1   **Riverscape community genomics: A comparative analytical**

2   **approach to identify common drivers of spatial structure**

3   Running title: Riverscape Community Genomics

4

5   Zachery D. Zbinden[1*], Marlis R. Douglas[1], Tyler K. Chafin[1,2], Michael E. Douglas[1]

6

7   [1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA

8   [2]Biomathematics and Statistics Scotland, Edinburgh, Scotland, UK

9   [*]Corresponding Author: zdzbinde@uark.edu

10 **ABSTRACT**

11 Genetic differentiation among local groups of individuals, i.e., genetic β-diversity, is a key component of

12 population persistence related to connectivity and isolation. However, most genetic investigations of

13 natural populations focus on a single species, overlooking opportunities for multispecies conservation

14 plans to benefit entire communities in an ecosystem. We present an approach to evaluate genetic β-

15 diversity within and among many species and demonstrate how this *riverscape community genomics*

16 approach can be applied to identify common drivers of genetic structure. Our study evaluated genetic β-

17 diversity in 31 co-distributed native stream fishes sampled from 75 sites across the White River Basin

18 (Ozarks, USA) using SNP genotyping (ddRAD). Despite variance among species in the degree of genetic

19 divergence, general spatial patterns were identified corresponding to river network architecture. Most

20 species (N=24) were partitioned into discrete sub-populations (K=2–7). We used partial redundancy

21 analysis to compare species-specific genetic β-diversity across four models of genetic structure: Isolation

22 by distance (IBD), isolation by barrier (IBB), isolation by stream hierarchy (IBH), and isolation by

23 environment (IBE). A significant proportion of intraspecific genetic variation was explained by IBH

24 ($\bar{x}$=62%), with the remaining models generally redundant. We found evidence for consistent spatial

25 modularity in that gene flow is higher within rather than between hierarchical units (i.e., catchments,

26 watersheds, basins), supporting the generalization of the Stream Hierarchy Model. We discuss our

27 conclusions regarding conservation and management and identify the 8-digit Hydrologic Unit (HUC) as

28 the most relevant spatial scale for managing genetic diversity across riverine networks.

29

30 **KEYWORDS:**

31 Comparative population genetics; Hydrologic units; Population structure; RADseq; Spatial modularity;

32 Stream hierarchy model

33 **1 | INTRODUCTION**

34 Genetic diversity is a biodiversity measure that may be quantified across geography and through time

35 (Huber et al., 2010; Leonard et al., 2017). It is tied to species' past and future evolutionary trajectories

36 (Shelley et al., 2021). Genetic diversity is a barometer for population-level persistence in accurately

37 reflecting demography, connectivity, and adaptive potential (Davis et al., 2018; DeWoody et al., 2021;

38 Paz-Vinas et al., 2018). But genetic diversity is often underutilized in conservation planning (Laikre,

39 2010; Paz-Vinas et al., 2018), in part due to a suite of affiliated necessities (i.e., specialized equipment,

40 technical expertise), all of which expand its bottom line (Blanchet et al., 2020). Moreover, when

41 assessment does occur, it is most often limited to populations within a single species or a small cadre of

42 entities within a species-group, thus minimizing the potential for much-needed generalizations

43 (Anthonysamy et al., 2018).

44 Therefore, when the concept of genetic diversity is applied in a comparative sense across co-

45 distributed species, it provides a solid framework from which community-wide management and policy

46 can be defined (Hanson et al., 2020). Systematic conservation planning (SCP; Margules & Pressey,

47 2000), based on managing complementary sites within a region containing unique biodiversity, could

48 benefit from focusing on intraspecific genetic diversity measured across community members (Paz-Vinas

49 et al., 2018; Xuereb et al., 2021). For example, multispecies assessments can reveal common dispersal

50 barriers (Pilger et al., 2017; Roberts et al., 2013), congruent distributions of genetic diversity (Hotaling et

51 al., 2019; Ruzich et al., 2019), relevant spatial scales for management (Blanchet et al., 2020), and

52 associations among species characteristics and genetic diversity (Bohonak, 1999; De Kort et al., 2021;

53 Pearson et al., 2014). Despite the potential complexity, a comprehensive and systematic management

54 strategy can emerge, one more appropriately aligned towards managing numerous species, with long-term

55 conservation goals beneficial to entire communities (Blanchet et al., 2017). In addition, it also tacitly

56 encourages support by stakeholders for an overarching management plan, one representing a consensus

57 across multiple species and ecosystems (Douglas et al., 2020).

2

58   The spatial structure—or pattern—of genetic variation within a species is primarily dictated by

59   the interplay between gene flow and genetic drift (Holderegger et al., 2006; Hutchinson & Templeton,

60   1999). Different observable patterns of spatial genetic structure (i.e., genetic β-diversity) are used to infer

61   the influence of different underlying processes (Orsini et al., 2013). Spatial uniformity of genetic diversity

62   (i.e., panmixia; Rosenberg et al., 2005) is the implicit null model of population structure indicative of the

63   differentiating effects of genetic drift being overwhelmed by homogenizing gene flow. The *de facto*

64   alternative is spatially continuous genetic divergence driven by an equilibrium between gene flow and

65   drift occurring within stable, dispersal-limited populations (i.e., isolation by distance, IBD; Wright, 1943).

66   For most species, a significant relationship between genetic dissimilarity and geographic distance is

67   expected (Meirmans, 2012), yet the strength of this association may vary due to the intrinsic

68   characteristics of a species (Bohonak, 1999; Singhal et al., 2018) or the extrinsic factors experienced by a

69   species (environmental or historical) that affect dispersal (gene flow) or effective population size (genetic

70   drift) (Orsini et al., 2013; Paz-Vinas & Blanchet, 2015). Additional models to explain spatial structure

71   have been introduced to explore other processes. For example, genetic divergence may be further

72   promoted by environmental dissimilarities *across* sites that promote local adaptation or limited/biased

73   dispersal (isolation by environment, IBE; Bradburd et al., 2013; Wang & Bradburd, 2014). Dispersal

74   resistance induced by physical and environmental characteristics in *between* sites (isolation by resistance,

75   IBR; McRae, 2006) or barriers to dispersal (i.e., isolation by barrier, IBB; Cushman et al., 2006; Ruiz-

76   Gonzalez et al., 2015) can also amplify the amount of genetic dissimilarity observed over a given

77   geographic distance.

78   For aquatic biodiversity, patterns of genetic divergence will also be governed by the structure and

79   architecture of the riverine network (in contemporary and past representations). Organisms within such

80   dendritic networks are demonstrably impacted by the physical structure of the habitat, which constrains

81   their movement (Peterson et al., 2013; White et al., 2020; Paz-Vinas et al., 2015) and leads to

82   correspondence between genetic relatedness and the underlying structural hierarchy (Hughes et al., 2009).

