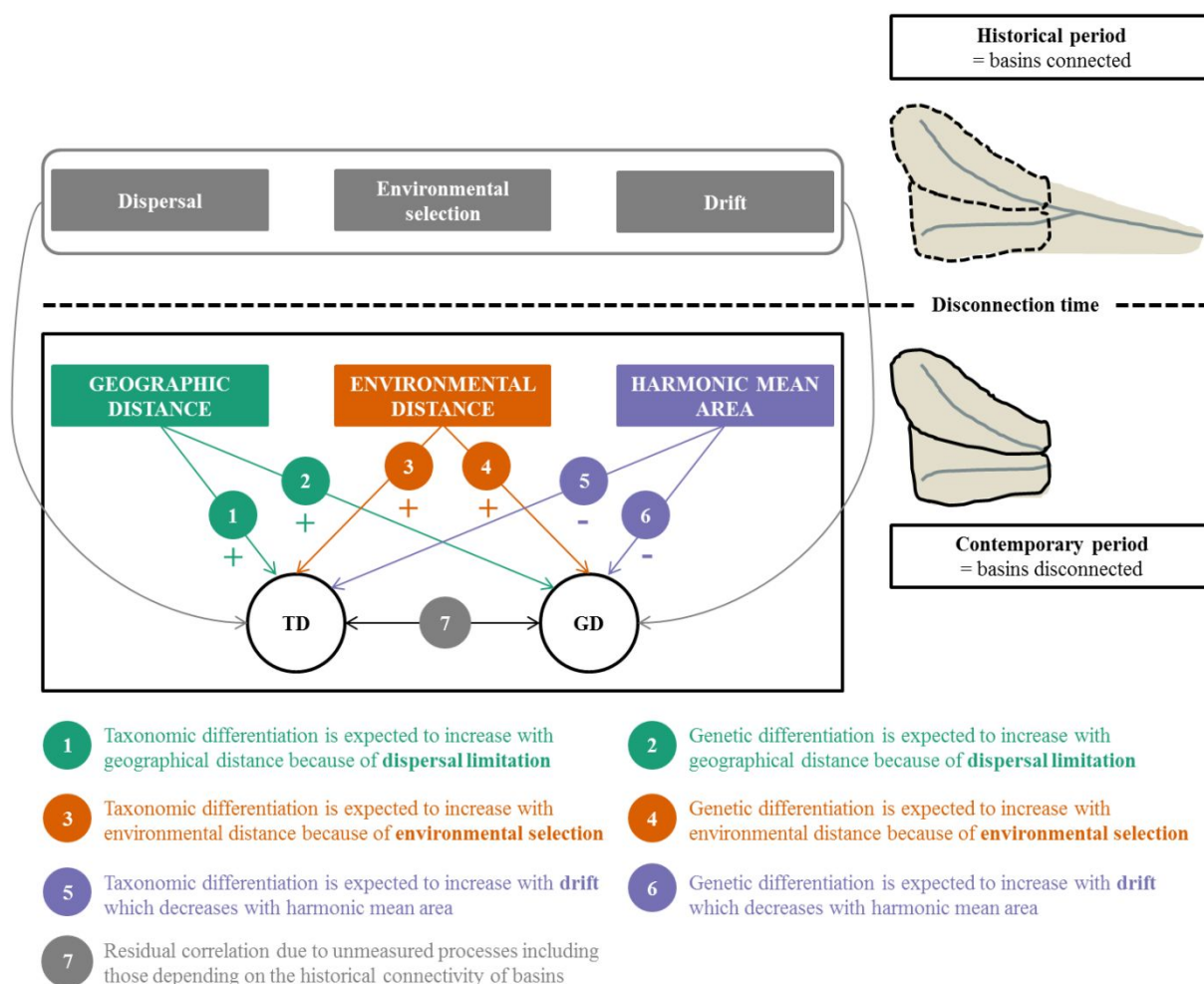
464   Our results also indicate that the sign and magnitude of β-SGDCs varied greatly

465   between the different species we examined. We provided two examples. On the one hand, we

466   showed that the β-SGDC was strong for *Rutilus rutilus*, suggesting that historical processes

467   largely shaped the patterns of genetic differentiation observed in this species. This is

468   consistent with a previous study indicating that *R. rutilus* is composed of two historically

469   isolated, independently evolving sets of populations (Larmuseau et al., 2009). On the other

470   hand, our results indicate that genetic differentiation in *Squalius cephalus* was not correlated

471   to taxonomic differentiation, but was well explained by geographic distance. This suggests

472   that dispersal limitation has a more important effect than any other processes on broad-scale

473   genetic differentiation in *S. cephalus*. This might seem surprising, since previous analyses of

474   the mitochondrial gene cytochrome b in that species indicated four main lineages originating

475   from multiple glacial refugia (Durand, Persat, & Bouvet, 1999; Seifertová et al., 2012).

476   However, the use of a combination of both mitochondrial and nuclear DNA, Seifertová et al.

477   (2012) also suggested that genetic differentiation in *S. cephalus* resulted from a combination

478   of different factors, i.e. post-glacial colonisation from different refugia or recent evolutionary

479   processes such as drift or dispersal limitation. Our results based on the analysis of the

480   mitochondrial gene COI strongly support the dispersal limitation hypothesis although, as we

481   did not design the sampling, but analysed existing data, we might not have captured the whole

482   picture. As *R. rutilus* and *S. cephalus* have similar dispersal abilities, this difference of β-

483   SGDC between the two species does not seem to be related to their difference in dispersal

484   abilities. Beyond these two species and this single trait, our analyses did not permit us to

485   uncover any trait that could explain the variability of β-SGDC over the 22 species we

486   examined. This is surprising because life-history traits of species are known to shape spatial

487   patterns of genetic differentiation (e.g. Duminil et al., 2007; Kelly & Palumbi, 2010), and

488   therefore β-SGDC. This unexpected finding may be due to different reasons. First, the

489   variability of β-SGDC may be due to species traits that we did not consider. Secondly, the β-

490   SGDC at the specific scale was sometimes examined for a small number of basins, which may

491   result in wide variations in the estimated correlation coefficients. Therefore, single species

492   outcomes should be interpreted with great caution.

493   To conclude, this study of broad-scale β-SGDC for multiple freshwater fish species

494   suggests that historical processes greatly contributed to the shape of present-day patterns of

495   taxonomic and genetic differentiations, independently of contemporary processes. This

496   implies that taxonomic differentiation may be an appropriate proxy to explain the role of

497   historical processes on patterns of genetic differentiation, which remains a challenging issue

498   in landscape genetics (Dyer, Nason, & Garrick, 2010). We believe that further studies of

499   broad-scale β-SGDCs comparing organisms whose dispersal is limited within habitat patches

500   with organisms that are able to disperse across habitat patches, will help to investigate the

