1   Content type: Article

2   **THE SYSTEMATIC CONSERVATION PLANNING FOR INTRASPECIFIC**

3   **GENETIC DIVERSITY.**

4   Ivan PAZ-VINAS[1,2,3*], Géraldine LOOT[1,4], Virgilio HERMOSO[6], Charlotte

5   VEYSSIERE[1,4], Nicolas POULET[7], Gaël GRENOUILLET[1,4] and Simon

6   BLANCHET[1,5]

7

8   [1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS),

9   École Nationale de Formation Agronomique (ENFA); UMR 5174 EDB (Laboratoire

10 Évolution & Diversité Biologique), 118 route de Narbonne, 31062 Toulouse cedex 4,

11 France.

12 [2] Aix-Marseille Université, CNRS, IRD, Avignon Université; UMR 7263 IMBE, Équipe

13 EGE, Centre Saint-Charles, Case 36, 3 place Victor Hugo, 13331 Marseille cedex 3,

14 France.

15 [3] Université de Lyon, CNRS, ENTPE; UMR 5023 LEHNA, Laboratoire d'Écologie des

16 Hydrosystèmes Naturels et Anthropisés, 6 rue Raphaël Dubois, 69622 Villeurbanne,

17 France.

18 [4] Université de Toulouse, UPS; UMR 5174 EDB, 118 route de Narbonne, 31062

19 Toulouse cedex 4, France.

20 [5] CNRS, Station d'Écologie Théorique et Expérimentale; UMR 5321, 09200 Moulis,

21 France.

22 [6] Centre Tecnologic Forestal de Catalunya, Crta. Sant Llorenc de Monunys, Km 2,

23 25280, Solsona. Lleida, Spain.

24 [7] Pôle Écohydraulique ONEMA-IMFT, Office National de l'Eau et des Milieux

25 Aquatiques, France.

26

27 **Short title:** Conservation planning of genetic diversity

28

29 **Keywords:** biodiversity, dendritic ecological networks, fish, freshwater, genetic

30 diversity, geostatistical tools, microsatellites, river basin, spatial patterns, systematic

31 conservation planning

32

33 ***Corresponding author:**

34
35 Ivan PAZ-VINAS
36 Université Lyon 1
37 CNRS, ENTPE, UMR 5023 LEHNA
38 Laboratoire d'Écologie des Hydrosystèmes Naturels et Anthropisés
39 3, rue Raphaël Dubois - Bât. Darwin C
40 69622 Villeurbanne Cedex France
41 Phone: (+33) 04 72 43 26 03
42 E-mail: ivanpaz23@gmail.com

43
44 **Co-author's e-mails:**

45
46 Simon Blanchet: simon.blanchet@sete.cnrs.fr
47 Virgilio Hermoso: virgilio.hermoso@gmail.com
48 Charlotte Veyssière: veyssiere.charlotte@gmail.com
49 Nicolas Poulet: nicolas.poulet@onema.fr
50 Gaël Grenouillet: gael.grenouillet@univ-tlse3.fr
51 Géraldine Loot: geraldine.loot@univ-tlse3.fr
52

53 **Abstract**
54

55   Intraspecific diversity is fundamental for species' adaptation to environmental

56 changes, and should hence be a main target for biodiversity conservation. However,

57 attempts to identify priority conservation areas for intraspecific diversity remain scarce.

58 Here, we used molecular data on six freshwater fish species sampled at a large spatial

59 scale, to determine hot- and cold-spots of genetic diversity, and to identify priority

60 conservation areas using a systematic conservation planning approach. We demonstrate

61 that the systematic conservation planning is an efficient and relevant approach for

62 preserving intraspecific diversity, although we identify weak congruencies and

63 surrogacies among conservation solutions found for each species. These weak

64 congruencies are due to among-species variation in the spatial distribution of hot-spots

65 of genetic diversity. We finally provide operational guidelines to efficiently use

66 systematic conservation planning methods with intraspecific genetic diversity data, and

67 to identify priority conservation areas for intraspecific diversity.

68

69

70   Biodiversity conservation is a major challenge that is often addressed by

71 identifying protected areas with high biodiversity and/or landscape values[1].

72 Conservation areas are generally identified as areas with high proportions of endemic,

73 rare or iconic species[2]. Alternatively, conservation planning can be based on the concept

74 of complementarity between conservation areas[3], and on cost-effectiveness analyses

75 such as systematic conservation planning methods (hereafter SCP[4]). SCP aims at

76 identifying a number of sites (i.e. *irreplaceable sites* that should be managed for

77 conservation in priority) best representing a predefined proportion of the biodiversity

78 observed in a region, at a minimum cost.

79   There have been attempts to include information on the phylogenetic history of

80 species assemblages into SCP approaches to preserve both species identities and their

81 macro-evolutionary history[5,6]. However, genetic diversity observed at the *population*

82   *level* (i.e. within species) has rarely been considered in SCP. *Intraspecific genetic*

83   *diversity* is a fundamental facet of biodiversity, as it is the fuel for species to adapt to

84   global and environmental changes[7–10]. Conservation geneticists have classically

85   considered this facet of biodiversity in conservation plans, for instance by identifying

86   "Evolutionary Significant Units" or unique genetic lineages[11]. However, conservation

87   geneticists have almost ignored the possibility to combine genetic data (e.g. allele

88   identities) with dedicated planning tools such as SPC[12].

89   The relative lack of genetic datasets at *large* spatial scales may partly explain

90   why SCP has yet rarely been applied to intraspecific genetic diversity[13,14]. Particularly,

91   conservation geneticists have been generally restricted by the amount and spatial range

92   of datasets. However, our capacity to compile genetic datasets at large spatial, temporal

93   and taxonomic scales has greatly increased in the last decades[15,16], so that it is now

94   possible to identify priority areas for the conservation of genetic diversity using

95   dedicated conservation planning tools.

96   Here, we tested the potential of SCP analyses to identify priority conservation

97   areas accounting for intraspecific genetic diversity measured at a large spatial scale. We

98   first considered a set of four common and representative freshwater fish species

99   (*Squalius cephalus*, *Gobio occitaniae*, *Barbatula barbatula* and *Phoxinus phoxinus*) to

100   test the influence of conservation targets (proportion of the total amount of genetic

101   diversity to be covered by irreplaceable sites) and analytical strategies (analysing each

102   species independently or all species pooled) on final conservation solutions (number

103   and identity of irreplaceable sites). We then included two rare species of particular

104   conservation interest (*Leuciscus burdigalensis* and *Parachondrostoma toxostoma*) to

105   test the relevance of the SCP approach in a "real conservation-oriented study". For these

106   two species (and the four common species), we ran SCP analyses considering a typical

107 conservation target[17,18] to (i) explore the spatial distribution of irreplaceable sites in a

108 riverscape, and (ii) test for congruency and surrogacy in irreplaceable sites among

109 species, and more particularly between rare and common species. We finally tested

110 whether –or not– irreplaceable sites were correctly predicted by classical indices of

111 genetic diversity (e.g. allelic richness). We demonstrate that preserving the genetic

112 diversity of a species assemblage is a feasible –yet complex– task necessitating

113 appropriate analyses to assist the decision-making process.

114

115 **Results**

116

117 *Descriptive statistics.* We assessed for the six species within-sites intraspecific genetic

118 diversity (i.e. α-IGD) by calculating both allelic richness (AR) and richness in private

119 alleles (PA). Overall, *P. toxostoma* (one of the two rare species) showed the lowest α-

120 IGD. Mean AR ranged from 2.114 for *P. toxostoma* to 5.821 for *P. phoxinus*, and mean

121 *PA* ranged from 0.036 for *P. toxostoma* to 0.162 for *L. burdigalensis* (Table S1). We

122 also assessed among-sites intraspecific genetic diversity (i.e. β-IGD) by quantifying (for

123 each species) how much a site is genetically unique compared to all others (using the

124 D_{est} index[19], see the Methods section). *Parachondrostoma toxostoma* also showed the

125 lowest mean D_{est} value (0.069), while the highest mean value was found for *B.*

126 *barbatula* (0.383; Table S1).