83   While this is most apparent within the contemporary structure of river networks, their historic structure,

84   i.e., paleohydrology, also serves to bookmark genetic diversity, and the effects of past connectivity or

85   isolation can still be observed within contemporary spatial patterns (Mayden, 1988; Strange & Burr,

86   1997). Moreover, the hierarchical complexity of these networks will likewise dictate population processes

87   related to colonization/extinction and effective sizes, as reflected within genetic diversities and

88   divergences (Chiu et al., 2020; Hopken et al., 2013; Thomaz et al., 2016). Thus, spatial genetic structuring

89   within riverine biodiversity should reflect isolation by stream hierarchy (IBH; *sensu* Stream Hierarchy

90   Model (SHM); Meffe & Vrijenhoek, 1988). The initial genesis for the SHM was narrowly defined within

91   desert stream fishes of the American West (Meffe and Vrijenhoek, 1988). The model is an example of the

92   more general principle of 'spatial modularity,' which occurs when certain sets of habitat patches are more

93   tightly connected through individual movement than they are to others (Fortuna et al., 2009). Spatial

94   modularity can reveal fundamental scales common among populations and possibly species, which can

95   inform conservation strategies (Fletcher et al., 2013). Therefore, an assessment of the SHM's generality,

96   as compared to alternative isolating regimes, was thus imperative (Brauer et al., 2018; Hopken et al.,

97   2013).

98   The factors that cause genetic structure can be correlated and confounding (Meirmans, 2012;

99   Perez et al., 2018; Wang & Bradburd, 2014). Different mechanisms can mask the occurrence of major

100   drivers by promoting those more ancillary with regard to single-species assessments. Variation in

101   intraspecific genetic β-diversity across space is driven by the balance between gene flow and genetic drift

102   (Hutchinson & Templeton, 1999), which is, in turn, tied to dispersal, life history, and biogeography

103   (Avise, 1992; Comte & Olden, 2018). While co-occurring species with similar histories and environments

104   should display similar *patterns* of genetic structure that reflect their shared set of extrinsic factors, the

105   *degree* of genetic divergence across space can vary due to differences in the species' intrinsic

106   characteristics (Riginos et al., 2014). Weaker genetic β-diversity can reduce the power to model genetic

107   structure accurately (Jones & Wang, 2012). The emerging results are twofold: Potentially erroneous

108   conclusions, which in turn beget ineffective management strategies. These issues can be mitigated using

109   replicated multispecies assessments to allow influential major processes to surface, thus effectively

4

110   categorizing both 'signal and noise' components with the former driving patterns of regional biodiversity

111   (Roberts et al., 2013).

112   Our objective was to establish an approach from which the generality of the SHM could be tested

113   across species of a riverscape fish community. This approach would allow key drivers to be identified,

114   with a concurrent expectation of common processes re-emerging within these ecological networks. We

115   accomplish this by comparing patterns of genetic diversity across 31 fish species within the White River

116   Basin of the Ozarks (AR/MO, USA). For each, we compared four models representing major drivers of

117   genetic structure: Isolation by distance (IBD), isolation by stream hierarchy (IBH), isolation by barrier

118   (IBB), and isolation by environment (IBE). We predicted that IBH would consistently explain genetic β-

119   diversity for most species and that variation explained by the alternative models would mostly be

120   captured by IBH. While we expected this consistent *pattern* of congruence between river network

121   architecture and genetic β-diversity, we expected that the *degree* of the strength of this association would

122   vary among species due to their unique histories and intrinsic characteristics that influenced dispersal and

123   effective population sizes. Our data represent thousands of SNPs (single nucleotide polymorphisms)

124   derived via recent advances in high-throughput sequencing (Peterson et al., 2012). This technology has, in

125   turn, allowed thousands of individuals to be genotyped as a financially and logistically practical research

126   endeavor across multiple non-model species (da Fonseca et al., 2016). Our study and the data we present

127   are relatively novel and provide an additional perspective compared to that established with meta-analytic

128   frameworks with similar aims (e.g., Pas-Vinas et al., 2015). We offer our approach as a potential blueprint

129   for developing more comprehensive genetic management plans at the community level.

130

131   **2 | MATERIALS AND METHODS**

132   **2.1 | Study system**

133   Our study system, the White River Basin, is located within the Western Interior Highlands of North

134   America, a previous component of the more extensive pre-Pleistocene Central Highlands extending north

135   and east but subsequently subsumed by numerous glacial advances into two disjunct sub-components:

136   Western Interior Highlands (i.e., Ozarks, Ouachita Mountains), and Eastern Highlands (i.e., Appalachian

137   Plateau, Blue Ridge, Appalachian Highlands) (Mayden, 1985). The Ozark Plateau remained an

138   unglaciated refugium with elevated endemism and diversity (Warren et al., 2000). The White River Basin

139   was established by at least Late Pliocene (>3 MYA; Jorgensen, 1993), but its eastern tributaries were

140   captured by the Mississippi River when it bisected the basin during the Pleistocene (Mayden, 1988;

141   Strange & Burr, 1997). This paleohydrologic signature may remain in contemporary patterns of

142   population divergence in the White River Basin, as manifested by replicated patterns of genetic structure

143   between eastern and western populations.

144

145   **2.2 | Sampling**

146   The sampling region for our study is composed of the White River and St. Francis River basins (AR/MO)

147   (Figure 1). Both are tributaries to the Mississippi River, draining 71,911 km^2 and 19,600 km^2,

148   respectively. Five sub-basins are apparent: St. Francis, Upper White, Black, Lower White, and Little Red

149   rivers (Figure 1). These are further subdivided into the following hierarchical Hydrologic Units (HUC)

150   (USGS & USDA-NRCS, 2013; USGS, 2021) representing different spatial scales: HUC-4 Subregions

151   (N=2); HUC-6 Basins (N=3); HUC-8 Subbasins (N=19); HUC-10 Watersheds (N=129) (Figure 1).

152   Sampling was approved by the University of Arkansas Institutional Animal Care and Use

153   Committee (IACUC: #17077), with collecting permits as follows: Arkansas Game & Fish Commission

154   (#020120191); Missouri Department of Wildlife Conservation (#18136); US National Parks Service

155   (Buffalo River Permit; BUFF-2017-SCI-0013). Fishes were sampled using seine nets in wadable streams

156   during low flow between June 2017 and September 2018. Time spent sampling a site ranged from 30–60

157   mins, with a target of 5-10 individuals/species encountered. Individuals were euthanized by immersion in

158   tricaine methanesulfonate (MS-222) at a concentration of 500 mg/L, buffered to pH=7 with subsequent

159   preservation in 95% ethanol. Formal species diagnosis occurred in the laboratory, and the right pectoral

160   fin was removed from each specimen and stored in 95% ethanol at -20 °C prior to subsequent DNA

161 extraction. Specimens are housed at the Arkansas Conservation and Molecular Ecology Lab, University

162 of Arkansas, Fayetteville.

163

164 **2.2 | Genomic data collection and filtering**

165 Genomic DNA was isolated (Qiagen Fast kits; Qiagen Inc.) and quantified by fluorometry (Qubit;

166 Thermo-Fisher Scientific). Individuals were genotyped using double-digest restriction site-associated

167 DNA (ddRAD) sequencing (Peterson et al., 2012), with procedures modified appropriately (Chafin et al.,

168 2019). Standardized DNA amounts (1,000 ng) were digested at 37°C with high-fidelity restriction

169 enzymes *Msp*I (5'-CCGG-3') and *Pst*I (5'-CTGCAG-3') (New England Biosciences), bead-purified

170 (Ampure XP; Beckman-Coulter Inc.), standardized to 100 ng, and then ligated with custom adapters

171 containing in-line identifying barcodes (T4 Ligase; New England Biosciences). Samples were pooled in

172 sets of 48 and size-selected from 326-426 bp, including adapter length (Pippin Prep; Sage Sciences).