501   extent to which our findings can be generalised.
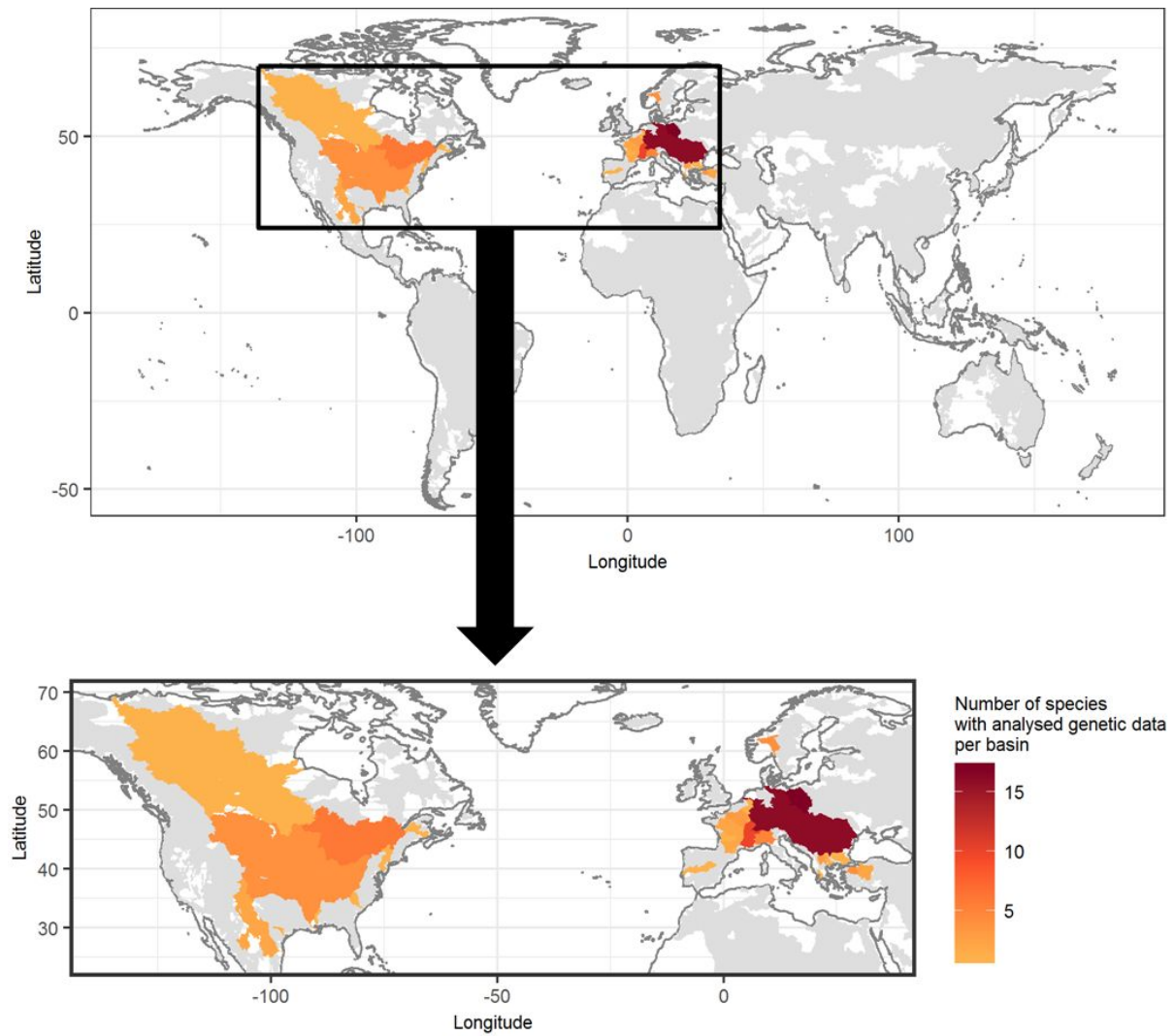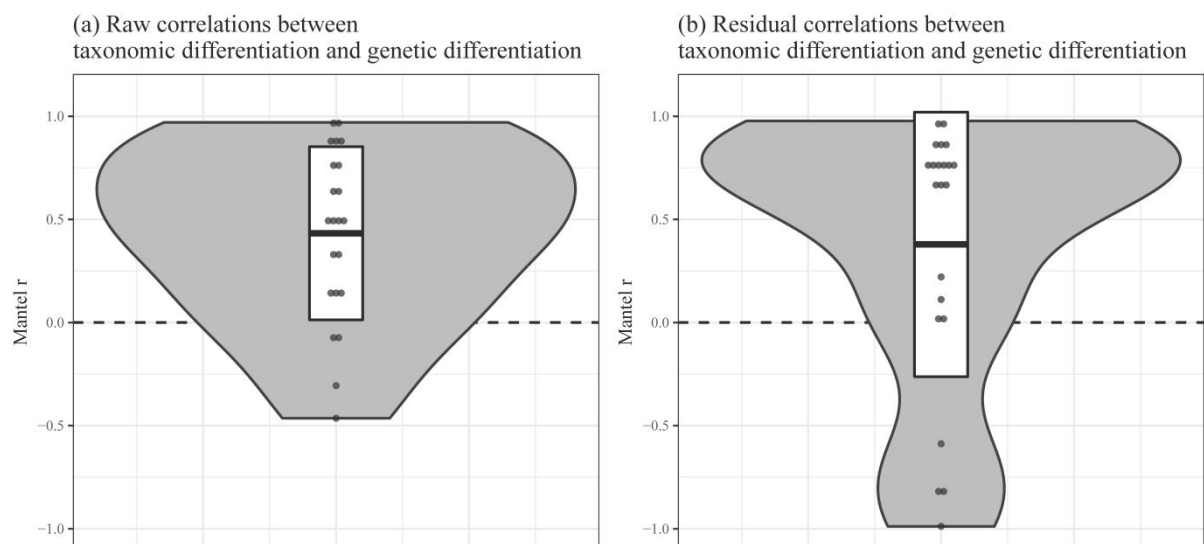
503   **Figures**

504

505   **Figure 1.** Illustration of the different processes expected to act in parallel on between-basin

506   taxonomic differentiation of freshwater fish communities (TD) and on between-basin genetic

507   differentiation of freshwater fish populations (GD) at different periods and the proxies used in

508   this study to infer those processes. The black frame encompasses the causal diagram depicting

509   the relationships that we actually tested in this study, i.e. the effects of geographic distance

510   (proxy of contemporary dispersal limitation), environmental distance (proxy of contemporary

511   environmental selection) and pairwise harmonic mean area between basins (proxy of

512   contemporary drift) on TD and GD. The double arrow between TD and GD represent the

513   residual correlation, i.e. the correlation between TD and GD which is not explained by the

514   parallel influence of contemporary dispersal limitation, contemporary environmental selection

515   and contemporary drift on TD and GD. The numbers on the arrows and the signs below them

516   correspond to our theoretical expectations regarding the different effects which are detailed in

517   the text below the black frame. The grey arrows represent the putative effects of historical

518   dispersal limitation, historical environmental selection and historical drift on TD and GD that

519   we did not directly test in our statistical approach. The grey polygons on the right side

520   represent the evolution of the basin configuration through time from one unique basin to two

521   distinct and disconnected basins, with almost no contemporary dispersal of individuals

522   between them (dispersal may occasionally occur between adjacent basins).
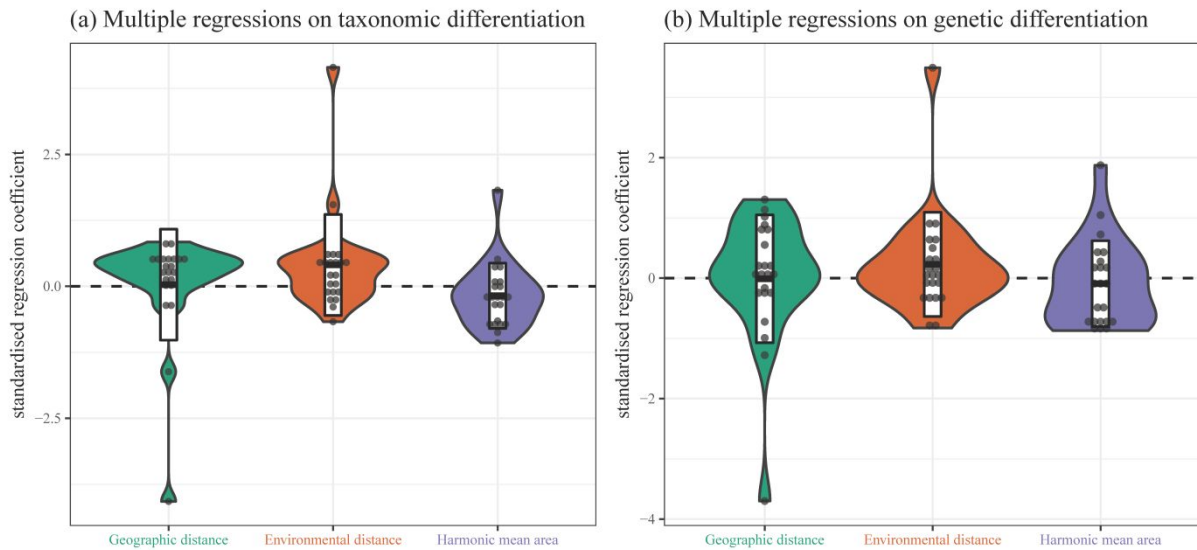
524

Figure 2. Spatial distribution of the number of species with analysed genetic data per basin

studied in our analyses at the global scale. Top: the rectangle indicates the location of the

enlarged area. Bottom: enlarged area including Europe and North America. In total, our study

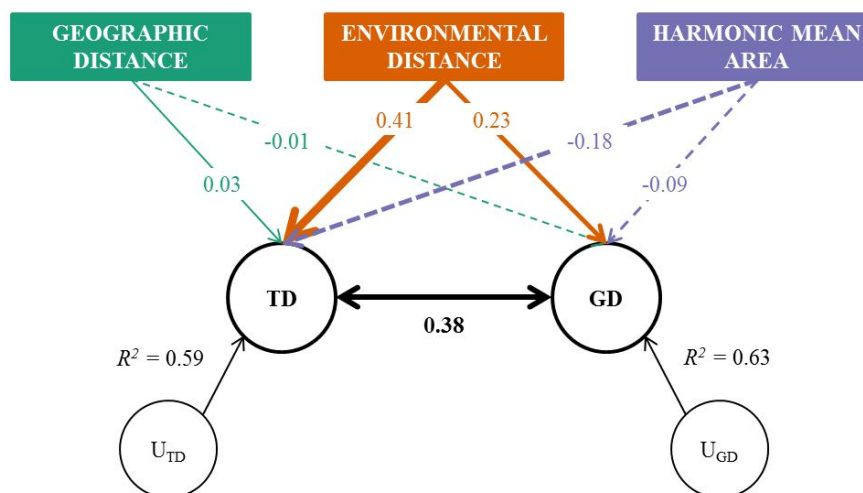encompasses 22 species with analysed genetic data in 38 basins.