127

128 Testing the suitability of SCP analyses for intraspecific genetic diversity

129

130   In this first step we focused on the four common species for which genetic data

131   were available at a large spatial scale, allowing a thorough exploration of the suitability

132   of SCP for intraspecific genetic diversity.

133

134   *Spatial patterns of genetic diversity.* Using Generalized Linear Models for Stream

135   Networks (GLMSSN)[20,21], we showed that spatial patterns of genetic diversity largely

136   varied and were actually poorly congruent among the four common fish species (Figure

137   1). As extreme examples, hotspots of *AR* for *S. cephalus* where mainly found in the

138   Western part of the network and on the core streams, whereas these same areas were

139   identified as coldspots of *AR* for *B. barbatula* (Figure 1A1-1A3). Similarly, hotspots of

140   *PA* were inversely related between *G. occitaniae* and *P. phoxinus* (Figure 1B2-1B4).

141   Similar conclusions were reached for D_{est} (Figure 1C). For instance, hotspots of D_{est}

142   were observed in opposite areas of the river basin for the species pair *B. barbatula/P.*

143   *phoxinus* (Figure 1C3-1C4). As a consequence, the sign, slope and significance of

144   GLMSSNs explanatory variables strongly varied among species (Figure 1). This

145   incongruence in spatial patterns of genetic diversity among species was also reflected by

146   the low to moderate correlation coefficients measured among all possible pairs of

147   species and for each index of genetic diversity (i.e. *AR*, *PA* and D_{est}, Table S2). Indeed,

148   Pearson's correlation coefficients were lower than 0.6 for all comparisons but two (i.e.

149   between *B. barbatula*/*G. occitaniae* and between *P. phoxinus*/*B. barbatula* for *AR*, see

150   Table S2).

151

152   *The influence of conservation targets and analytical strategies to identify irreplaceable*

153   *sites for genetic conservation.* We used alleles' presence/absence data combined with

154   SCP procedures to test the influence of conservation targets and analytical strategies on

155   the number and identity of irreplaceable sites (i.e. sites that were selected in all the

156   solutions in SCP analyses; see the Methods section for more details).

157   When species were analysed independently, we found that the number of

158   irreplaceable sites increased as the conservation target (i.e. the percentage of alleles to

159   be covered by irreplaceable sites) increased, with a steep increase for conservation

160   targets higher than 75% of the total number of alleles present at the river basin scale

161   (Figure 2). However, the percentage of irreplaceable sites strongly varied among

162   species. For instance, for a 90% conservation target –which is a common target in

163   conservation genetics[17]–, the proportion of irreplaceable sites ranged from 3.61% of the

164   total number of sampled sites for *G. occitaniae* to 28.57% for *P. phoxinus* (Table S3;

165   Figure 2). For extreme conservation targets (100% of alleles to be covered), the

166   proportion of irreplaceable sites varied from 25.30% for *G. occitaniae* to 68.26% for *P.*

167   *phoxinus* (Table S3; Figure 2).

168   When alleles from the four common species were analysed in a single pooled

169   analysis, we similarly found that the proportion of irreplaceable sites increase as the

170   conservation target increases (Figure 2). Interestingly, we did not identify irreplaceable

171   sites for the 30% conservation goal, and only 3 irreplaceable sites were found for the

172   50% conservation goal (Figure 2). The proportion of irreplaceable sites increased

173   moderately to 17.39% for the 75% target, and then steeply increased for higher

174   conservation targets to reach 55.70% for the 90% target and 76.08% for the 100% target

175   (Figure 2). This later result suggests that almost all the river basin should be protected

176   to reach high conservation targets when adopting a pooled strategy.

177

178   A real conservation-oriented study using SCP approaches

179

180   We here focused on two rare species (in addition to the four common species) to

181   explore the usefulness of SCPs in a real case study.

182

183   *Identification of irreplaceable sites for genetic conservation.* We first visually explored

184   the spatial distribution of irreplaceable sites. Overall, the localization of irreplaceable

185   sites in the riverscape strongly varied among species, being spread all over the river

186   basin (Figure 3). We failed to identify areas (e.g. upstream or downstream locations)

187   clustering irreplaceable sites for any species (Figure 3A-F). This apparent lack of

188   clustering was statistically confirmed by our generalized linear models (Table S4).

189   Indeed, the two positional indices we used as explanatory variables (i.e. distance from

190   the outlet of sampling sites and the betweenness centrality of each sampling site, see the

191   Methods section) were not significant predictors of the irreplaceability of sites for all

192   species, except for distance from the outlet for *P. phoxinus* (Table S4). This indicates

193   that neither the position of sites in the riverscape, nor the positional importance of these

194   sites, determine the irreplaceability of sites.

195   Second, we tested whether conservation solutions found for each species were

196   spatially congruent among species. Over the six fish species, we identified forty-two

197   sites (out of the ninety-two sites, i.e. 45.65%) as irreplaceable at the 90% conservation

198   target for at least one species (Figure 4). Thirty-two of these forty-two sites were

199   irreplaceable for at least one of the four common species (Figure 4), and fourteen of the

200   forty-two sites were irreplaceable for at least one of the two rare species (Figure 4).

201   Among the six species, only eight out of these forty-two sites were irreplaceable for at

202   least two species, and only one of these sites was irreplaceable for three species (Figure

203   4).

204

205   *Surrogacy in irreplaceable sites among species.* We tested whether solutions found for

206   one species can be used as a surrogate for other species by calculating the percentage of

207   the total number of alleles observed for a given species that is covered by irreplaceable

208   sites identified for another species. Levels of surrogacy were generally low to moderate,

209   and strongly varied among species pairs (Table 1). For instance, irreplaceable sites

210   identified for *G. occitaniae* failed to cover the genetic diversity of the two rare species,

211   and covered only 17.66 to 42.60% of the total number of alleles of the other common

212   species. This indicates that irreplaceable sites identified for *G. occitaniae* (i.e. the most

213   widespread species) are poor surrogates for preserving the intraspecific genetic diversity

214   of other species (18.38% of surrogacy; Table 1). Conversely, irreplaceable sites found

215   for *P. phoxinus* and *L. burdigalensis* are better surrogates for *G. occitaniae*, as 79.58%

216   and 70.27% of the total number of alleles of *G. occitaniae* was covered by irreplaceable

217   sites identified for *P. phoxinus* and *L. burdigalensis* respectively (Table 1). Overall,

218   irreplaceable sites best covering genetic diversity of other species were those identified

219   for *B. barbatula*, which covered in average 68.28% of the total number of alleles of

220   other species (Table 1).

221   Interestingly, the thirty-two irreplaceable sites identified for the four common

222   species covered 79.87% and 90% of the total number of alleles of *L. burdigalensis* and

223   *P. toxostoma* respectively, suggesting that irreplaceable sites identified for a set of

224   common species can be good surrogates for intraspecific genetic diversity of rare

225   species. Conversely, the fourteen irreplaceable sites identified for the two rare species

226   covered a total number of alleles ranging from 65.18% for *S. cephalus* to 79.40% for *B.*

227   *barbatula*.

228

229   *Relationships between genetic indices and irreplaceable sites*. For all species but *G.*

230   *occitaniae* and *P. toxostoma* (for which none of the variables were significant

231   predictors), *PA* was identified as the only variable that significantly predicted the

232   probability for a site to be identified as an irreplaceable site (Table S5). This probability

233   increased with the number of *PA* in a site.

234

235   **Discussion**

236

237   Intraspecific diversity constitutes the fuel for species and populations to cope

238   with environmental changes[7–9]. This biodiversity facet is hence the first that should

239   respond to global change, allowing populations and species to respond adaptively to

240   these changes[10]. However, this biodiversity facet has so far been poorly integrated in

241   dedicated optimization planning tools. We here fill this gap by demonstrating that the

242   systematic conservation planning of intraspecific genetic diversity is a feasible –yet

243   complex– task necessitating careful considerations.