173 Illumina adapters and i7 index were added via 12-cycle PCR with Phusion high-fidelity DNA polymerase

174 (New England Biosciences). Three libraries (3x48=144 individuals/lane) were pooled per lane and single-

175 end sequenced on the Illumina HiSeq 4000 platform (1x100bp; Genomics & Cell Characterization Core

176 Facility; University of Oregon, Eugene). Quality control checks, including fragment analysis and

177 quantitative real-time PCR, were performed at the core facility before sequencing. Raw sequence reads

178 are deposited in the NCBI Sequence Read Archive (Zbinden, Douglas, et al., 2022a).

179   Raw Illumina reads were demultiplexed, clustered, filtered, and aligned in IPYRAD v.0.9.62

180 (Eaton & Overcast, 2020). Reads were first demultiplexed, allowing up to one barcode mismatch,

181 yielding individual FASTQ files containing raw reads (N=3,060 individual files). Individuals averaged >2

182 million reads, with those extremely low removed ($< \bar{x} - 2s$) to reduce errors from poor-quality

183 sequencing. Individuals were previously screened for admixture among species using a combination of

184 standard analyses for assigning ancestry proportions based on SNP data (Zbinden, Douglas, et al., 2022b).

185 There were 70 putatively admixed individuals removed, most (N=66) belonging to the minnow family

186 Leuciscidae. Raw sequence reads were partitioned by species (N=31) and aligned *de novo* in IPYRAD

7

187 (Eaton & Overcast, 2020). Adapters/primers were removed, and reads with >5 bases having Phred quality

188 <20 or read length <35 bases (after trimming) were discarded. Clusters of homologous loci were

189 assembled using an 85% identity threshold. Putative homologs were removed if any of the following were

190 met: <20x and >500x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of

191 nucleotides polymorphic; >8 indels present; or presence in <15% of individuals. Paralogs were identified

192 (and subsequently removed) as those clusters exhibiting either >2 alleles per site in consensus sequence or

193 excessive heterozygosity (>5% of consensus bases or >50% heterozygosity/site among individuals).

194   Biallelic SNP panels for each species were then visualized and filtered with the R package

195 RADIATOR (Gosselin, 2020). To ensure high data quality, loci were removed if: Monomorphic; minor

196 allele frequency <3%; Mean coverage <20 or >200; Missing data >30%; SNP position on read >91; and if

197 HWE lacking in one or more sampling sites ($\alpha = 0.0001$). To reduce linkage disequilibrium, only one

198 SNP per locus was retained (that which maximized minor allele count). Finally, singleton individuals per

199 species at a sampling site and those with >75% missing data in the filtered panel were removed. Species-

200 level SNP panel alignments, metadata, and R code are available on Open Science Framework (Zbinden,

201 Douglas, et al., 2022c).

202

203 **2.3 | Exploring genetic structure**

204 Intraspecific genetic structure was first assessed among sites for visualization, with subsequent modeling

205 done among *individuals*. For each species (*N*=31), pairwise F_{ST} (Weir & Cockerham, 1984) was

206 calculated among sites (HIERFSTAT; Goudet et al., 2017). Jost's *D* was also quantified among sites and

207 globally, as it is based on the effective number of alleles rather than heterozygosity and hence less biased

208 by sampling differences (Jost, 2008). Additional global intraspecific F_{ST} analogs were also quantified for

209 comparison: Multi-allelic G_{ST} (Nei, 1973) and unbiased G''_{ST} (Meirmans & Hedrick, 2011) (MMOD;

210 Winter, 2012). We tested for isolation by distance (IBD) using both linearized F_{ST} and Jost's *D*. Their

211 relationships with river distance (log-transformed) were assessed using the Mantel test (Mantel & Valand,

212 1970) (ECODIST; Goslee & Urban, 2020) and visualized using linear regression (Rousset, 1997).

8

213   Admixture analysis of population structure and ancestry coefficients were estimated using sparse

214   non-negative matrix factorization (sNMF) (Frichot et al., 2014). We ran sNMF for each species, with 20

215   repetitions per *K* value (1 to *N* sites or 20, whichever was smallest) and α=100 (LEA; Frichot & François,

216   2015). The best *K* (i.e., the number of distinct gene pools) from each sNMF run minimizes the cross-

217   validation entropy criterion (Alexander & Lange, 2011). The best *K* was then used to impute missing data

218   (*impute* function using method='mode' in LEA). The sNMF algorithm was then repeated (as above) using

219   imputed genotypes. The resulting Q-matrices of ancestry coefficients per cluster were used to map

220   population structure and served as the "IBB" (isolation by barrier) model below. The indirect inference of

221   dispersal barriers based on discrete population structure was previously used to represent IBB (Ruiz-

222   Gonzalez et al., 2015) and is consistent with the use of population structure maps to infer the presence of

223   dispersal barriers.

224   We further assessed among-site genetic variation between Hydrologic Units (HUCs) and discrete

225   population clusters (determined via sNMF) using analysis of molecular variance (AMOVA) (Excoffier et

226   al., 1992). The purpose was to determine which spatial scale of HUC explained the most variance and to

227   compare that to the variance explained by discrete structure across *K* populations inferred above with

228   sNMF. AMOVA was performed for each species at four HUC levels (4-, 6-, 8-, and 10-digit) to compare

229   the amount of genetic variation among HUCs, all sites, and sites within HUCs. The Watershed Boundary

230   Dataset (USGS, 2021) was used to assign HUC classifications to each site. AMOVA was then performed

231   for each species with genetic clusters *K*>1 to compare the genetic variation among discrete populations,

232   all sites, and sites within populations. Individuals with admixed ancestry between different *K* populations

233   were assigned to a population based on their highest ancestry proportion. The variance components were

234   used to estimate Φ-statistics (analogous to *F*-statistics): Φ_{CT} = the genetic variation among groups (either

235   HUCs or discrete populations); Φ_{ST} = the genetic variation among sites across all groups; and Φ_{SC} = the

236   genetic variation among sites within groups. The wrapper R package POPPR (Kamvar et al., 2015) was

237   used to implement the PEGAS (Paradis, 2010) version of AMOVA with default settings.

238

239 **2.4 | Modeling genetic β-diversity**

240 To compare four models of spatial genetic variation among groups of individuals (i.e., β-Diversity), we

241 employed a variation partitioning framework based on partial redundancy analysis (Capblancq &

242 Forester, 2021; Chan & Brown, 2020). We elected to analyze individual genetic values (i.e., individual x

243 SNP data matrix) rather than decomposing them into among-site distances (e.g., F_{ST}) to allow the use of

244 more powerful multivariate methods rather than derived forms of the Mantel test (e.g., multiple regression

245 on distance matrices) (Legendre & Fortin, 2010). For each species, we partitioned a matrix representing

246 *individual genetic variation* among four explanatory matrices based on: IBD, IBB, IBH, and IBE (Figure

247 2).

248 Individual genetic variation within each species (i.e., individual x SNP data matrix) was reduced

249 to major axes of variation using principal components analysis (PCA) on each SNP panel (Xuereb et al.,

250 2018). The appropriate number of PCs retained for each species was determined by testing the

251 significance of observed component eigenvalues versus random eigenvalues generated by randomly

252 shuffling the genetic data matrix and performing a PCA 999 times. The *p*-value for each PC axis was

253 estimated as: (number of random eigenvalues equal to or larger than the observed + 1)/1000 (Rnd-

254 Lambda; Peres-Neto et al., 2005). This method was implemented using the R package PCDIMENSION

255 (Coombes & Wang, 2019). For each species, individual scores on the retained PCs represented individual

256 genetic variation (i.e., response matrix) which was modeled using the explanatory matrices (i.e., models)

257 described below (Figure 2).