530

Figure 3. Violin plots showing the distribution and the probability density of (a) raw correlations coefficients and (b) residual correlation coefficients between taxonomic differentiation and genetic differentiation for the 22 fish species examined at global scale. Crossbars indicate mean ± standard deviation.
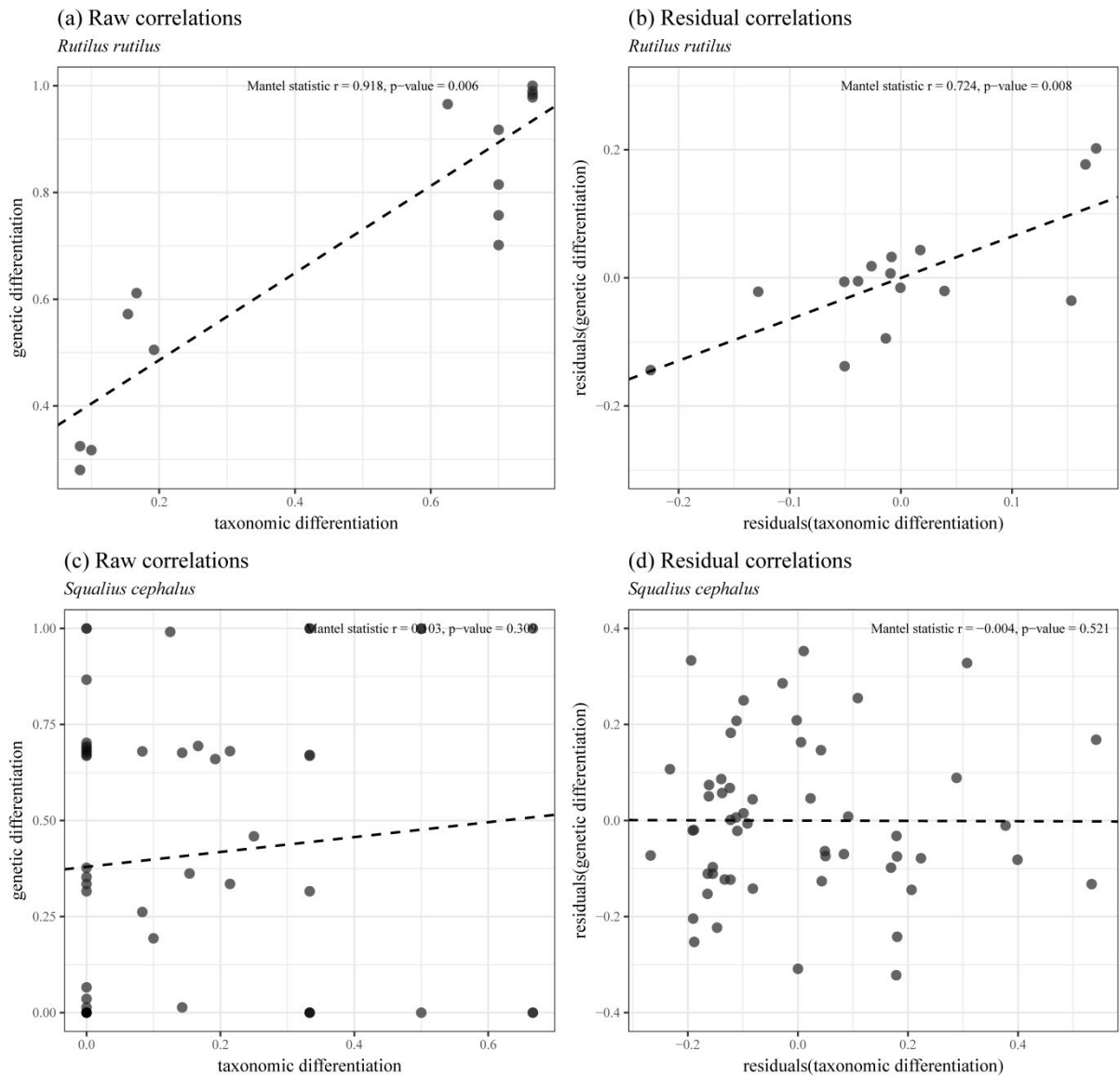
536

Figure 4. Violin plots showing the distribution and the probability density of the effect of geographic distance, environmental distance and harmonic mean area (as standardised regression coefficients) for the 22 fish species examined at global scale on (a) taxonomic differentiation and (b) genetic differentiation. Crossbars indicate mean ± standard deviation.

542

543　**Figure 5.** Causal diagram depicting the results of multiple regressions of geographic distance,

544　environmental distance and harmonic mean area on TD and GD (taken separately). Arrows

545　represent putative causal effects through the standardised regression coefficients (positive

546　values: solid arrows; negative values: dotted arrows; arrow width is proportional to coefficient

547　value). The R^2 represent the part of variance explained by the multiple regressions on TD

548　(left) and GD (right). U_{TD} and U_{GD} represent unspecified factors influencing TD and GD

549　respectively. The double arrow between SD and GD corresponds to the residual correlation.

550　All numerical values represent means for the 22 fish species examined at global scale.

Figure 6. Observed (dots) and predicted (dashed line) raw values (a, c) and residual values from the multiple regressions (b, d) of genetic differentiation as a function of taxonomic differentiation for the two fish species (a, b) *Rutilus rutilus* and (c, d) *Squalius cephalus*.

References

556   Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity.

558   *Zoobiology*, 5(2), 181–190.

559   Almany, G. R. (2003). Priority effects in coral reef fish communities. *Ecology*, 84(7), 1920–

560   1935.

561   April, J., Hanner, R. H., Mayden, R. L., & Bernatchez, L. (2013). Metabolic rate and climatic

562   fluctuations shape continental wide pattern of genetic divergence and biodiversity in

563   fishes. *PLoS ONE*, 8(7), e70296.

564   Arnold, J.B. (2018). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package

565   version 4.0.1. Retrieved from https://CRAN.R-project.org/package=ggthemes.

566   Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012).

567   Distance decay of similarity in freshwater communities: do macro-and microorganisms

568   follow the same rules?. *Global Ecology and Biogeography*, 21(3), 365–375.

569   Baselga, A. (2012). The relationship between species replacement, dissimilarity derived from

570   nestedness, and nestedness. *Global Ecology and Biogeography*, 21(12), 1223–1232.

571   Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., &

572   Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal

573   continuum of biodiversity at species and genetic levels. *Nature Communications*,

574   4(1892).

575   Baselga, A., Gómez-Rodríguez, C., & Vogler, A. P. (2015). Multi-hierarchical macroecology

576   at species and genetic levels to discern neutral and non-neutral processes. *Global*

577   *Ecology and Biogeography*, 24(8), 873–882.

578   Baselga, A., Orme, D., Villeger, S., De Bortoli, J., & Leprieur, F. (2018). betapart:

579   Partitioning Beta Diversity into Turnover and Nestedness Components. R package

580   version 1.5.0. Retrieved from https://CRAN.R-project.org/package=betapart.

581   Bivand, R., & Lewin-Koh, N. (2018). maptools: Tools for Reading and Handling Spatial

582   Objects. R package version 0.9-3. Retrieved from https://CRAN.R-

583   project.org/package=maptools.

584   Bivand, R., Keitt, T., & Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data

585   Abstraction Library. R package version 1.3-4. Retrieved from https://CRAN.R-

586   project.org/package=rgdal.

587   Bivand, R., & Rundel, C. (2018). rgeos: Interface to Geometry Engine - Open Source

588   ('GEOS'). R package version 0.3-28. Retrieved from https://CRAN.R-

589   project.org/package=rgeos.

590   Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to go bigger: emerging patterns in

591   macrogenetics. *Trends in Genetics*, 33(9), 579–580.

592   Carvalho, G. R. (1993). Evolutionary aspects of fish distribution: genetic variability and

593   adaptation. *Journal of Fish Biology*, 43(Supplement A), 53–73.