244

245   *From idiosyncratic distributions of genetic diversity…*

246   Our results strongly suggest that the genetic diversity of the targeted species did

247   not follow a common spatial pattern, but rather species-specific (idiosyncratic) spatial

248   distributions. This conclusion holds true for all genetic diversity indices, and it

249   corroborates the few previous studies investigating simultaneously the spatial

250   distribution of genetic diversity at large spatial scales and for sympatric species

251   (e.g.[22,23]). This was however unexpected given that recent meta-analyses on freshwater

252   organisms demonstrated that α-IGD is generally higher in downstream than in upstream

253   areas[24], and that β-IGD tends to be higher in upstream than in downstream sections[25].

254   Overall, these very general spatial patterns of intraspecific genetic diversity were

255   verified in our datasets. For instance, a negative relationship between allelic richness

256   and distance from the outlet is expected in freshwater organisms[24], and was actually

257   observed for three out of the six species considered in this study (Figure S1). However,

258   when using a precise and novel approach to map genetic diversity across the network,

259   we demonstrated that the distribution of cold- and hot-spots of α- and β-IGD was subtler

260   and idiosyncratic. This probably reflects interactions between colonization histories,

261   life-history-traits of species and the network structure, which are expected to drive

262   patterns of genetic diversity in rivers[24,26,27].

263

264   *...to the systematic conservation planning of intraspecific genetic diversity.*

265   The spatial mismatch in intraspecific genetic diversity among species probably

266   explains why we found that the level of congruency and surrogacy of irreplaceable sites

267   identified for each species was extremely low. For instance, there was an extremely low

268   proportion of irreplaceable sites that were common to two or three species (and never

269   more than three species; Figure 4). In the same way, we detected no clear patterns in the

270   spatial distribution of irreplaceable sites. In riverscapes, it is generally assumed that

271   small upstream areas are the "source" of genetic uniqueness, and hence the primary

272   areas to protect (i.e. the "small but mighty" paradigm[25]). Our results did not confirm this

273   paradigm since irreplaceable sites (for any of the six species) were not particularly

274   situated in upstream areas and/or in areas of high connectivity (i.e. areas displaying high

275   centrality values[28,29]), and rather suggests that priority areas for the conservation of

276   intraspecific genetic diversity should cover the whole distribution range of species.

277   Finally, the level of surrogacy among irreplaceable sites was low to moderate, and never

278   attained the 90% threshold we assumed when we considered all species

279   independently[17,18]. Combined with our finding that the number of irreplaceable sites can

280   be relatively high for reasonable conservation targets (up to 46% sites are identified as

281   irreplaceable sites for at least one species at the conservation target of 90%), we

282   concluded that our ability to identify priority areas for intraspecific genetic diversity is

283   highly species-specific and depends on the capacity to tackle the trade-off between the

284   amount of genetic diversity to protect, and the extent of priority areas we can

285   realistically protect.

286   However, when surrogacy between all irreplaceable sites identified for the entire

287   set of common species and those identified for the rare species was tested, we reached

288   reasonable proportions of the total number of alleles to protect (~80-90%) for the two

289   rare species (i.e. *P. toxostoma* and *L. burdigalensis*). This result suggests that, in some

290   cases, genetic data obtained for a set of widely-distributed, "easier-to-sample" common

291   species displaying varying life-history traits can be used for identifying protection areas

292   for the intraspecific genetic diversity of other sympatric rare species that can be more

293   problematic to sample.

294   Overall, our results suggest that two different analytical strategies can be

295   employed in real-case SCP studies aiming at preserving intraspecific genetic diversity

296   (i) identification of conservation areas for each rare species independently or (ii)

297   identification of conservation areas for a set of representative common species. Both

298   strategies have their own advantages and inconveniences. The first strategy optimally

299   preserves genetic diversity of rare species at competitive costs (e.g. 14 irreplaceable

300   sites to protect in our demonstrative study), but this at the expense of the genetic

301   diversity of other sympatric species. Conversely, the second strategy will optimally

302   preserve genetic diversity of a set of common species while maintaining high levels of

303   genetic diversity for rare species, but this at a higher cost (e.g. 32 irreplaceable sites to

304   protect in our study). Whether to choose one of these two strategies will therefore

305   depend on many factors such as how difficult is the sampling of rare species compared

306   to common species, or the extent of the resources available for setting new protected

307   areas. We recommend however to adopt the second strategy when possible, since it

308   allows to simultaneously maintain genetic diversity from rare and common species.

309   Indeed, genetic diversity of common species is vital for ensuring ecosystem stability, as

310   it ultimately influence species interactions, population dynamics and ecosystem

311   functions[30], and we argue that it should be considered in conservation plans.

312

313   **Conclusions**

314

315   Our study provides novel, insightful and promising knowledge on the setting of

316   priority conservation areas for intraspecific diversity. It shows that systematic

317   conservation planning methods are useful objective tools for conservation geneticists

318   whose conservation solutions will strikingly depend on the species to be preserved and

319   the quantity of genetic information that managers aim at preserving in a landscape.

320   Given our results, we suggest that two strategies could be employed in real-case

321   conservation programs: (i) identification of priority conservation areas for each rare

322   species independently or (ii) identification of priority conservation areas on the basis of

323   the analysis of a set of representative common species that may serve as "umbrellas" for

324   rare sympatric species.

325   Our study also raises many additional questions that should be considered in the

326   near future. Among others, we believe that the next steps will be to formally identify

327   sound conservation targets for intraspecific diversity, to test whether neutral

328   intraspecific diversity appropriately mirrors quantitative and adaptive diversity[9], and to

329   quantify the influence of intraspecific diversity on ecosystem functioning and services,

330   so as to better evaluate the added value of preserving such a facet of biodiversity[9,31].

331

332   **Methods**

333

334   Data collection

335

336   *Biological models.* We focused on an assemblage of six Cyprinid freshwater fish

337   species. Three of them are widespread in Europe (i.e. *Squalius cephalus*, *Phoxinus*

338   *phoxinus* and *Barbatula barbatula*) whereas three of them (*Gobio occitaniae*, *Leuciscus*

339   *burdigalensis* and *Parachondrostoma toxostoma*) are endemic to Southern France[32].

340   This set of species covers a large functional trait space that is representative of many

341   freshwater fish communities. For instance, *S. cephalus* is a large-bodied fish species

342   with long lifespan (i.e. it can be 60 cm long and live up to 15 years[32]) whereas at the

343   extreme *P. phoxinus* is a small-bodied species with shorter lifespan (i.e. it is less than 12

344   cm long and usually lives up to 4-5 years[32]). From an ecological perspective, *G.*

345   *occitaniae, P. toxostoma* and *B. barbatula* are bottom feeders, whereas *S. cephalus* and

346   *P. phoxinus* are water column feeders and *L. burdigalensis* is more opportunistic.

347   Further, *B. barbatula* is mainly active during night, while the other species are

348   particularly active during the day. Four of these species are relatively abundant (i.e. *S.*

349   *cephalus*, *G. occitaniae*, *P. phoxinus* and *B. barbatula*), although they greatly vary in

350   their ecological niche and hence their spatial occupancy in the river network (see Figure

351   S2 for maps representing the spatial distribution of sampling sites for each species,

352   which roughly corresponds to their spatial distribution in the Garonne-Dordogne river

353   basin). We will hereafter refer to this set of species as the "common" species. The two

354   other species are rare (*L. burdigalensis*) to very rare (*P. toxostoma;* see Figure S2) in the

355   Garonne-Dordogne river basin, and are of particular interest for conservation. *Leuciscus*

356   *burdigalensis* is a recently described species that is locally experiencing both

357   demographic and genetic bottlenecks in many populations[33,34]. *Parachondrostoma*

358   *toxostoma* is a vulnerable species[35] listed in the IUCN red list, in the Annex II of the

359   European Union Habitats Directive and in Appendix III of the Bern Convention[35].

360

361   *Sampling design.* During Spring/Summer 2010-2011, we used electric-fishing to sample

362   ninety-two sites distributed across thirty-five rivers from a large river basin, the

363   Garonne-Dordogne River basin (>100,000 km^2, South-Western France; Figure S2;

364   Table S6). Sampling sites were chosen to cover the whole distribution range of each

365   species at the riverscape scale, and to allow characterising spatial patterns of genetic

366   diversity for all these species. Up to 25 individuals *per* species *per* site were sampled

367   when possible. Not all species were present at all sampling sites (Figure S2; Table S6),

368   and some species were at a density that did not allow reaching the 25 individuals

369   threshold. In these cases, we captured as many individuals as possible. We anesthetized

370   each individual and then we collected and stored in 90% ethanol a fragment of their

371   pelvic fin. All individuals were released alive at their sampling location.