258 The first model (IBD) relied on river network distance measured between individuals (RIVERDIST;

259 Tyers, 2017). The distance matrix was decomposed into positively correlated spatial eigenvectors using

260 distance-based Moran's eigenvector maps (Borcard & Legendre, 2002; Chan & Brown, 2020; Dray et al.,

261 2006) within the R package ADESPATIAL (Dray et al., 2020). Each individual was assigned a score for

262 each eigenvector positively correlated with genetic diversity (Figure 2). Eigenfunction analysis is an

263 alternative means to assess the contribution of geographic distance on patterns of genetic variation and is

264 more powerful than Mantel correlations at detecting fine-scale structure (Xuereb et al., 2018).

265   The second model (IBB) was based on the discrete population structure inferred above using the

266   admixture analysis sNMF, which was represented as the individual population coefficients (i.e., Q-

267   matrix). The assumption was that discrete population structure indicates a reduction of gene flow between

268   populations due to a barrier (or high resistance) to dispersal. Similar clustering analysis output was

269   previously used to represent IBB (Ruiz-Gonzalez et al., 2015). However, IBB often implicitly refers to

270   instream barriers, such as dams and weirs. Our indirect approach based on population coefficients should

271   capture the effects of those barriers as well as natural barriers to gene flow. While testing for significant

272   genetic differences among populations defined by clustering methods is entirely meaningless due to

273   circularity, it is reasonable to assess the strength of separation among clusters (e.g., F-statistics from

274   AMOVA) or variation explained by clustering (e.g., adjusted R^2 from RDA) (Meirmans 2015). Note: the

275   IBB model could not be incorporated for species in which population structure was not apparent (K=1),

276   and these species were thus tested using only three models of genetic structure.

277   The third model (IBH) was constructed using four levels of HUCs (4-, 6-, 8-, and 10-digit) that

278   characterized an individual's position within the stream hierarchy, i.e., hydrologic unit (USGS, 2021).

279   Each site and individual within site could be classified by its unique HUC code at the four levels. These

280   codes were nominal categories and were decomposed into N-1 binary 'dummy variables.' This

281   decomposition produces a 'new' variable for each HUC within a spatial level, and an individual is either

282   within (=1) or outside (=0) a given HUC. Note: the dummy variables produced were one less than the

283   number of original categories to avoid introducing collinearity.

284   The fourth model (IBE) relied on contrasting environmental variation across sites—represented

285   using principal component scores (Figure 2). Environmental variables were taken from a compendium of

286   281 factors distinguished by five major classes: (i) hydrology/physiography, (ii) climate, (iii) land cover,

287   (iv) geology, and (v) anthropogenic (HYDRORIVERSv.1.0; Linke et al., 2019). We aimed to reduce the

288   environmental factors to include only those demonstrating a significant relationship with genetic

289   variation. Variables were first analyzed within major classes, with invariant factors and those exhibiting

290   collinearity being removed in a stepwise manner (USDM; Naimi, 2013) until each had a variation inflation

11

291   factor (VIF) <10 to reduce collinearity. Standardization occurred by subtracting means and dividing by

292   standard deviations to ensure variables had a common scale. Variables within each major class (e.g.,

293   climate) were selected for subsequent analyses using forward selection (Blanchet et al., 2008). Variables

294   within the class being considered (e.g., climate) were first tested for a significant relationship with the

295   response data (individual genetic variation) using redundancy analysis (RDA; Rao, 1964). If the overall

296   relationship was significant ($\alpha < 0.05$), a stepwise forward procedure was carried out such that specific

297   variables were selected if the adjusted R^2 of the model increased significantly ($\alpha < 0.05$) and the adjusted

298   R^2 did not exceed that of the overall model (Blanchet et al., 2008). This procedure was employed using

299   the *ordiR2step* function in the R package VEGAN (Oksanen et al., 2020). The selected variables from each

300   of the five classes (hydrology/physiography, climate, land cover, geology, and anthropogenic) were

301   combined into a single environmental matrix, then reduced to a set of PCs using robust principal

302   components analysis (ROBPCA; Hubert et al., 2005). The number of PCs retained was determined

303   following Hubert and coworkers (2005), as implemented in the R package ROSPCA (Hubert et al., 2016).

304   The environmental matrix of individuals x PCs was used to model IBE.

305   For each species, individual genetic variation (individuals x PCs matrix) was partitioned among

306   the four explanatory models of genetic structure (IBD, IBB, IBH, IBE) using variation partitioning based

307   on partial redundancy analysis (pRDA; Anderson & Legendre, 1999; Borcard et al., 1992). Redundancy

308   analysis (Rao, 1964) is a constrained ordination technique and an extension of multiple regression that

309   summarizes the relationship between linear combinations of multiple response variables and linear

310   combinations of multiple explanatory variables (Capblancq & Forester, 2021). The overall variance in the

311   response matrix is partitioned into constrained and unconstrained fractions. The former is interpreted as

312   the amount of response variation 'explained' by the explanatory set(s), expressed as a proportion

313   equivalent to R^2 in multiple regression, which is adjusted to reduce bias introduced by multiple predictors

314   (Peres-Neto et al., 2006). The formula for the full RDA model herein was: Genetic β Diversity ~ IBD +

315   IBB + IBH + IBE (Figure 2). Variation partitioning consists of running sequential pRDAs to adjust the

316   linear effects of an explanatory set on the response by accounting for other explanatory sets first, e.g.,

317   Genetic β Diversity ~ IBD | (IBB + IBH + IBE). (In other words, here we estimate variation explained by

318   distance after removing that explained by the other three response matrices.) This set of tests allows

319   partitioning the variation explained into that 'purely' attributable to a given explanatory set (e.g., IBD) and

320   into shared components attributable to two or more sets (e.g., IBD + IBH). Variation partitioning allows

321   the correlation structure among competing models to be visualized, typically using Venn diagrams

322   showing adjusted R^2 for different intersectionalities of the explanatory data sets (i.e., pure and correlated

323   components). The variation partitioning was performed using the *varpart* function in the R package VEGAN

324   (Oksanen et al., 2020). The models were judged based on their significance, adjusted coefficients of

325   determination R^2, and correlation with other models. The 'best' model should explain the most variation in

326   the genetic response matrix (i.e., highest adj. R^2), should account for variation after accounting for that

327   explained by other models (i.e., 'pure' adj. $R^2 > 0$), and account for most variation explained by competing

328   models (i.e., other models are redundant).