594   Derry, A. M., Arnott, S. E., Shead, J. A., Hebert, P. D., & Boag, P. T. (2009). Ecological

595   linkages between community and genetic diversity in zooplankton among boreal shield

596   lakes. *Ecology*, 90(8), 2275–2286.

597   Dias, M. S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J. F., … Tedesco,

598   P. A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity.

599   *Ecology Letters*, 17(9), 1130–1140.

600 Dray, S., & Dufour, A. (2007). The ade4 package: Implementing the Duality Diagram for

601 Ecologists. *Journal of Statistical Software*, 22(4), 1–20.

602 Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J.

603 (2007). Can population genetic structure be predicted from life-history traits?. *The*

604 *American Naturalist*, 169(5), 662–672.

605 Durand, J. D., Persat, H., & Bouvet, Y. (1999). Phylogeography and postglacial dispersion of

606 the chub (*Leuciscus cephalus*) in Europe. *Molecular Ecology*, 8(6), 989–997.

607 Dyer, R. J., Nason, J. D., & Garrick, R. C. (2010). Landscape modelling of gene flow:

608 improved power using conditional genetic distance derived from the topology of

609 population networks. *Molecular Ecology*, 19(17), 3746–3759.

610 Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J. G., & Blanchet, S. (2016). Lessons from the

611 fish: a multi-species analysis reveals common processes underlying similar species-

612 genetic diversity correlations. *Freshwater Biology*, 61(11), 1830–1845.

613 Fraser, C. I., Davies, I. D., Bryant, D., Waters, J. M. (2018). How disturbance and dispersal

614 influence intraspecific structure. *Journal of Ecology*, 106,1298–1306.

615 Froese, R., & Pauly, D. (2017). FishBase. World Wide Web electronic publication.

616 www.fishbase.org, version (06/2017).

617 Gilbert, B., & Levine, J. M. (2017). Ecological drift and the distribution of species diversity.

618 *Proceedings of the Royal Society B: Biological Sciences*, 284(1855), 20170507.

619 Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics.

620 R package version 0.04-22. Retrieved from https://CRAN.R-

621 project.org/package=hierfstat

622   Hahsler, M., Buchta, C., & Hornik, K. (2018). seriation: Infrastructure for Ordering Objects

623   Using Seriation. R package version 1.2-3.Retrieved from https://CRAN.R-

624   project.org/package=seriation.

625   Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8),

626   1633–1638.

627   Hugueny, B. (1989). West African rivers as biogeographic islands: species richness of fish

628   communities. *Oecologia*, 79, 236–243.

629   Kahilainen, A., Puurtinen, M., & Kotiaho, J. S. (2014). Conservation implications of species-

630   genetic diversity correlations. *Global Ecology and Conservation*, 2, 315–323.

631   Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern

632   Pacific rocky intertidal community. *PLoS ONE*, 5(1), e8594.

633   Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species-

634   genetic diversity correlations to the understanding of community assembly rules. *Oikos*,

635   126(6), 759–771.

636   Larmuseau, M. H. D., Freyhof, J., Volckaert, F. A. M., & Van Houdt, J. K. J. (2009).

637   Matrilinear phylogeography and demographical patterns of *Rutilus rutilus*: implications

638   for taxonomy and conservation. *Journal of Fish Biology*, 75(2), 332–353.

639   Laroche, F., Jarne, P., Lamy, T., David, P., & Massol, F. (2014). A neutral theory for

640   interpreting correlations between species and genetic diversity in communities. *The*

641   *American Naturalist*, 185(1), 59–69.

642   Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior:

643   a permutational regression approach. *Evolution*, 48(5), 1487–1499.

untypeset proof

644 Leprieur, F., Olden, J. D., Lek, S., & Brosse, S. (2009). Contrasting patterns and mechanisms

645   of spatial turnover for native and exotic freshwater fish in Europe. *Journal of*

646   *Biogeography*, 36(10), 1899–1912.

647 Leprieur, F., Tedesco, P. A., Hugueny, B., Beauchard, O., Dürr, H. H., Brosse, S., &

648   Oberdorff, T. (2011). Partitioning global patterns of freshwater fish beta diversity reveals

649   contrasting signatures of past climate changes. *Ecology Letters*, 14(4), 325–334.

650 Lichstein, J. W. (2007). Multiple regression on distance matrices: a multivariate spatial

651   analysis tool. *Plant Ecology*, 188(2), 117–131.

652 Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

653 Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid

654   them. *Molecular Ecology*, 24, 3223–3231.

655 Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F_{ST} and related

656   measures. *Molecular Ecology Resources*, 11(1), 5–18.

657 Moritz, C., Meynard, C. N., Devictor, V., Guizien, K., Labrune, C., Guarini, J. M., &

658   Mouquet, N. (2013). Disentangling the role of connectivity, environmental filtering, and

659   spatial structure on metacommunity dynamics. *Oikos*, 122(10), 1401–1410.

660 Oberdorff, T., Tedesco, P. A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr,

661   H. H. (2011). Global and regional patterns in riverine fish species richness: a review.

662   *International Journal of Ecology*, 201(967631).

663 Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., …Wagner, H.

664   (2018). vegan: Community Ecology Package. R package version 2.5-2. Retrieved from

665   https://CRAN.R-project.org/package=vegan.

666 Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of

667   population genetic differentiation in the wild: isolation by dispersal limitation, isolation

668   by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999.

669 Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the*

670   *National Academy of Sciences*, 70(12), 3321–3323.

671 Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and

672   ecology. *Journal of Biogeography*, 26(4), 867–878.

673 Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.

674   Retrieved from https://CRAN.R-project.org/package=RColorBrewer.

675 Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2018). Biostrings: Efficient

676   manipulation of biological strings. R package version 2.48.0.

677 Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis,

678   A., & Vogler, A. P. (2011). Testing the species–genetic diversity correlation in the

679   Aegean Archipelago: toward a haplotype-based macroecology? *The American*

680   *Naturalist*, 178(2), 241–255.

681 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution

682   in R language. *Bioinformatics*, 20, 289–290.

683 Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part II. *Ecology*,

684   43(3), 410–432.

685 Prunier, J. G., Dubut, V., Chikhi, L., & Blanchet, S. (2017). Contribution of spatial

686   heterogeneity in effective population sizes to the variance in pairwise measures of

687   genetic differentiation. *Methods in Ecology and Evolution*, 8(12), 1866–1877.

688   Qian, H., & Ricklefs, R. E. (2012). Disentangling the effects of geographic distance and

689   environmental dissimilarity on global patterns of species turnover. *Global ecology and*

690   *Biogeography*, 21(3), 341–351.

691   R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna,

692   Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-

693   project.org/.

694   Radinger, J., & Wolter, C. (2014). Patterns and predictors of fish dispersal in rivers. *Fish and*

695   *fisheries*, 15(3), 456–473.

696   Robinson, D., & Hayes, A. (2018). broom: Convert Statistical Analysis Objects into Tidy

697   Tibbles. R package version 0.5.0. Retrieved from https://CRAN.R-

698   project.org/package=broom.

699   Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge, UK: Cambridge

700   University Press.

701   Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under

702   isolation by distance. *Genetics*, 145(4), 1219-–1228.

703   Seifertová, M., Bryja, J., Vyskočilová, M., Martínková, N., & Šimková, A. (2012). Multiple

704   Pleistocene refugia and post-glacial colonization in the European chub (*Squalius*

705   *cephalus*) revealed by combined use of nuclear and mitochondrial markers. *Journal of*

706   *Biogeography*, 39(6), 1024–1040.

707   Sepkoski, J. J., & Rex, M. A. (1974). Distribution of freshwater mussels: coastal rivers as

708   biogeographic islands. *Systematic Zoology*, 23(2), 165–188.

709   Serrouya, R., Paetkau, D., McLellan, B. N., Boutin, S., Campbell, M., & Jenkins, D. A.

710   (2012). Population size and major valleys explain microsatellite variation better than

711   taxonomic units for caribou in western Canada. *Molecular Ecology*, 21(11), 2588–2601.

712   Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in

713   ecological communities. *Ecography*, 30(1), 3–12.

714   Sørensen, T. J. (1948). A method of establishing groups of equal amplitude in plant sociology

715   based on similarity of species content, and its application to analyses of the vegetation on

716   Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, 5, 1–

717   34.

718   Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., … Oberdorff,

719   T. (2017). A global database on freshwater fish species occurrence in drainage basins.

720   *Scientific Data*, 4(170141).