372

373   *Genotyping.* Genomic DNA was extracted using a salt-extraction protocol[36]. We used

374   multiplexed Polymerase Chain Reactions (PCRs) to co-amplify 8 to 15 microsatellite

375   loci depending on the species (8 for *G. occitaniae*, 9 for *B. barbatula*, 10 for *S. cephalus*

376   and *P. phoxinus*, 14 for *L. burdigalensis* and 15 for *P. toxostoma*). We used 5-20 ng of

377   genomic DNA and QIAGEN® Multiplex PCR Kits (Qiagen, Valencia, CA, USA) to

378   perform PCR amplifications. Details on loci, primer concentrations, PCR conditions and

379   multiplex sets can be found in Table S7. The genotyping was conducted on an ABI

380   PRISM™ 3730 Automated Capillary Sequencer (Applied Biosystems, Foster City, CA,

381   USA). The scoring of allele sizes was done using GENEMAPPER® v.4.0 (Applied

382   Biosystems).

383

384   *Genetic diversity assessment.* Given the good spatial resolution of the sampling obtained

385   for the four common species (i.e. *S. cephalus*, *G. occitaniae*, *P. phoxinus* and *B.*

386   *barbatula*), descriptive genetic analyses were conducted for sampling sites displaying a

387   minimum sample size of N=10 individuals for these species, so as to maximize

388   consistency on subsequent allelic frequency-based genetic analyses (see Figure S2 and

389   Table S6 for details on sample sizes). For the two rare species, for which the sampling

390   was more restricted (i.e. *L. burdigalensis* and *P. toxostoma*), genetic analyses were

391   conducted for sampling sites displaying a minimum sample size of N=6 individuals, so

392   as to maximize the number of sampling sites included in the SCP procedures. We then

393   determined for each of the six species the occurrence of null alleles and potential

394   scoring errors with the program MICROCHECKER 2.3[37]. We tested for departures

395   from Hardy-Weinberg (HW) equilibrium with the 'adegenet' R package v1.6-2[38]. The

396   program GENEPOP v4.0[39] was used to assess linkage disequilibrium among loci within

397   sites. We found significant deviations from HW for a few locus/population pairs for

398   each six species (see Appendix A1 and Supplementary File 1 for details and raw tables),

399   and significant linkage disequilibrium and homozygote excesses for only the four

400   common species (Appendix A1; Supplementary File 1). However, no clear patterns

401   were observed for any species across loci and populations for these deviations. Given

402   the small extent of these deviations and given the large spatial extent of the databases,

403   we assumed that they weakly affected our main findings (Appendix A1; Supplementary

404   File 1).

405   To assess within-sites intraspecific genetic diversity (i.e. α-IGD), we applied

406   rarefaction procedures implemented in ADZE v1.0[40] to calculate both allelic richness

407   (AR[41]) and private allelic richness (PA[42]) at the sampling site level (based on a

408   minimum of N=10 individuals for common species or N=6 individuals for rare species).

409   To assess the among-sites component of intraspecific genetic diversity (i.e. β-IGD), we

410   used the R package 'mmod'[43] to calculate –for each species– a pairwise genetic

411   differentiation index (i.e. D_{est}[19]). For each site (and species), we then derived the

412   averaged value of all pairwise D_{est} values estimated between one given site and all the

413   remaining sites, so as to obtain a single value *per* site.

414

415   <u>Testing the suitability of SCP analyses for intraspecific genetic diversity</u>

416

417   In the first step, we tested the influence of conservation targets and analytical

418   strategies on final conservation solutions. In this step, we focused specifically on data

419   from the four common species, as their large coverage of the sampling area is more

420   suited for the demonstrative exercise done in this step.

421

422   *Spatial patterns of genetic diversity.* We first used geostatistical modelling tools to

423   explore spatial patterns of α and β genetic diversity for the four common species at the

424   riverscape scale by predicting the distributions of AR, PA and D_{est} from the observed

425   empirical values using Generalized Linear Models for Spatial Stream Networks

426   (GLMSSN[20,21]). This was done using the 'STARS' toolset of ARCGIS v10.2 and the R

427   package 'SSN'[20,21]. We conducted a model selection procedure based on a comparison

428   of Akaike Information Criteria (AIC) estimated for several competing GLMSSNs.

429   These models were built by (i) assuming three geographic descriptors (i.e. topological

430   distance from the outlet, longitude and latitude) as explanatory variables, and (ii)

431   choosing a tail-down covariance structure model among the following ones:

432   exponential, Mariah, spherical, linear-with-sill and Epanechnikov. As the number of

433   explanatory variables differed among the GLMSSNs we built, we used the maximum

434   likelihood estimation method for each GLMSSN, so as to allow AIC-based model

435   comparisons. For each common species and genetic index, the best model had the

436   lowest AIC score (see Supplementary File 1 for raw results). This model was used to

437   estimate the slope and the significance of the relationships between explanatory

438   variables and each genetic index, so as to test whether or not spatial patterns of

439   intraspecific genetic diversity can be detected. We finally used predictions from the best

440   models to produce krigged maps for each common species and each genetic index, so as

441   to visually represent the spatial distribution of intraspecific genetic diversity across the

442   whole river drainage, and to visually highlight hot- and cold-spots of intraspecific

443   genetic diversity. We also calculated Pearson's correlation coefficients between values

444   calculated at the site level for each genetic index (i.e. *AR*, *PA* and D_{est}) for each pair of

445   common species, so as to test for spatial congruency in patterns of genetic diversity

446   among common species.

447

448   *Identification of irreplaceable sites.* We then tested whether conservation targets (i.e.

449   the percentage of total number of conservation units to be present in the final

450   conservation solution) and analytical approaches (i.e. species-specific or species-pooled

451   analyses) influence the identification of irreplaceable sites. SCP methods traditionally

452   use species presence/absence data as input data to identify irreplaceable sites for the

453 conservation of taxonomic diversity at the community level (i.e. sites that cannot be

454 excluded from an optimal selection of sites for conservation)[44]. Here, we replaced

455 species presence/absence data by alleles' presence-absence data to identify irreplaceable

456 sites for the conservation of intraspecific genetic diversity of each species in the river

457 drainage. We used the program Marxan v2.1[44] and genetic data from the common

458 species to identify, for each common species independently, an optimal set of sites that

459 best represent at least 50, 75, 90 or 100% of the total number of alleles present in the

460 whole riverscape at a minimum "cost", which corresponds to four conservation targets.

461 Given the lack of ground estimates for conservation cost, we used a constant cost *per*

462 site, so our objective translated into identifying the minimum number of sites that

463 represent a given proportion of intraspecific genetic diversity (i.e. 50, 75, 90 or 100% of

464 the total pool of alleles represented at least once[45]). We arbitrarily choose the 50, 75 and

465 100% conservation targets to explore how the proportion of alleles to protect affects the

466 selection of irreplaceable sites. We additionally tested the 90% conservation target, as it

467 corresponds to a threshold target being typically assumed in *ex-situ* conservation

468 plans[17,18]. In order to estimate the relevance of each site to preserve a given proportion

469 of the allelic diversity in the river basin, we used two different methods. In case of

470 100% of allelic diversity, we used the traditional irreplaceability measure reported by

471 Marxan, which ranges between 100% (highly irreplaceable) and 0% (not irreplaceable).

472 This measure is estimated by running the optimization algorithm a number of times

473 (N=100 runs in our case) and then computing the frequency of selection of each site

474 within the solutions obtained. Sites with unique allelic composition will be selected

475 across all runs and reported as highly irreplaceable, whereas sites with more common

476 alleles, replaceable by other sites with the same alleles, will appear poorly irreplaceable.