329

330   **3 | RESULTS**

331   **3.1 | Sampling and data recovery summarized**

332   Collections (*N*=75; Figure 1) yielded *N*=72 species and *N*=3,605 individuals. On average, we collected

333   ~11 species/site, typical for streams sampled with seine nets in North America (Matthews, 1998) and

334   similar highland streams within the Mississippi Basin (Zbinden, Geheber, Lehrter, & Matthews, 2022;

335   Zbinden, Geheber, Matthews, & Marsh-Matthews, 2022).

336   We genotyped *N*=3,060 individuals across *N*=31 species, with at least two individuals collected

337   at ≥5 sampled sites. Simulations and empirical evaluations underscore the accuracy of F_{ST} estimates when

338   large numbers of SNPs (≥1,500) are employed across a minimum of two individuals (Nazareno et al.,

339   2017; Willing et al., 2012). After removing samples with missing data >75% and those as singletons of

340   their species at a site, the remaining *N*=2,861 were analyzed for genetic structure (Table 1). The number

341   of individuals analyzed per species ranged from 15–358 (x=92.3; *s*=80.8), and the sites at which each

342   species was collected ranged from 5–50 (x=16.8; *s*=11.2). The number of individuals/species/site

343   ranged from 2–15 ($x\square$=5.1; s=1.5). The mean number of raw reads/individual/species spanned from 1.65

344   million to 3.22 million ($x\square$=2,289,230.0; s=341,159.5). The mean N of loci/species recovered by IPYRAD

345   ranged from 14,599–30,509 ($x\square$=20,081.7; s=4,697.6) with a mean sequencing depth/locus of 73.6x

346   (s=12.0x). After filtering loci and retaining one SNP per locus, the panels for each species contained

347   2,168–10,033 polymorphic sites ($x\square$=4,486.7; s=1,931.1) with mean missing data/species at 12% (s=2%).

348

349   **3.2 | Genetic structure**

350   **3.2.1 | Among-site genetic divergence**

351   Distributions of among-site F_{ST} and D varied widely among species (Figure 3; Supplement S1), as did

352   global indices of genetic divergence (Table 2). All three global indices of fixation or genetic divergence

353   (G_{ST}, G''_{ST}, D) were negatively correlated with within-site heterozygosity (H_S), positively correlated with

354   total heterozygosity (H_T), and highly, positively correlated with each other (Table 3).

355   A significant relationship was found between linearized among-site F_{ST} and log-transformed

356   among-site river network distance for 23 (74%) of the N=31 species (Figure 4). Mantel coefficients

357   ranged from 0.11–0.88 ($x\square$=0.51; s=0.19). Results were largely similar when IBD was tested with Jost's

358   D, again with the same 23 species showing a significant relationship, along with two additional taxa:

359   Smallmouth Bass (*Micropterus dolomieu*; Lacepède, 1802) and Largemouth Bass [*Micropterus*

360   *salmoides*; (Lacepède, 1802)]. Mantel correlation coefficients ranged from 0.15–0.92 ($x\square$=0.51; s=0.19).

361

362   **3.2.2 | Population structure**

363   An apparent lack of discrete genetic structure emerged across seven species, suggesting continuous

364   structuring at the spatial scale of our study (Figure 5). For the remaining 24 species, at least two and up to

365   seven discrete sub-populations were identified (Figure 6). This structure corresponded at the broadest

366   hierarchical level to the two major northern basins: Upper White and Black rivers, for all species sampled

367   in both sub-basins (N=22). There was also evidence of fine-scale structure for five species within the

14

368   Little Red River Basin. Smaller catchments with distinct gene pools across multiple species included:

369   North Fork (4 spp.), Buffalo (3 spp.), Upper Black (4 spp.), Current (3 spp.), and Spring rivers (4 spp.).

370

371   **3.2.3 | AMOVA**

372   Discrete genetic structuring was also supported via AMOVA. Genetic variation among HUCs was

373   significant for 24 species (Table 4). For the other seven species, variation among HUCs was $\leq 1\%$, save

374   for Ozark Sculpin (*Cottus hypselurus*; Robins & Robison, 1985) and Creek Chub [*Semotilus*

375   *atromaculatus*; (Mitchill, 1818)]. HUC differences for these accounted for >80% of the genetic variance

376   but were non-significant due to a lack of power. Southern Redbelly Dace [*Chrosomus erythrogaster*;

377   (Rafinesque, 1820)] could not be tested due to a lack of repeated samples within HUC levels. Further

378   evidence of genetic structure among HUCs was revealed in the pattern of Φ_{SC} (genetic divergence among

379   sites within HUCs) < Φ_{ST} (divergence among all sites) found across 26 species. The 8-digit HUC level

380   explained the greatest genetic variance across 21 species (Table 4).

381   Genetic variation among discrete population clusters (based on sNMF) was significant for 21 of

382   the N=31 species (Table 4). Seven species were best described as single populations (K=1) and were not

383   tested. The three species without significant structure, despite K>1 via sNMF, could likely be explained

384   by low power resulting from a small number of sample sites. Again, as with HUCs, $\Phi_{SC} < \Phi_{ST}$ was

385   observed. However, all tested species showed this pattern (i.e., sites within the same population were less

386   differentiated than sites across all populations).

387

388   **3.3 | Modeling genetic β-diversity**

389   Variability in genetic β-diversity was partitioned across four models of genetic structure for the N=31

390   species. Principal components of SNP panel variation served as representatives of genetic variation.

391   Across species, the number of genetic PCs ranged from 2–93 (x□=20.0; s=20.1; Table 1).

392   Combining the four models (IBD, IBB, IBH, IBE) explained between 3–100% of the genetic β-

393   diversity across species (x□=63.0%; s=35.3%; Figure 7). Isolation by stream hierarchy (IBH; x□=62.0%;

15

394   *s*=34.7%) and barrier (IBB; $x\square$=49.3%; *s*=30.0%) contributed most to the total variation explained, while

395   distance (IBD; $x\square$=32.1%; *s*=25.1%) and environment (IBE; $x\square$=33.0%; *s*=21.4%) explained less (Figure

396   7; Supplement S1). Variation explained by "pure" models, after accounting for that explained by the other

397   three, was >0 only for stream hierarchy and barrier (Figure 7; Supplement S1), suggesting that distance

398   and environment are encapsulated by the former. Indeed, correlative structure among models revealed

399   most genetic variance was explained by stream hierarchy, with the other models largely redundant (Figure

400   8; Supplement S1).

401

402   **4 | DISCUSSION**

403   Genetic diversity is an essential metric for inferring evolutionary processes and guiding conservation.

404   Single-species estimates of genetic diversity are standard given practical constraints, e.g., funding

405   mandates for species of conservation concern. However, adopting a multispecies approach for analyzing

406   genetic diversity could allow for more comprehensive and systematic management plans to be developed

407   by focusing on commonalities (rather than differences) among species. The Stream Hierarchy Model

408   (Meffe & Vrijenhoek, 1988) posits that the dispersal of stream-dwelling organisms is more limited

409   between hierarchical units (basins, sub-basins, watersheds) than within (i.e., 'spatial modularity; Fortuna

410   et al., 2009). If this model was generalizable, it could determine relevant scales and regions for managing

411   genetic diversity that may harbor complementary biodiversity and aid in systematic conservation planning

412   (Margules & Pressey, 2000; Paz-Vinas et al., 2018; Xuereb et al., 2021).

413   Few studies have analytically compared the spatial structure of genetic β-diversity. Paz-Vinas et

414   al. (2018) compared genetic α- and β-diversity among six river-dwelling fish species in France using

415   microsatellites. In contrast to our results, their study found that species did not conform to common

416   spatial patterns, although their approach differed from that herein (i.e., 'hot-' and 'coldspots' of α- and β-

417   diversity). Fortuna et al. (2009) used isozymes to demonstrate that three of four Mediterranean shrubs in

418   their study displayed similar patterns of network connectivity (modularity) in that the same sets of

16

419   sampling sites formed consistent patches based on estimates of genetic connectivity (i.e., higher

420   connectivity among sites *within* a patch) that likely represent fundamental scales for populations.