721   Tedesco, P. A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H. H., Beauchard, O., …

722   Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global

723   Ecology and Biogeography*, 21(10), 977–987.

724   Tisseuil, C., Cornu, J. F., Beauchard, O., Brosse, S., Darwall, W., Holland, R., … Oberdorff,

725   T. (2013). Global diversity patterns and cross-taxa convergence in freshwater systems.

726   *Journal of Animal Ecology*, 82(2), 365–376.

727   Vellend, M. (2003). Island biogeography of genes and species. *The American Naturalist*,

728   162(3), 358–365.

729   Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic

730   diversity. *Ecology Letters*, 8(7), 767–781.

731   Vellend, M., & Orrock, J. L. (2009). Ecological and genetic models of diversity. *The theory of*

732   *Island biogeography revisited*. Princeton Univ. Press, Princeton, NJ, 439–462.

733   Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., & Garant, D. (2014).

734   Drawing ecological inferences from coincident patterns of population- and community-

735   level biodiversity. *Molecular Ecology*, 23(12), 2890–2901.

736   Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and

737   geography in spatial genetic divergence. *Ecology Letters*, 16(2), 175–182.

738   Whitlock, M. C. (2005). Combining probability from independent tests: the weighted Z-

739   method is superior to Fisher's approach. *Journal of Evolutionary Biology*, 18(5), 1368–

740   1373.

741   Whittaker, R. H. (1960). Vegetation of the Siskiyou mountains, Oregon and California.

742   *Ecological monographs*, 30(3), 279–338.

743   Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 213-251.

744   Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical

745   Software*, 21(12), 1-20.

746   Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New-York, NY: Springer-

747   Verlag.

748   Winter, D. J. (2012). mmod: an R library for the calculation of population differentiation

749   statistics. *Molecular Ecology Resources*, 12(6), 1158-1160.

750   Wright, S. (1943). Isolation by distance. *Genetics*, 28(March), 114–138.

751   Wright, S. (1951). The genetical structure of natural populations. *Annals of Eugenics*, 15(4),

752   323–354.

753   **Biosketch**

754   The authors are particularly interested in documenting and explaining spatial patterns of

755   biodiversity at large spatial scales. They try to disentangle the interplay between historical and

756   contemporary factors (including anthropogenic pressures) in shaping patterns of alpha and

757   beta diversities in aquatic populations and communities and value this knowledge to guide

758   conservation strategies of aquatic ecosystems.

759   **Supporting information**

760   Supporting information for this article may be found online.

761   **Data availability**

762   The data and the script used for the analyses are available from the Dryad Digital Repository:

763   https://doi.org/10.5061/dryad.8vb6160.

Journal of Biogeography

SUPPORTING INFORMATION

Correlations between broad-scale taxonomic and genetic differentiations suggest a dominant imprint of historical processes on beta diversities

Marine Robuchon, Boris Leroy, Céline Jézéquel and Bernard Hugueny

Appendix S1 Characteristics of the drainage basins analysed in our study

aet = mean annual actual evapotranspiration; pet = mean annual potential evapotranspiration; tmoy = mean annual temperature ; precsum = mean annual precipitation; runoff = mean annual surface runoff; Surface_area_km = surface area; N_species = number of species with genetic data. See main text for the source of data. Mean and standard deviation (SD) over the 38 basins are given at the end of the table.

Basins	aet	pet	tmoy	precsum	runoff	Surf_area_km	N_species
Acheloos	640	1066.1	123.27929	924.009	557.74899	5821.728156	1
Agly	630	906.3	125.49053	770.632	116.367	1125.200682	3
Arachthos	655	1034.7	113.31274	1004.44	742.65767	2473.939349	1
Aude	625	928.5	123.53473	781.976	130.21899	5283.136409	1
Axios	493	986.3	101.52936	601.732	154.31825	24583.23488	3
Charente	644	911.6	120.03306	867.967	233.5864	9561.56072	1
Churchill	398	582	-23.90383	475.655	118.85188	304248.866	1
Danube	574	850.7	87.342907	750.564	295.57456	802142.8624	16
Delaware	813	991.2	91.341375	1148.11	623.8048	28618.74225	1
Dordogne	682	935.2	111.27602	886.614	350.67982	24113.64285	1
Elbe	507	721.7	79.989063	623.004	178.65205	143864.5137	16
Evinos	629	1068.6	124.6515	898.565	462.55499	1105.990108	1
Evros	525	994.7	112.64752	622.026	233.69195	53005.94877	1
Garonne	653	932	111.66166	822.437	258.36741	55217.94818	2
Glomma	404	469.2	8.6601708	730.294	415.933	41930.43322	4
Grande.USA	333	1517.8	163.21044	395.802	11.279145	480465.5835	2
Guadalupe	714	1510	198.33098	844.73	31.73177	15777.64583	1
Hudson	756	924.8	71.827872	1118.13	526.25893	34795.07612	1
Loire	602	854.6	105.10719	750.497	196.85747	116661.6899	2
Louros	663	1121.3	145.51997	1040.27	713.57898	1080.177613	1
Mackenzie	329	521.7	-43.57659	392.294	160.29758	1680044.211	1
Meuse	588	732.3	91.633783	886.959	285.14163	32856.16342	1
Mississippi	618	1096.6	95.416708	759.456	222.75135	3017672.487	4
Nelson	418	726	12.974033	492.429	95.228112	1019864.883	1
Odra	489	737.8	81.363189	581.098	155.37879	118991.636	17
Pearl	1119	1437.3	178.64183	1495.1	483.74908	21790.63225	1

Po	635	788.2	92.085106	1020.21	519.70202	73831.63159	5
Rhine	588	738.1	82.161385	872.964	475.19941	163633.2382	16
Rhone	629	802.8	85.285926	1019.07	479.94488	96827.05122	10
Rihios	414	1063.9	139.98324	484.53	1.7229	2088.952236	2
Saint.Jean.USA	635	762.6	36.171725	1049.03	608.68723	54888.62971	1
Saint.Laurent	629	811.7	49.93017	884.414	343.00659	1056005.322	6
Sakarya	421	1033.3	106.13612	497.924	62.963135	70666.64849	2
Santee	985	1322.8	155.71445	1239.84	132.91453	39531.78697	1
Seine	559	811.4	101.43787	690.465	173.8974	75984.27645	3
Strymon	459	957.1	104.39717	545.061	93.719109	16826.18541	1
Susurluk.Cayi	578	1056.4	120.13898	727.301	111.40774	23768.99736	4
Tagus	459	1097.2	136.86304	565.262	214.13121	71216.07416	1
Mean	591.9	942.2	97.9	796.3	288.8	257588.6	3.6
SD	158.5	237.6	50.0	245.9	204.0	584802.5	4.8

Appendix S2 Descriptive statistics of the datasets used for analyses

N_basins_genet = number of basins with genetic data analysed; N_basins_native = number of basins where the species is recoreded as native; Prop_basins_genet/native = proportion of basins with genetic data analysed over basins where the species is recorded as native; N_indiv_tot = total number of individuals; N_indiv_mean = mean number of individuals by basin; N_indiv_min = minimum number of individuals in a basin; N_indiv_max = maximum number of individuals in a basin. Mean and standard deviation (SD) over the 22 species are given at the end of the table.

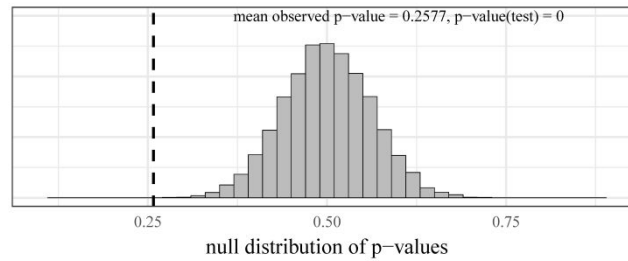| Species	N_basins_genet	N_basins_native	Prop_basins_genet/native	N_indiv_tot	N_indiv_mean	N_indiv_min	N_indiv_max
Abramis brama	4	8	0.500	46	12	6	16
Alburnoides bipunctatus	5	5	1.000	31	6	4	13
Alburnus alburnus	5	8	0.625	43	9	3	18
Barbatula barbatula	4	4	1.000	55	14	5	22
Campostoma anomalum	4	6	0.667	30	8	5	15
Chondrostoma nasus	4	5	0.800	71	18	5	35
Esox lucius	11	16	0.688	97	9	3	15
Gobio gobio	7	8	0.875	71	10	3	22
Lampetra planeri	4	6	0.667	27	7	3	10
Leuciscus aspius	4	5	0.800	34	9	3	17
Leuciscus leuciscus	4	5	0.800	47	12	4	19
Lota lota	5	7	0.714	37	7	5	13
Notropis atherinoides	4	4	1.000	25	6	4	11
Phoxinus phoxinus	4	4	1.000	234	59	5	207
Rhinichthys atratulus	5	5	1.000	28	6	3	11
Rhinichthys cataractae	4	5	0.800	39	10	5	17
Rutilus rutilus	6	11	0.545	58	10	3	17
Scardinius erythrophthalmus	4	11	0.364	26	7	3	10
Squalius cephalus	11	16	0.688	87	8	3	21
Telestes pleurobipunctatus	4	6	0.667	20	5	3	9
Thymallus thymallus	6	7	0.857	41	7	4	11
Tinca tinca	5	7	0.714	32	6	3	13
Mean	5.2	7.2	0.762	53.6	11.0	3.9	24.6
SD	2.1	3.4	0.175	45.3	11.0	1.0	41.1