477 For the other conservation targets, we selected a random pull of 50, 75 and 90% of the

478   total number of alleles existing at the basin level and for each species. We then ran the

479   optimization algorithm to identify the minimum set of sites representing this particular

480   selection of alleles. Given that the analyses were run only on a subset of alleles, we

481   replicated this process 100 times to minimize the effect of the arbitrary selection of

482   alleles. For each subset of alleles, we ran the Marxan procedure as explained above (e.g.

483   constant cost and 100 runs for each) and retained the best solution for subsequent

484   analyses (solution with the lowest value for the objective function). We then calculated

485   the irreplaceability as the frequency of selection of each site within the 100 random pull

486   of alleles. For a given species and a given conservation target, we considered a site as

487   irreplaceable for genetic diversity conservation when it displayed an irreplaceability

488   value of 100%. We selected such a high threshold so as to be conservative. We tested

489   and compared visually how the proportions of irreplaceable sites vary among species

490   and conservation targets.

491   To test how pooling data from several species affect the identification of

492   irreplaceable sites, we further performed a "pooled" analysis, in which all alleles found

493   for each common species at a site were pooled together in a single input dataset. We

494   then selected a random pull of 30, 50, 75, 90 and 100% of the total number of alleles

495   existing at the basin level (all common species confounded), and performed 100 Marxan

496   runs *per* conservation targets to identify the minimum set of sites representing these

497   particular selections of alleles.

498

499   <u>A real conservation-oriented study using SCP approaches</u>

500

501   In this second step, we (i) explored the spatial distribution of irreplaceable sites,

502   (ii) tested whether conservation solutions found for each species are congruent among

503 species, (iii) tested whether solutions found for one species can be used as a surrogate

504 for other species, and (iv) tested whether –or not– irreplaceable sites are correctly

505 predicted by classical indices of genetic diversity. In this step, we included the two rare

506 species, since congruency and surrogacy are particularly important to measure for rare

507 species for which data are more difficult to collect. We therefore focus more specifically

508 on the comparisons implying common *vs.* rare species.

509

510 *Identification of irreplaceable sites for genetic conservation.* We focused only on

511 irreplaceable sites identified for the 90% target and used the program Marxan as

512 described above to identify these sites for *L. burdigalensis* and *P. toxostoma*

513 independently, in addition of the four common species.

514   We mapped these irreplaceable sites (for the six species pooled or independently)

515 on the river network, so as to test (i) whether or not specific areas harboured more

516 irreplaceable sites (e.g. upstream areas that are generally thought to be of high

517 conservation priority[25]) and (ii) if irreplaceable sites are spatially congruent among

518 species and, most notably, among common and rare species. In addition, we ran GLMs

519 (assuming a binomial error terms distribution) including whether or not a site has been

520 designated as an irreplaceable site at the 90% target as a binomial dependent variable,

521 and distance to the outlet of sampling sites and betweenness centrality values[46,47] for

522 each sites explanatory variables. Betweenness centrality is an index quantifying the

523 positional importance of each sampling site within the river basin[28,29]. The significance

524 of each term was tested at the $\alpha=0.05$ threshold.

525

526 *Surrogacy in irreplaceable sites among species.* We then estimated the levels of

527 surrogacy among species by calculating the percentage of the total number of alleles

528   observed for a given species that is covered by irreplaceable sites identified for another

529   species. Although surrogacy was calculated for all species pairs, we specifically focused

530   on rare species by calculating (i) the percentage of the total number of alleles observed

531   for rare species covered by all the irreplaceable sites identified for all the common

532   species, and (ii) the percentage of the total number of alleles observed for common

533   species covered by all the irreplaceable sites identified for the rare species.

534

535   *Relationships between irreplaceable sites and indices of genetic diversity*. To test the

536   ability of classical genetic indices to predict the propensity for a site to be irreplaceable

537   from a conservation viewpoint, we ran for each species Generalized Linear Models

538   (GLMs, with a binomial error terms distribution) including whether or not a site has

539   been designated as an irreplaceable site at the 90% target as a binomial dependent

540   variable, and AR, PA and D_{est} as explanatory variables. We tested the significance of

541   each term at the α=0.05 threshold. Explanatory variables were centred and scaled, in

542   order to compare the relative strength of the predictors among species.

543

544   **Acknowledgements**

545

564

565   **References**

566

567   1. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. The performance and

568   potential of protected areas. *Nature* **515,** 67–73 (2014).

569   2. Filipe, A. F. *et al.* Selection of priority areas for fish conservation in Guadiana River

570   Basin, Iberian Peninsula. *Conserv. Biol.* **18,** 189–200 (2004).

571   3. Kirkpatrick, J. B. An iterative method for establishing priorities for the selection of

572   nature reserves: An example from Tasmania. *Biol. Conserv.* **25,** 127–134 (1983).

573   4. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405,**

574   243–253 (2000).

575   5. Asmyhr, M. G., Linke, S., Hose, G. & Nipperess, D. A. Systematic conservation

576   planning for groundwater ecosystems using phylogenetic diversity. *PLoS ONE* **9,**

577   e115132 (2014).

578   6.   Buerki, S. *et al.* Incorporating evolutionary history into conservation planning in

579   biodiversity hotspots. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** 20140014–20140014

580   (2015).

581   7.   Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation.

582   *Nature* **470,** 479–485 (2011).

583   8.   Carroll, S. P. *et al.* Applying evolutionary biology to address global challenges.

584   *Science* **346,** 1245993–1245993 (2014).

585   9.   Mittell, E. A., Nakagawa, S. & Hadfield, J. D. Are molecular markers useful

586   predictors of adaptive potential? *Ecol. Lett.* **18,** 772–778 (2015).

587   10. Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. Adaptation to Global

588   Change: A Transposable Element–Epigenetics Perspective. *Trends Ecol. Evol.* **31,**

589   514–526 (2016).

590   11. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure

591   using multilocus genotype data. *Genetics* **155,** 945–959 (2000).

592   12. Diniz-Filho, J. A. F. *et al.* Planning for optimal conservation of geographical

593   genetic variability within species. *Conserv. Genet.* **13,** 1085–1093 (2012).

594   13. Grantham, H. S., Pressey, R. L., Wells, J. A. & Beattie, A. J. Effectiveness of

595   Biodiversity Surrogates for Conservation Planning: Different Measures of

596   Effectiveness Generate a Kaleidoscope of Variation. *PLoS ONE* **5,** e11430 (2010).

597   14. Hermoso, V., Januchowski-Hartley, S. R. & Pressey, R. L. When the suit does not

598   fit biodiversity: Loose surrogates compromise the achievement of conservation

599   goals. *Biol. Conserv.* **159,** 197–205 (2013).

600   15. Andrew, R. L. *et al.* A road map for molecular ecology. *Mol. Ecol.* **22,** 2605–2626

601   (2013).

602   16. Pauls, S. U. *et al.* Integrating molecular tools into freshwater ecology: developments

603   and opportunities. *Freshw. Biol.* **59,** 1559–1576 (2014).

604   17. Frankham, R., Briscoe, D. A. & Ballou, J. D. *Introduction to conservation genetics.*

605   (Cambridge University Press, 2002).

606   18. Neel, M. C. & Cummings, M. P. Effectiveness of conservation targets in capturing

607   genetic diversity. *Conserv. Biol.* **17,** 219–229 (2003).

608   19. Jost, L. G(ST) and its relatives do not measure differentiation. *Mol. Ecol.* **17,** 4015–

609   4026 (2008).

610   20. Peterson, E. & Ver Hoef, J. M. V. STARS: STARS: An ArcGIS toolset used to

611   calculate the spatial information needed to fit spatial statistical models to stream

612   network data. *J. Stat. Softw.* **56,** 1–17 (2014).