421   Our multispecies approach yielded two salient points: 1) From a macro-perspective, river network

422   topology and complexity are manifested in common patterns of genetic structure across species

423   (consistent modularity); and 2) on a finer scale, the degree of intraspecific genetic divergence varies

424   widely among co-distributed species. Most species showed significant IBD patterns but also discrete

425   population sub-structure, as reflected most strongly by sub-basin delineations (e.g., HUC-8). These

426   patterns were corroborated by AMOVA and variation partitioning and are generalized across species.

427   Overall, stream fish genetic structure patterns indicated dispersal limited primarily among *versus* within

428   river catchments.

429

430   **4.1 | Drivers of isolation at the basin-wide scale**

431   4.1.1   *Isolation by Distance and river networks*

432   IBD is expected when a genetic study's spatial extent is greater than individuals' average dispersal

433   distance, i.e., distance moved from natal habitat to breeding habitat. We examined species from seven

434   families of fishes, most of which are small-bodied, highland species that presumably do not disperse great

435   distances (Matthews & Robison, 1988). Indeed, significant IBD patterns were detected in 81% of the

436   species in our study. However, the strength of the relationship was generally weak (Mantel r =0.47 &

437   0.51 for linearized F_{ST} and D, respectively).

438   While IBD may primarily explain genetic variation along a single stream or river, i.e., linear

439   scale, it fails to incorporate the spatial structure of riverine networks (Thomaz et al., 2016). Therefore,

440   IBD may not be an appropriate general model for fish genetic structure at the network scale (Hopken et

441   al., 2013). IBD plots for many species (Figure 4) showed high genetic divergence even among relatively

442   proximate localities, with apparent clusters indicating discrete rather than continuous structure (Guillot et

443   al., 2009). This evidence suggests that—at the network scale—a more nuanced pattern occurs, with high

444   residual variation resulting. The failure of IBD to account for large amounts of variation in genetic

17

445   divergence reflects additional resistance to dispersal, as caused by longitudinal changes in habitat

446   characteristics such as slope, depth, volume, and predator composition. For example, two river reaches of

447   equal length can have very different habitat matrices, and these can be more influential on gene flow than

448   space alone (Guillot et al., 2009; Lowe et al., 2006; Ruiz-Gonzalez et al., 2015).

449

450   4.1.2 *Stream Hierarchy Model*

451   Our results show that individual genetic variation is best explained by the Stream Hierarchy Model

452   (Brauer et al., 2018; Hopken et al., 2013; Meffe & Vrijenhoek, 1988). In other words, the majority of

453   variation explained by IBD, IBE, and IBB could be accounted for by IBH alone. This was corroborated

454   via variation partitioning, in which IBD, IBE, and IBB models were redundant with IBH. A concordance

455   of population structure with stream hierarchy yielded a similar percentage of among-site genetic variation,

456   as explained by among-HUC and among-population groupings. In short, the variance explained by

457   distance and environment was due to differences among HUC drainages. These results highlight the

458   necessity of accounting for population structure prior to exploring the relationship between genotypes and

459   environmental heterogeneity, e.g., within genotype by environment frameworks (Lawson et al., 2020).

460

461   4.1.3 *Disentangling cumulative effects*

462   Our analyses also revealed complex spatial patterns of genetic diversity. We evaluated competing

463   isolation models using an approach that identified distance and barriers as putative drivers, with strong

464   genetic divergence identified even across short geographical distances (Chan & Brown, 2020; Ruiz-

465   Gonzalez et al., 2015). This interaction can confound analyses that incorporate either alone. For example,

466   if sampling is clustered, discrete genetic groups can be spuriously inferred along an otherwise continuous

467   gradient of genetic variation (Frantz et al., 2009). Furthermore, a continuous pattern can be erroneously

468   extrapolated when the underlying reality is described by distinct clusters separated by geographic distance

469   (Meirmans, 2012). Here we echo the importance of testing various hypotheses concerning genetic

18

470   structure (Perez et al., 2018). Idiosyncrasies and complex interactions cannot be discerned by testing

471   single models in isolation (e.g., discrete structure or IBD).

472

473   **4.2 | Drivers of variation within and among species**

474   The species assayed herein display marked differences concerning dispersal capability (Shelley et al.,

475   2021). Given this, we expected the *degree* of genetic structure to vary widely among species across our

476   study region (Comte & Olden, 2018; Husemann et al., 2012; Pilger et al., 2017). Dispersal-related traits

477   drive gene flow among localities and determine the spatial scale at which patterns of genetic structure

478   emerge (Bohonak, 1999; Riginos et al., 2014). The physical structure of the river network then further

479   modulates these patterns by dictating dispersal pathways of metapopulations and their colonization and

480   extinction probabilities (Falke et al., 2012; Labonne et al., 2008; Fagan, 2002). These superimposed

481   processes influence genetic divergence among distal populations (Thomaz et al., 2016; Chiu et al., 2020).

482   Similar patterns emerge when analyzing community diversity via species composition. Headwater

483   streams tend to have very different communities due to dispersal limitations (Finn et al., 2011; Zbinden &

484   Matthews, 2017; Zbinden, Geheber, Lehrter, & Matthews, 2022). The interaction between traits and the

485   environment is an overarching influence that unites ecology and evolution.

486   Many species studied herein are small-bodied with aggregate distributions in upland and

487   headwater streams (Robison & Buchanan, 2020). Thus, species-specific dispersal limitations, as imposed

488   by unsuitable large riverine habitats (Radinger & Wolter, 2015; Schmidt & Schaefer, 2018), explain

489   considerable variation in genetic structuring within the White River. Large rivers are hypothesized as

490   inhospitable habitats to upland fishes (e.g., resources, depth, turbidity, substrates) and impose resistance

491   to successful migration (e.g., higher discharge, greater density of large-bodied predators). These

492   characteristics constrain migration and limit gene flow amongst basins that drain into large rivers (Fluker

493   et al., 2014; Schmidt & Schaefer, 2018; Turner & Robison, 2006). The results are asymmetric gene flow

494   and source-sink metapopulation dynamics, with susceptible species, those smaller and less tolerant,

495   diverging most rapidly (Campbell Grant et al., 2007). Thus it is not surprising that the species for which

19

496   we found no evidence of discrete genetic structure (i.e., K=1) are also among the larger, more generalist

497   species.

498   Life-history traits likely play an important role as well. For example, those that directly influence

499   effective population size (Nei & Tajima, 1981; Waples, 2022) may generate differences among species

500   regarding the rate at which genetic differences arise due to genetic drift (Blanchet et al., 2020). Species

501   with 'slow' life histories, characterized by longer generations and delayed maturity, show an increased

502   probability of local extirpation, inflating genetic drift concomitant with global extinction risk (Hutchings

503   et al., 2012; Pearson et al., 2014; Chafin et al., 2019). Similar contingencies exist for ecological traits,

504   such as highly specialized trophic adaptations, narrow environmental tolerances, or any other following

505   the same general mechanism of influence on dispersal or effective population sizes (Olden et al., 2008).

506   Ecological traits are mirrored by morphology (Douglas & Matthews, 1992), underscoring an interaction

507   of trait effects that are difficult to disentangle. Ultimately, intraspecific genetic divergence is driven by a

508   combination of factors that influence population size, demographic history, and connectivity (Zbinden et

509   al., 2022d). Clearly, these complex interactions among drivers require more comparative multispecies

510   assessments as they shape genetic diversity and structure within and among species (microevolutionary

511   scale) and, thus, ultimately lead to speciation and extinction (macroevolutionary scale).