Appendix S3 Results of the Mantel tests between taxonomic differentiation and genetic differentiation (raw correlations) and between the residuals of taxonomic differentiation and genetic differentiation (residual correlations)

r is the Mantel r coefficient; mean and standard deviation (SD) over the 22 species are given at the end of the table

Species	Raw correlations		Residual correlations	
	r	pval	r	pval
Abramis brama	0.539	0.167	0.949	0.042
Alburnoides bipunctatus	0.963	0.100	0.885	0.008
Alburnus alburnus	-0.098	0.333	-0.781	0.975
Barbatula barbatula	0.327	0.458	0.716	0.083
Campostoma anomalum	0.333	0.250	-0.857	1.000
Chondrostoma nasus	0.841	0.042	0.784	0.042
Esox lucius	0.685	0.077	0.040	0.406
Gobio gobio	0.586	0.174	0.112	0.354
Lampetra planeri	-0.306	0.583	0.900	0.083
Leuciscus aspius	0.448	0.167	0.825	0.042
Leuciscus leuciscus	0.511	0.167	-0.989	1.000
Lota lota	0.918	0.050	0.793	0.008
Notropis atherinoides	0.725	0.208	0.221	0.375
Phoxinus phoxinus	-0.049	0.708	0.802	0.042
Rhinichthys atratulus	0.488	0.033	0.682	0.025
Rhinichthys cataractae	0.099	0.625	-0.588	1.000
Rutilus rutilus	0.918	0.006	0.724	0.008
Scardinius erythrophthalmus	0.970	0.250	0.978	0.042
Squalius cephalus	0.103	0.304	-0.004	0.507
Telestes pleurobipunctatus	-0.464	0.750	0.798	0.042
Thymallus thymallus	0.799	0.001	0.730	0.008
Tinca tinca	0.187	0.217	0.618	0.017
Mean	0.433	0.258	0.379	0.278
SD	0.420	0.229	0.641	0.376

Appendix S4. Tests of the overall correlation (over the 22 species) between taxonomic differentiation and genetic differentiation for (a) the raw correlations and (b) the residual correlations (i.e. the correlations between the residuals of the multiple regressions). The p-values associated to these tests were obtained by comparing the mean observed p-value over the species to a null distribution of p-values. Dashed vertical lines in test plots show the mean observed p-value over the species.

(a) Raw correlations between
taxonomic differentiation and genetic differentiation

mean observed p–value = 0.2577, p–value(test) = 0

0.25   0.50   0.75
null distribution of p–values

(b) Residual correlations between
taxonomic differentiation and genetic differentiation

mean observed p–value = 0.2777, p–value(test) = 3e–04

0.25   0.50   0.75
null distribution of p–values
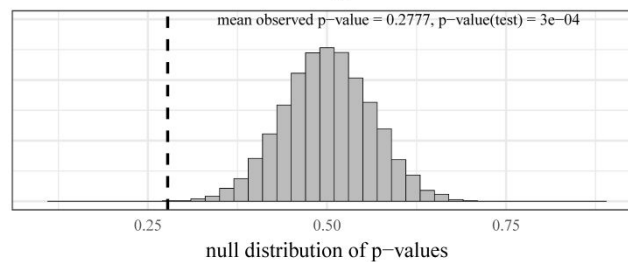
Appendix S5 Results of multiple regressions on taxonomic differentiation

All the coefficients are standardised partial regression coefficients; R2 is the total variance explained by the multiple regression; mean and standard deviation (SD) over the 22 species are given at the end of the table.

Species	Intercept	pval.Intercept	Geo	pval.Geo	Env	pval.Env	Area	pval.Area	R2	pval.R2
Abramis brama	0.659	0.701	0.521	0.398	0.475	0.298	-0.202	0.360	0.521	0.624
Alburnoides bipunctatus	2.261	0.207	0.338	0.207	-0.084	0.550	-0.763	0.207	0.784	0.020
Alburnus alburnus	2.194	0.199	0.416	0.079	-0.669	0.813	-0.670	0.164	0.388	0.367
Barbatula barbatula	0.659	0.628	0.521	0.326	0.475	0.138	-0.202	0.323	0.521	0.624
Campostoma anomalum	0.672	0.882	0.305	0.461	0.570	0.329	0.324	0.700	0.797	0.288
Chondrostoma nasus	0.659	0.705	0.521	0.406	0.475	0.301	-0.202	0.366	0.521	0.624
Esox lucius	0.317	0.092	0.558	0.090	0.073	0.267	0.410	0.989	0.702	0.000
Gobio gobio	1.402	0.445	0.150	0.331	0.015	0.438	-0.711	0.099	0.602	0.001
Lampetra planeri	1.962	0.844	0.470	0.387	0.195	0.503	-0.389	0.330	0.978	0.032
Leuciscus aspius	-0.253	0.908	-0.333	0.540	0.641	0.460	0.083	0.464	0.085	0.975
Leuciscus leuciscus	0.659	0.618	0.521	0.326	0.475	0.126	-0.202	0.331	0.521	0.624
Lota lota	0.421	0.794	0.770	0.233	0.568	0.142	0.036	0.628	0.853	0.007
Notropis atherinoides	0.605	0.710	0.101	0.461	0.426	0.379	-0.303	0.369	0.594	0.542
Phoxinus phoxinus	0.659	0.629	0.521	0.375	0.475	0.131	-0.202	0.332	0.521	0.624
Rhinichthys atratulus	0.344	0.859	0.061	0.510	-0.287	0.684	0.513	0.791	0.314	0.488
Rhinichthys cataractae	-9.883	0.373	-4.072	0.882	4.147	0.079	1.821	0.830	0.576	0.563
Rutilus rutilus	1.819	0.391	0.312	0.164	-0.132	0.559	-0.875	0.089	0.873	0.000
Scardinius erythrophthalmus	-0.590	0.748	-0.385	0.662	-0.219	0.670	-0.712	0.167	0.789	0.299
Squalius cephalus	0.034	0.956	-0.025	0.527	0.228	0.196	-0.044	0.423	0.052	0.434
Telestes pleurobipunctatus	-87.769	0.253	-1.617	0.829	1.548	0.171	0.086	0.423	0.867	0.192
Thymallus thymallus	1.903	0.715	0.215	0.368	-0.089	0.542	-0.654	0.120	0.353	0.172
Tinca tinca	7.394	0.550	0.839	0.397	-0.389	0.603	-1.071	0.196	0.802	0.016
Mean	-3.358	0.600	0.032	0.407	0.405	0.381	-0.178	0.395	0.592	0.342
SD	19.067	0.254	1.051	0.204	0.956	0.214	0.617	0.250	0.249	0.288

Appendix S6 Results of multiple regressions on genetic differentiation

All the coefficients are standardised partial regression coefficients; R2 is the total variance explained by the multiple regression; mean and standard deviation (SD) over the 22 species are given at the end of the table.
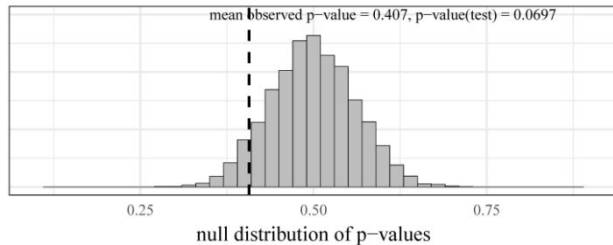
Table S6.1 Results when genetic differentiation is measured with G''$_{ST}$