613   21. Ver Hoef, J., Peterson, E., Clifford, D. & Shah, R. SSN: An R Package for Spatial

614   Statistical Modeling on Stream Networks. *J. Stat. Softw.* **56,** 1–43 (2014).

615   22. Fortuna, M. A., Albaladejo, R. G., Fernandez, L., Aparicio, A. & Bascompte, J.

616   Networks of spatial genetic variation across species. *Proc. Natl. Acad. Sci.* **106,**

617   19044–19049 (2009).

618   23. Taberlet, P. *et al.* Genetic diversity in widespread species is not congruent with

619   species richness in alpine plant communities. *Ecol. Lett.* **15,** 1439–1448 (2012).

620   24. Paz-Vinas, I., Loot, G., Stevens, V. M. & Blanchet, S. Evolutionary processes

621   driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Mol.*

622   *Ecol.* **24,** 4586–4604 (2015).

623   25. Finn, D. S., Bonada, N., Múrria, C. & Hughes, J. M. Small but mighty: headwaters

624   are vital to stream network biodiversity at two levels of organization. *J. North Am.*

625   *Benthol. Soc.* **30,** 963–980 (2011).

626   26. Paz-Vinas, I. & Blanchet, S. Dendritic connectivity shapes spatial patterns of

627   genetic diversity: a simulation-based study. *J. Evol. Biol.* **28,** 986–994 (2015).

628   27. Thomaz, A. T., Christie, M. R. & Knowles, L. L. The architecture of river networks

629   can drive the evolutionary dynamics of aquatic populations. *Evolution* **70,** 731–739

630   (2016).

631   28. Altermatt, F. Diversity in riverine metacommunities: a network perspective. *Aquat.*

632   *Ecol.* **47,** 365–377 (2013).

633   29. Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J. G. & Blanchet, S. Lessons from the

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

635   genetic diversity correlations. *Freshw. Biol.* (2016). doi:10.1111/fwb.12826

636   30. Mimura, M. *et al.* Understanding and monitoring the consequences of human

637   impacts on intraspecific variation. *Evol. Appl.* **10,** 121–139 (2017).

638   31. Siefert, A. *et al.* A global meta-analysis of the relative extent of intraspecific trait

639   variation in plant communities. *Ecol. Lett.* **18,** 1406–1419 (2015).

640   32. Kottelat, M. & Freyhof, J. *Handbook of European freshwater fishes.* (Publications

641   Kottelat, 2007).

642   33. Poulet, N., Beaulaton, L. & Dembski, S. Time trends in fish populations in

643   metropolitan France: insights from national monitoring data. *J. Fish Biol.* **79,** 1436–

644   1452 (2011).

645   34. Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G. & Blanchet, S. The demographic

646   history of populations experiencing asymmetric gene flow: combining simulated

647   and empirical data. *Mol. Ecol.* **22,** 3279–3291 (2013).

648   35. Paz-Vinas, I. *et al.* Combining genetic and demographic data for prioritizing

649   conservation actions: insights from a threatened fish species. *Ecol. Evol.* **3,** 2696–

650   2710 (2013).

651   36. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality

652   genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25,** 4692–4693

653   (1997).

654   37. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. micro-

655   checker: software for identifying and correcting genotyping errors in microsatellite

656   data. *Mol. Ecol. Notes* **4,** 535–538 (2004).

657   38. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers.

658   *Bioinformatics* **24,** 1403–1405 (2008).

659   39. Rousset, F. genepop'007: a complete re-implementation of the genepop software for

660   Windows and Linux. *Mol. Ecol. Resour.* **8,** 103–106 (2008).

661   40. Szpiech, Z. A., Jakobsson, M. & Rosenberg, N. A. ADZE: a rarefaction approach

662   for counting alleles private to combinations of populations. *Bioinformatics* **24,**

663   2498–2504 (2008).

664   41. Petit, R. J., El Mousadik, A. & Pons, O. Identifying Populations for Conservation

665   on the Basis of Genetic Markers. *Conserv. Biol.* **12,** 844–855 (1998).

666   42. Kalinowski, S. T. Counting alleles with rarefaction: Private alleles and hierarchical

667   sampling designs. *Conserv. Genet.* **5,** 539–543 (2004).

668   43. Winter, D. J. mmod: an R library for the calculation of population differentiation

669   statistics. *Mol. Ecol. Resour.* **12,** 1158–1160 (2012).

670   44. Ball, I. R., Possingham, H. P. & Watts, M. in *Spatial conservation prioritisation:*

671   *Quantitative methods and computational tools* (eds. Moilanen, A., Wilson, K. A. &

672   Possingham, H. P.) **Chapter 14,** 185–195 (Oxford Iniversity Press, Oxford, UK,

673   2009).

674   45. Hermoso, V. & Kennard, M. J. Uncertainty in coarse conservation assessments

675   hinders the efficient achievement of conservation goals. *Biol. Conserv.* **147,** 52–59

676   (2012).

677   46. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry*

678   **40,** 35 (1977).

679   47. Estrada, E. & Bodin, Ö. Using network centrality measures to manage landscape

680   connectivity. *Ecol. Appl.* **18,** 1810–1825 (2008).

681

682   **Author Contributions:** The study was designed by SB, GL and IPV. IPV and SB

683   wrote the manuscript with the help of VH, GG, NP and GL. IPV, SB, GL and CV

684   collected the samples. GL and CV produced the genetic data. IPV, VH and SB

685   conducted the population genetic and statistical analyses.

686

687   **Competing Financial Interests statement:** Authors declare not having any competing

688   financial interest.

689

690 **Figures legends**

691

692 **Figure 1**

693 Spatial distribution of observed (coloured circles) and interpolated (coloured lines)

694 values of AR, PA and D_{est} (A, B, C respectively in the figure) for *Squalius cephalus*,

695 *Gobio occitaniae*, *Barbatula barbatula,* and *Phoxinus phoxinus* (1, 2, 3, 4 respectively

696 in the figure) obtained with GLMSSNs. The width of coloured lines is inversely related

697 to the prediction standard error. The cursor on the vertical coloured scale indicates the

698 mean value of AR, PA and D_{est}. The slope (β) of each explanatory variable (i.e.

699 topological distance to the outlet, longitude and latitude) and its significance is also

700 reported. N.I. indicates that the explanatory variable has not been included in the model;

701 * p-value <0.05; ** p-value <0.01; *** p-value <0.001.

702

703 **Figure 2**

704 Percentage of irreplaceable sites identified by Marxan for conservation targets of 50, 75,

705 90 and 100% of the total number of alleles present in the river basin for each common

706 species and for a pooled analysis in which all alleles from all common species were