512

513   **4.3 | Disentangling historic and contemporary drivers**

514   4.3.1   *Paleohydrology in the White River system*

515   In this study, discrete population structure coincides with major topological divides within the White

516   River stream network, such as a consistent east/west divide between Upper White and Black rivers,

517   mirroring prior community composition studies (Matthews & Robison, 1988; 1998). Similar patterns

518   were observed at smaller scales among drainages within the study region, as reported for White River

519   crayfish (Fetzner & DiStefano, 2008). While the Lower White and Black rivers are certainly

520   contemporary large-river habitats, both would have been much larger pre-Pleistocene when together they

521   represented the main channel of the Old Mississippi River (Mayden, 1988; Strange & Burr, 1997). This

522   large-river habitat would have separated the eastern and western highland tributaries, with inhospitable

523   habitat for upland species. Pronounced limitations regarding historic dispersal induced by the Old

524   Mississippi could explain the greater isolation of the Little Red River and Black River tributary

525   populations compared to those in the Upper White River. Here, additional work should incorporate

526   coalescent perspectives (e.g., Oaks, 2019) that test the role of past geomorphic events in driving co-

527   divergence and co-demographic patterns, such as the Pleistocene incursion by the Old Mississippi into the

528   modern Black River channel.

529

530   4.3.2   *Contemporary drivers*

531   Spatial discontinuities in genetic structure can also reveal contemporary barriers to migration/gene flow

532   (Lee et al., 2018; Ruiz-Gonzalez et al., 2015). The Upper White River dams (e.g., Norfork, Bull Shoals,

533   Table Rock, and Beaver dams) represent the most apparent anthropogenic barriers to gene flow. Dams

534   elsewhere have demonstrated discrete populations above and below the structure (Roberts et al., 2013).

535   However, observable genetic impacts can be limited due to the relatively short period these dams have

536   been in place (Ruzich et al., 2019). Those on the White River were constructed between 1912

537   (Taneycomo Powersite Dam) and 1966 (Beaver Dam).

538   Our study was not explicitly designed to assess impoundment effects on diversity, nor were they

539   directly tested. Nevertheless, evidence of discrete population structure has emerged, corresponding to the

540   location of such dams. Four species showed discrete populations within the North Fork River above the

541   Norfork Dam: Southern Redbelly Dace [*Chrosomus erythrogaster*; (Rafinesque, 1820)]; Yoke Darter

542   (*Etheostoma juliae*; Meek, 1891); Northern Studfish [*Fundulus catenatus*; (Storer, 1846)]; and

543   Blackspotted Topminnow [*Fundulus olivaceus*; (Storer, 1845)] (sites colored magenta; Figure 6). One

544   species, Orangethroat Darter [*Etheostoma spectabile*; (Agassiz, 1854)], showed a distinct population in

545   the James River above Table Rock Dam (sites colored gold; Figure 6). However, both North Fork and

546   James rivers drain eight-digit HUC watersheds, which explains high amounts of genetic variation across

547   the study region, regardless of dams. This highlights the importance of understanding 'natural' network-

21

548   wide patterns of genetic structure prior to deriving conclusions regarding anthropogenic barriers,

549   particularly when they coincide with stream hierarchy. Differentiating dams as barriers *versus* stream

550   hierarchy could be accomplished through divergence time estimates (Hansen et al., 2014) or by

551   contrasting observed genetic differentiation with that expected in the face of a true obstacle for *N*

552   generations (Prunier et al., 2020). That aspect, as it now stands, is beyond the scope of our current study.

553

554   **5| CONCLUSIONS**

555   The multispecies comparative approach employed here revealed general patterns that could not have been

556   discerned from a singular study of any one species. Additionally, the variability among species regarding

557   intraspecific genetic structure provides specific information of interest that single-species studies cannot.

558   While meta-analytic approaches have some potential, they are limited by confounding effects that stem

559   from differences between studies, such as markers, sample sizes, environmental exigencies, and historical

560   context. This necessitates a community-level approach within a study region. Further work aimed at

561   modeling variables can lead to greater insight, ultimately improving our hypotheses regarding genetic

562   diversity for which contemporary data are unavailable.

563   Importantly, our comparative approach supports the Stream Hierarchy Model as a general model for

564   the genetic structure of lotic fish species and suggests that hydrologic units characterize regional genetic

565   diversity quite well. Out of this result emerged the potential for HUC units to serve as a 'rule of thumb' for

566   riverine biodiversity conservation. None of the species evaluated herein were panmictic. Genetic variation

567   among HUCs was apparent despite limited evidence of discrete population or continuous structure for

568   some species. Across a suite of commonly occurring fishes representing seven families, we identified

569   greater intraspecific gene flow within than among basins/sub-basins. Therefore, fish populations within

570   separate HUCs at the 8-digit+ level (e.g., HUC6, HUC4, HUC2) should be considered isolated until

571   proven otherwise (Shelley et al., 2021).

572   As previously recognized, independent populations warrant independent management (Hopken et

573   al., 2013). When migration is low or non-existent, management of one population is unlikely to impact

22

574   another. Genetic variation unique to hydrologic units could allow for adaptation to future environmental

575   change, while on the other hand, isolation of populations could underscore elevated extirpation risks

576   (Harrisson et al., 2014). Furthermore, efforts to propagate populations via stocking or translocation should

577   carefully assess the genetic landscape of the species in question, particularly before co-mingling diversity

578   from different sub-basins (Meffe & Vrijenhoek, 1988). Such uninformed mixing of genetic stocks could

579   promote outbreeding and the erosion of unique genetic diversity within river catchments. However, this

580   must be weighed against the risks of local extirpation (Pavlova et al., 2017).

581   Given this study's general and comparative nature, we refrain from designating populations within

582   species as potential management units (MUs). However, species showing high levels of genetic structure

583   (Table 2) could be assessed individually for such designation, possibly requiring more fine-scaled,

584   targeted sampling. Additional river/sub-basin-specific management efforts could also be justified, given

585   the presence of unique populations across multiple species (Hopken et al., 2013). These consistent spatial

586   modalities shared among multiple species point to areas of the river network that not only harbor unique

587   genetic diversity but also play a key role in metapopulation viability and metacommunity dynamics

588   (Fletcher et al., 2013). Here we specifically refer to: The Little Red, North Fork, Buffalo, Upper Black,

589   Current, and Spring rivers. These may indeed represent *evolutionarily significant catchments*, and this

590   insight underscores the potential for community-level genetic examination to elevate management to the

591   ecosystem scale (Hanson et al., 2020; Paz-Vinas et al., 2018).

592

600   represent those of the authors and do not necessarily represent the views of the NSF nor other affiliated or

601   contributing organizations.