Species	Intercept	pval.Intercept	Geo	pval.Geo	Env	pval.Env	Area	pval.Area	R2	pval.R2
Abramis brama	-7.853	0.142	-1.278	0.954	0.630	0.089	1.047	1.000	0.286	0.847
Alburnoides bipunctatus	1.667	0.375	0.156	0.349	0.045	0.496	-0.760	0.204	0.739	0.035
Alburnus alburnus	0.928	0.791	0.034	0.619	-0.084	0.442	-0.694	0.359	0.448	0.280
Barbatula barbatula	2.138	0.562	0.224	0.343	0.030	0.457	-0.854	0.174	0.466	0.682
Campostoma anomalum	3.536	0.001	1.136	0.000	-0.347	0.902	-0.042	0.310	0.931	0.101
Chondrostoma nasus	-0.458	0.952	-0.165	0.582	0.650	0.167	0.463	0.671	0.417	0.731
Esox lucius	1.025	0.097	1.021	0.097	-0.103	0.627	-0.138	0.335	0.927	0.000
Gobio gobio	5.610	0.120	0.768	0.132	-0.335	0.848	-0.511	0.144	0.737	0.000
Lampetra planeri	2.266	0.286	0.552	0.000	-0.742	1.000	0.397	1.000	0.335	0.806
Leuciscus aspius	-1.954	0.658	-0.992	0.711	0.147	0.437	-0.460	0.314	0.545	0.598
Leuciscus leuciscus	-2.365	1.000	-0.210	0.511	0.944	0.162	0.725	0.671	0.962	0.056
Lota lota	0.710	0.476	0.884	0.216	0.295	0.236	-0.094	0.507	0.757	0.028
Notropis atherinoides	-0.801	0.833	-0.255	0.594	0.500	0.244	-0.685	0.163	0.999	0.001
Phoxinus phoxinus	1.527	0.751	0.047	0.468	-0.304	0.754	-0.870	0.218	0.725	0.382
Rhinichthys atratulus	-1.992	0.417	-0.723	0.839	0.867	0.159	0.171	0.572	0.234	0.633
Rhinichthys cataractae	-1.374	0.198	-3.698	1.000	3.493	0.000	1.875	1.000	0.380	0.766
Rutilus rutilus	0.772	0.679	0.039	0.505	0.327	0.292	-0.682	0.130	0.872	0.000
Scardinius erythrophthalmus	-0.167	1.000	-0.275	0.523	-0.331	0.499	-0.799	0.242	0.929	0.105
Squalius cephalus	3.345	0.024	0.848	0.011	0.069	0.383	0.222	0.760	0.813	0.000
Telestes pleurobipunctatus	99.964	0.175	1.305	0.000	-0.826	1.000	0.275	0.825	0.657	0.467
Thymallus thymallus	2.520	0.570	0.255	0.330	0.175	0.378	-0.742	0.069	0.639	0.009
Tinca tinca	1.026	0.877	0.092	0.470	-0.054	0.496	0.131	0.591	0.036	0.972
Mean	5.003	0.499	-0.011	0.420	0.229	0.457	-0.092	0.466	0.629	0.341
SD	21.381	0.334	1.060	0.305	0.865	0.293	0.713	0.308	0.271	0.350

Table S6.2 Results when genetic differentiation is measured with F_{ST}

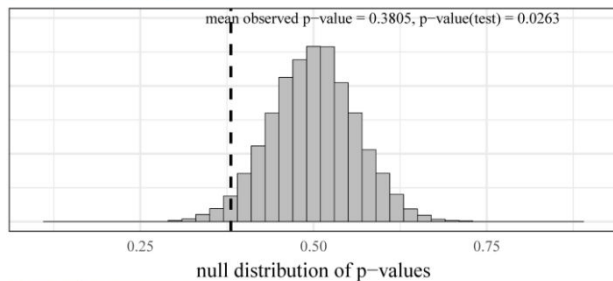| Species	Intercept	pval.Intercept	Geo	pval.Geo	Env	pval.Env	Area	pval.Area	R2	pval.R2
Abramis brama	0.808	0.810	0.170	0.481	-0.633	0.778	-0.813	0.277	0.698	0.416
Alburnoides bipunctatus	-1.278	0.198	-0.268	1.000	0.228	0.000	0.113	0.655	0.044	0.961
Alburnus alburnus	-3.214	0.108	-0.621	0.902	1.165	0.144	0.723	0.906	0.599	0.118
Barbatula barbatula	-4.092	0.543	-0.765	0.718	0.205	0.379	1.140	0.803	0.324	0.816
Campostoma anomalum	1.264	0.706	-0.093	0.530	-0.702	0.797	-0.207	0.627	0.675	0.446
Chondrostoma nasus	-0.127	0.870	-0.098	0.532	0.722	0.180	0.191	0.619	0.469	0.679
Esox lucius	0.230	0.357	-0.119	0.571	-0.155	0.726	0.051	0.538	0.038	0.578
Gobio gobio	1.649	0.540	0.183	0.298	-0.276	0.789	-0.120	0.338	0.047	0.839
Lampetra planeri	-4.038	0.379	-1.397	0.699	1.337	0.343	0.033	0.419	0.944	0.082
Leuciscus aspius	-1.608	0.309	-1.292	1.000	0.829	0.113	-0.268	0.340	0.380	0.766
Leuciscus leuciscus	-2.876	0.398	-0.536	0.763	0.103	0.326	-0.043	0.502	0.308	0.829
Lota lota	0.661	0.263	1.058	0.196	0.309	0.214	-0.275	0.262	0.971	0.000
Notropis atherinoides	-1.436	0.958	-0.390	0.505	0.609	0.495	-0.626	0.493	0.963	0.055
Phoxinus phoxinus	15.134	0.169	1.609	0.169	-0.234	0.792	-2.050	0.123	0.986	0.021
Rhinichthys atratulus	2.101	0.398	0.640	0.219	-0.560	0.778	-0.278	0.242	0.172	0.748
Rhinichthys cataractae	0.363	0.865	-0.235	0.553	0.169	0.443	0.630	0.756	0.282	0.850
Rutilus rutilus	-1.722	0.109	-0.469	0.991	1.028	0.033	0.102	0.603	0.582	0.019
Scardinius erythrophthalmus	6.458	0.254	0.950	0.000	-0.552	1.000	0.458	0.747	0.605	0.530
Squalius cephalus	0.188	0.808	-0.038	0.565	-0.065	0.511	0.019	0.453	0.010	0.913
Telestes pleurobipunctatus	100.388	0.339	1.331	0.159	-0.821	0.749	0.261	0.497	0.683	0.436
Thymallus thymallus	3.549	0.310	0.529	0.222	-0.608	0.855	-0.340	0.218	0.171	0.543
Tinca tinca	-0.616	0.847	-0.148	0.586	0.652	0.148	-0.214	0.375	0.512	0.202
Mean	5.081	0.479	0.000	0.530	0.125	0.481	-0.069	0.490	0.476	0.493
SD	21.673	0.277	0.774	0.295	0.648	0.310	0.623	0.209	0.324	0.333

Appendix S7. Tests of the overall effects (over the 22 species) of geographic distance, environmental distance and difference in area on (a) taxonomic differentiation, (b) genetic differentiation measured with G''_{ST} and (c) genetic differentiation measured with F_{ST}. The p-values associated to these tests were obtained by comparing the mean observed p-value over the species associated to each effect to a null distribution of p-values. Dashed vertical lines in test plots show the mean observed p-value over the species.

(a) Multiple regressions on taxonomic differentiation

Geographic distance

mean observed p−value = 0.407, p−value(test) = 0.0697

null distribution of p−values

Environmental distance

mean observed p−value = 0.3805, p−value(test) = 0.0263

null distribution of p−values

Harmonic mean area

mean observed p−value = 0.3952, p−value(test) = 0.0399

null distribution of p−values

(b) Multiple regression on genetic differentiation (G''_{ST})

Geographic distance

mean observed p−value = 0.4204, p−value(test) = 0.0992

null distribution of p−values

Environmental distance

mean observed p−value = 0.4574, p−value(test) = 0.2468

null distribution of p−values

Harmonic mean area

mean observed p−value = 0.466, p−value(test) = 0.2907

null distribution of p−values
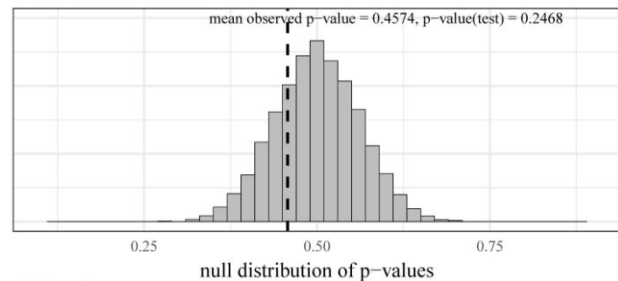
(c) Multiple regressions on genetic differentiation (F_{ST})

Geographic distance

mean observed p−value = 0.5297, p−value(test) = 0.6786

null distribution of p−values

Environmental distance

mean observed p−value = 0.4812, p−value(test) = 0.3803

null distribution of p−values

Harmonic mean area

mean observed p−value = 0.4902, p−value(test) = 0.4404

null distribution of p−values
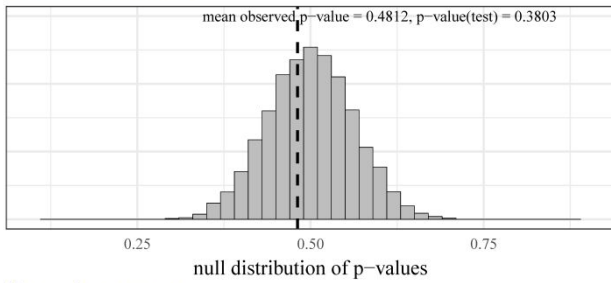
Appendix S8. Variations of the residual correlation coefficient between taxonomic differentiation and genetic differentiation (r) according to different species traits: (a) length, (b) dispersal ability, (c) longevity in the wild, (d) vulnerability, (e) use in aquaculture, (f) use as baits, (g) use in aquarium and (h) habitat.