707 pooled together. N_{SITES} represents the number of sites included in the Marxan analyses.
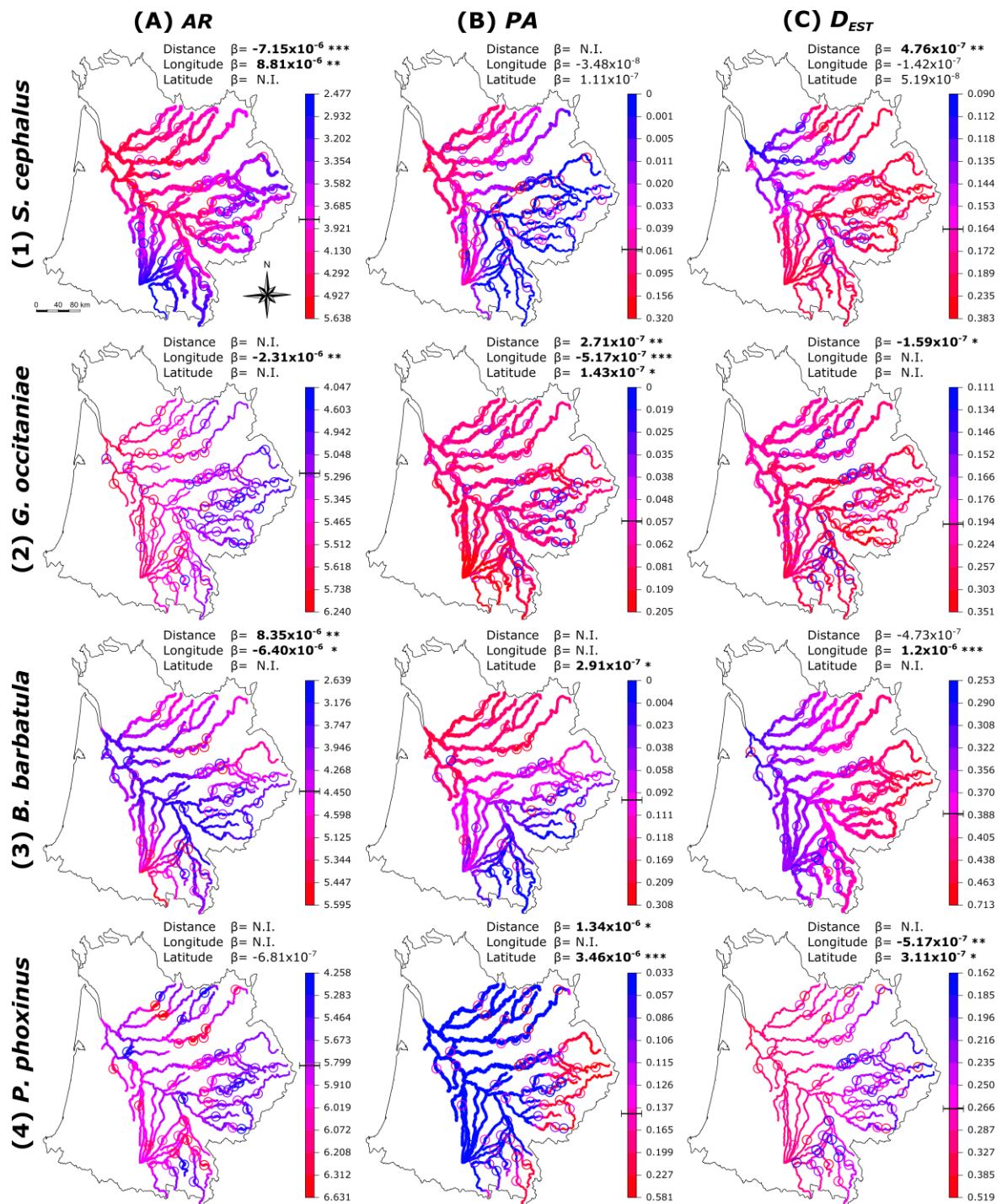
708

709 **Figure 3**

710 Irreplaceable sites that have been identified for each species by Marxan for preserving

711 90% of the total number of alleles present in the river basin (red-filled circles).

712

713 **Figure 4**

714   Irreplaceable sites that have been identified by Marxan for at least one (unicoloured-

715   filled points), two (black dotted circles surrounding bicoloured points) or three (black

716   bolded circle surrounding a tricoloured point) species, assuming a conservation target of

717   90% of the total number of alleles present in the river basin when considering the six
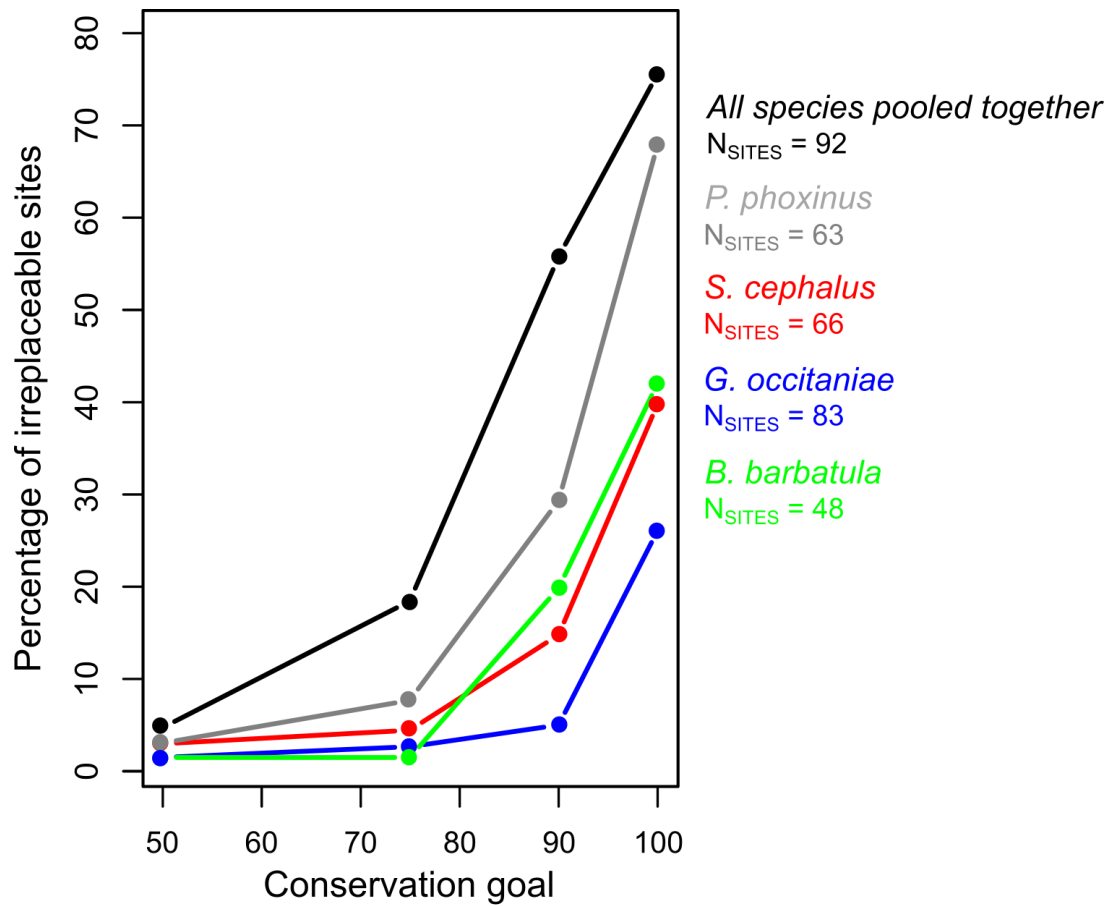
718   species.

719

720 **FIGURE 1**

721

722

723   **FIGURE 2**

All species pooled together
N_{SITES} = 92

P. phoxinus
N_{SITES} = 63

S. cephalus
N_{SITES} = 66

G. occitaniae
N_{SITES} = 83

B. barbatula
N_{SITES} = 48

724   (Percentage of the total number of alleles to protect)

725

726 **FIGURE 3**

727

FIGURE 4

Legend:
- ○ Non-irreplaceable site
- ● *S. cephalus* (7)
- ● *G. occitaniae* (3)
- ● *B. barbatula* (8)
- ● *P. phoxinus* (18)
- ● *P. toxostoma* (3)
- ● *L. burdigalensis* (11)

Irreplaceable site (for at least one species)

- ◌ Irreplaceable site (for at least two species)
- ◯ Irreplaceable site (for at least three species)