Historical period
= basins connected

Dispersal

Environmental selection

Drift

Disconnection time

GEOGRAPHIC DISTANCE

ENVIRONMENTAL DISTANCE

HARMONIC MEAN AREA

1 +

2 +

3 +

4 +

5 -

6 -

TD

7

GD
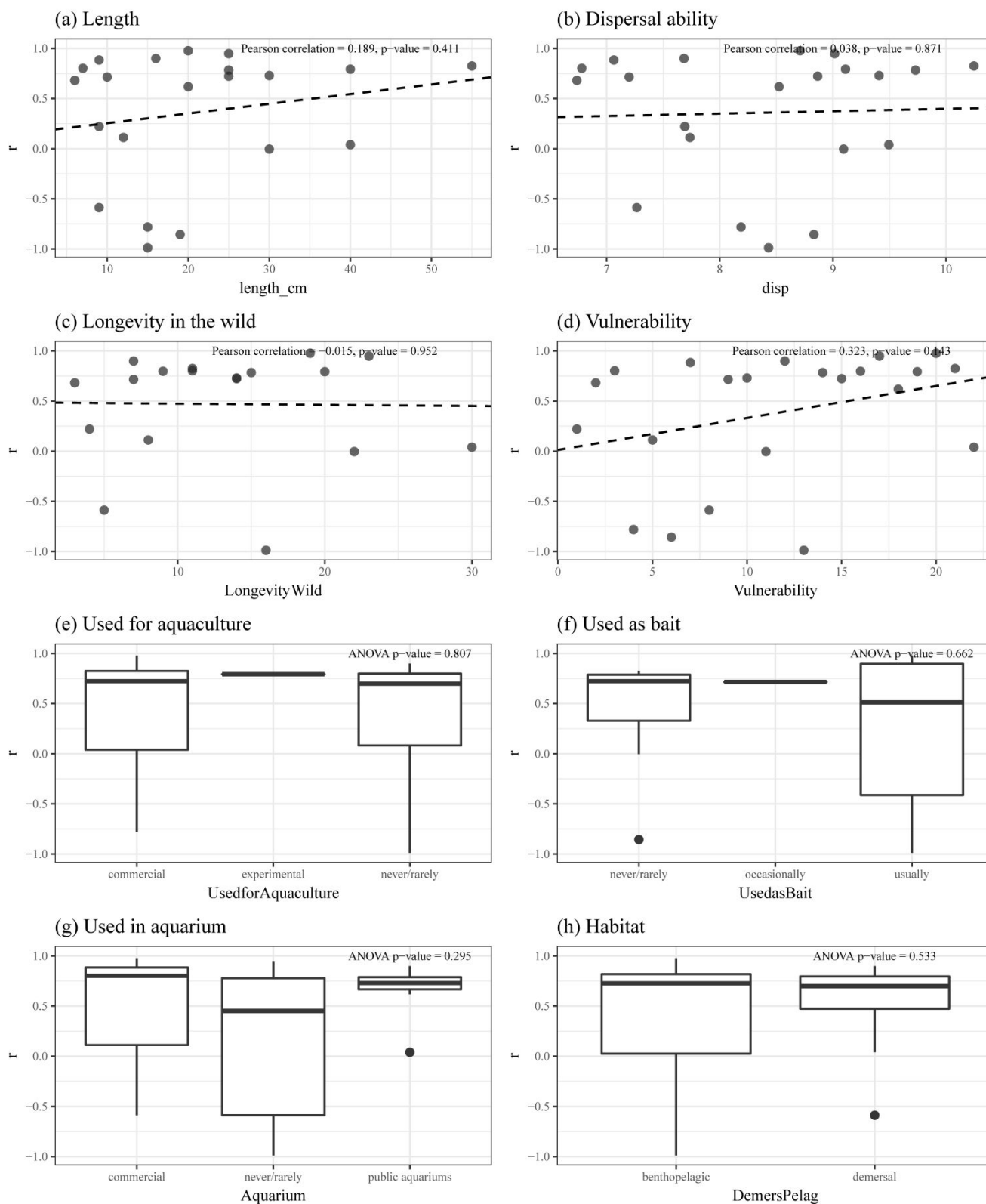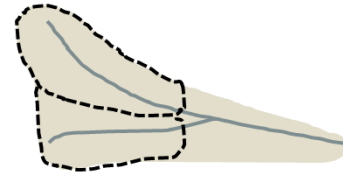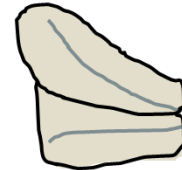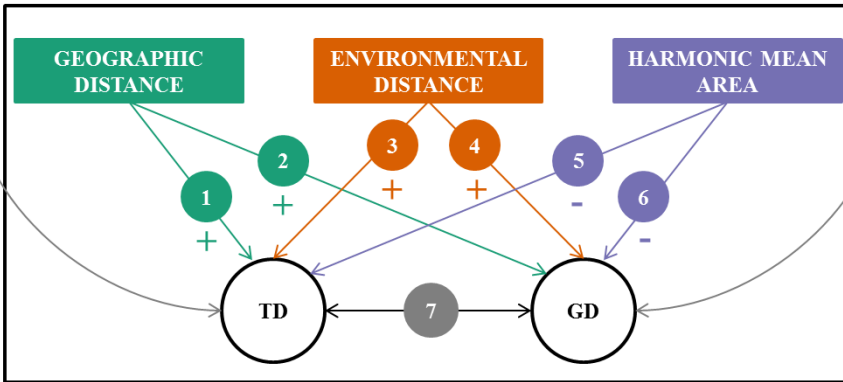
Contemporary period
= basins disconnected

1 Taxonomic differentiation is expected to increase with geographical distance because of **dispersal limitation**

2 Genetic differentiation is expected to increase with geographical distance because of **dispersal limitation**

3 Taxonomic differentiation is expected to increase with environmental distance because of **environmental selection**

4 Genetic differentiation is expected to increase with environmental distance because of **environmental selection**

5 Taxonomic differentiation is expected to increase with **drift** which decreases with harmonic mean area

6 Genetic differentiation is expected to increase with **drift** which decreases with harmonic mean area

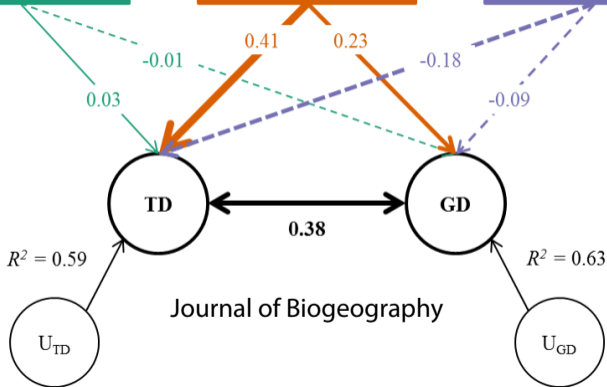7 Residual correlation due to unmeasured processes including those depending on the historical connectivity of basins

untypeset proof

(a) Multiple regressions on taxonomic differentiation

(b) Multiple regressions on genetic differentiation

untypeset proof

(a) Raw correlations

Rutilus rutilus

Mantel statistic r = 0.918, p-value = 0.006

(b) Residual correlations

Rutilus rutilus

Mantel statistic r = 0.724, p-value = 0.008

(c) Raw correlations

Squalius cephalus

Mantel statistic r = 0.103, p-value = 0.309

(d) Residual correlations

Squalius cephalus

Mantel statistic r = −0.004, p-value = 0.521