0 40 80 km

729
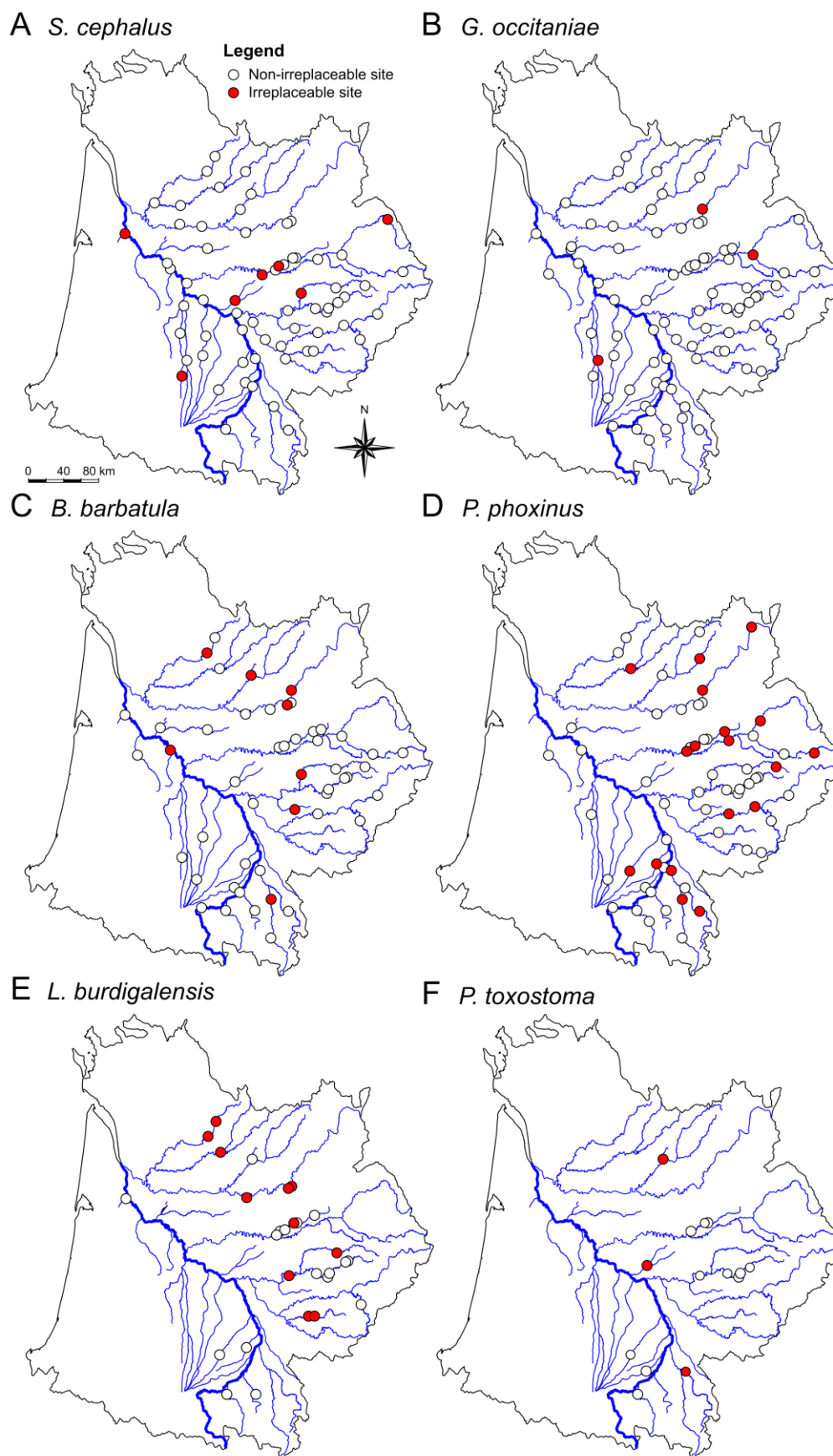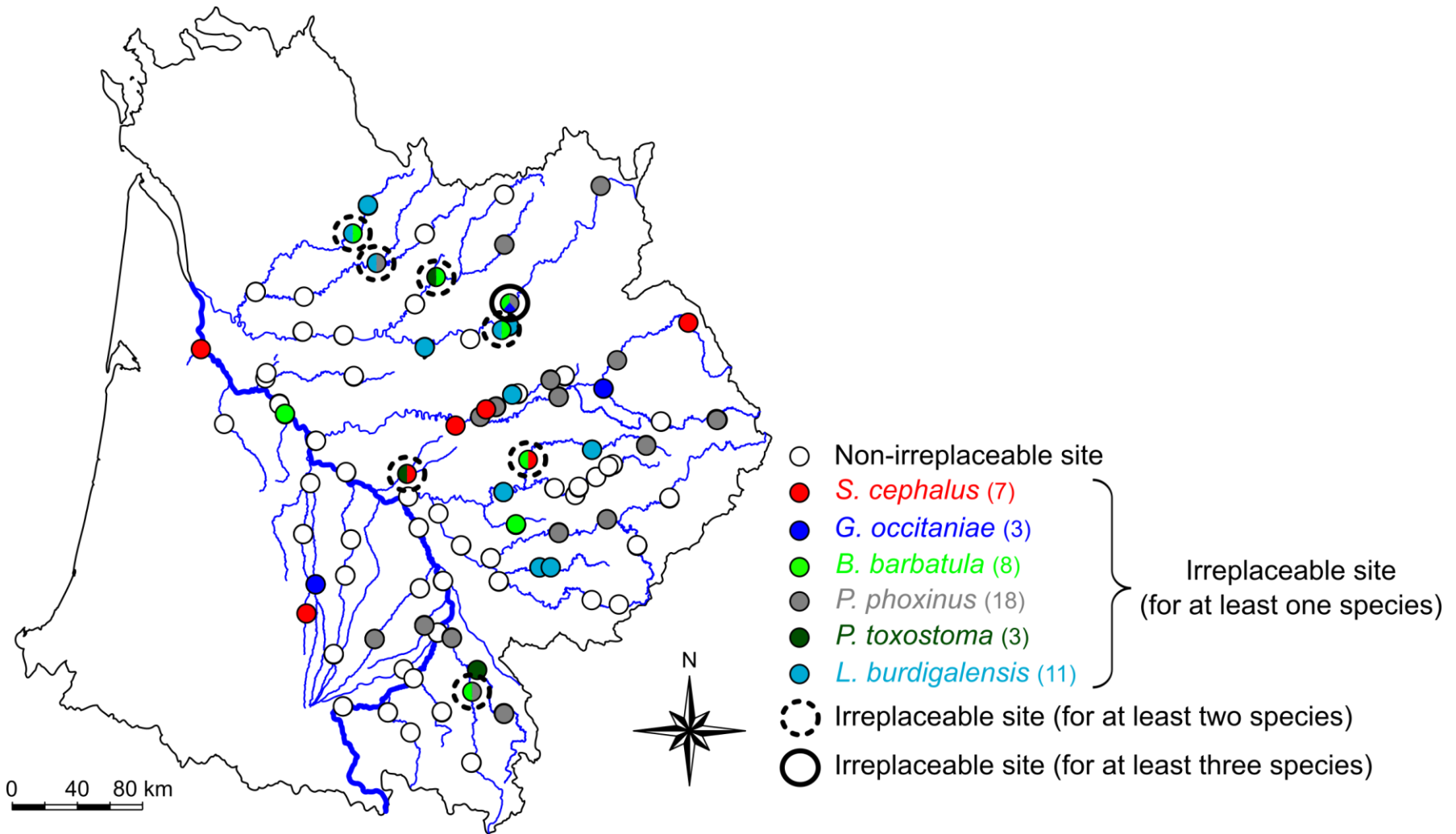
730 **TABLE 1:** Table reporting for each species the percentages of their total number of alleles that are covered by the sampling sites having

731 been identified as irreplaceable sites for the other species (considering the 90% conservation target).

732

Proportion of alleles from (below) that fell within irreplaceable sites from (right):	*Squalius cephalus*	*Gobio occitaniae*	*Barbatula barbatula*	*Phoxinus phoxinus*	*Leuciscus burdigalensis*	*Parachondrostoma toxostoma*	Common species	Rare species
Squalius cephalus	*90.00*	17.66	61.52	56.52	61.60	39.73	—	65.18
Gobio occitaniae	65.60	*90.00*	66.76	79.58	70.27	50.97	—	73.40
Barbatula barbatula	49.79	31.65	*90.00*	79.57	76.78	38.95	—	79.40
Phoxinus phoxinus	48.25	42.60	68.43	*90.00*	63.49	44.61	—	68.88
Leuciscus burdigalensis	44.73	0	66.13	67.09	*90.00*	36.42	79.87	—
Parachondrostoma toxostoma	78.55	0	78.55	50.00	67.14	*90.00*	90.00	—
Average over all other species	*55.58*	*18.38*	*68.28*	*66.55*	*67.86*	*42.14*		

733