REFERENCES

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 1–6. https://doi.org/10.1186/1471-2105-12-246

Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, *62*(3), 271–303. https://doi.org/10.1080/00949659908811936

Anthonysamy, W. J. B., Dreslik, M. J., Douglas, M. R., Thompson, D., Klut, G. M., Kuhns, A. R., Mauger, D., Kirk, D., Glowacki, G. A., Douglas, M. E., & Phillips, C. A. (2018). Population genetic evaluations within a co-distributed taxonomic group: a multispecies approach to conservation planning. *Animal Conservation*, *21*(2), 137–147. https://doi.org/10.1111/acv.12365

Avise, J. C. (1992). Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, 62-76. https://doi.org/10.2307/3545516

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, *89*(9), 2623–2632. https://doi.org/10.1890/07-0986.1

Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to go bigger: Emerging patterns in macrogenetics. *Trends in Genetics*, *33*(9), 579–580. https://doi.org/10.1016/j.tig.2017.06.007

Blanchet, S., Prunier, J. G., Paz-Vinas, I., Saint-Pé, K., Rey, O., Raffard, A., Mathieu-Bégné, E., Loot, G., Fourtune, L., & Dubut, V. (2020). A river runs through it: The causes, consequences, and management of intraspecific diversity in river networks. *Evolutionary Applications*, *13*(6), 1195–1213. https://doi.org/10.1111/eva.12941

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *The Quarterly Review of Biology*, *74*(1), 21–45. https://doi.org/10.1086/392950

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1-2), 51-68. https://doi.org/10.1016/S0304-3800(01)00501-4

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045–1055. https://doi.org/10.2307/1940179

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*(11), 3258-3273. https://doi.org/10.1111/evo.12193

Brauer, C. J., Unmack, P. J., Smith, S., Bernatchez, L., & Beheregaray, L. B. (2018). On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system. *Molecular Ecology*, *27*(17), 3484–3497. https://doi.org/10.1111/mec.14808

Campbell Grant, E. H., Lowe, W. H., & Fagan, W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, *10*(2), 165-175. https://doi.org/10.1111/j.1461-0248.2006.01007.x

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309. https://doi.org/10.1111/2041-210X.13722

Chafin, T. K., Douglas, M. R., Martin, B. T., & Douglas, M. E. (2019). Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity*, *123*(6), 759–773. https://doi.org/10.1038/s41437-019-0259-2

Chan, K. O., & Brown, R. M. (2020). Elucidating the drivers of genetic differentiation in Malaysian torrent frogs (Anura: Ranidae: Amolops): A landscape genomics approach. *Zoological Journal of the Linnean Society*, *190*(1), 65–78. https://doi.org/10.1093/zoolinnean/zlz151

Chiu, M. C., Li, B., Nukazawa, K., Resh, V. H., Carvajal, T., & Watanabe, K. (2020). Branching networks can have opposing influences on genetic variation in riverine metapopulations. *Diversity and Distributions*, *26*(12), 1813–1824. https://doi.org/10.1111/ddi.13160

Comte, L., & Olden, J. D. (2018). Fish dispersal in flowing waters: A synthesis of movement□ and genetic□based studies. *Fish and Fisheries*, *19*, 1063–1077. https://doi.org/10.1111/faf.12312

652 Coombes, K. R., Wang, M., & Coombes, M. K. R. (2019). Package 'PCDimension'. Retrieved from:
653   *https://cran.r-project.org/package=PCDimension*

654 Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex
655   landscapes: Testing multiple hypotheses with causal modeling. *American Naturalist*, *168*(4), 486–
656   499. https://doi.org/10.1086/506976

657 Cushman, S. A., Wasserman, T. N., Landguth, E. L., & Shirk, A. J. (2013). Re-evaluating causal
658   modeling with mantel tests in landscape genetics. *Diversity*, *5*(1), 51–72.
659   https://doi.org/10.3390/d5010051

660 da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L.,
661   Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation
662   biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*,
663   3–13. https://doi.org/10.1016/j.margen.2016.04.012

664 Davis, C. D., Epps, C. W., Flitcroft, R. L., & Banks, M. A. (2018). Refining and defining riverscape
665   genetics: How rivers influence population genetic structure. *Wiley Interdisciplinary Reviews:*
666   *Water*, *5*(2), e1269. https://doi.org/10.1002/wat2.1269

667 De Kort, H., Prunier, J. G., Ducatez, S., Honnay, O., Baguette, M., Stevens, V. M., & Blanchet, S. (2021).
668   Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and
669   animal populations. *Nature Communications*, *12*(1), 1-11. https://doi.org/10.1038/s41467-021-
670   20958-2

671 DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance
672   of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147-4154.
673   https://doi.org/10.1111/mec.16051

674 Douglas, M. R., Anthonysamy, W. J. B. Mussmann, S. M., Davis, M. A., Louis, W. and Douglas, M. E.
675   (2020). Multi-targeted management of upland game birds at the agroecosystem interface in
676   midwestern North America. *PLoS ONE* 15(4), e0230735.
677   https://doi.org/10.1371/journal.pone.0230735

678 Douglas, M. E., and Matthews, W. J. (1992). Does morphology predict ecology? Hypothesis testing
679   within a freshwater stream fish assemblage. *Oikos*, 65(2), 213-224. https://doi.org/10.2307/3545012

680 Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G.,
681   Legendre, P., & Wagner, H. (2020). R package "adespatial": Multivariate multiscale spatial analysis.
682   Retrieved from: https://cran.r-project.org/package=adespatial

683 Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for
684   principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3-4), 483-
685   493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

686 Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets.
687   *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

688 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from
689   metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction
690   data. *Genetics*, *131*(2), 479–491. https://doi.org/10.1093/genetics/131.2.479

691 Fagan, W. F. (2002). Connectivity, fragmentation, and extinction risk in dendritic
692   metapopulations. *Ecology*, *83*(12), 3243-3249. https://doi.org/10.1890/0012-
693   9658(2002)083[3243:CFAERI]2.0.CO;2

694 Falke, J. A., Bailey, L. L., Fausch, K. D., & Bestgen, K. R. (2012). Colonization and extinction in
695   dynamic habitats: an occupancy approach for a Great Plains stream fish assemblage. *Ecology*, *93*(4),
696   858-867. https://doi.org/10.1890/11-1515.1

697 Fetzner Jr, J. W., & DiStefano, R. J. (2008). Population genetics of an imperiled crayfish from the White
698   River drainage of Missouri, USA. *Freshwater Crayfish*, *16*, 131-146.

699 Finn, D. S., Bonada, N., Múrria, C., & Hughes, J. M. (2011). Small but mighty: Headwaters are vital to
700   stream network biodiversity at two levels of organization. *Journal of the North American*
701   *Benthological Society*, *30*(4), 963–980. https://doi.org/10.1899/11-012.1

702 Fletcher, R. J., Revell, A., Reichert, B. E., Kitchens, W. M., Dixon, J. D., & Austin, J. D. (2013).

703   Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature*
704   *communications*, *4*(1), 1-7. https://doi.org/10.1038/ncomms3572

705 Fluker, B. L., Kuhajda, B. R., & Harris, P. M. (2014). The influence of life-history strategy on genetic
706   differentiation and lineage divergence in darters (Percidae: Etheostomatinae). *Evolution*, *68*(11),
707   3199–3216. https://doi.org/10.1111/evo.12505

708 Fortuna, M. A., Albaladejo, R. G., Fernández, L., Aparicio, A., & Bascompte, J. (2009). Networks of
709   spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, *106*(45),
710   19044-19049. https://doi.org/10.1073/pnas.0907704106

711 Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to
712   determine the genetic structure of a continuously distributed population: Clusters or isolation by
713   distance? *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-
714   2664.2008.01606.x

715 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies.
716   *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

717 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation
718   of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.
719   https://doi.org/10.1534/genetics.113.160572

720 Goslee, S., & Urban, D. (2020). ecodist: Dissimilarity-based functions for ecological analysis. Retrieved
721   from: https://cran.r-project.org/package=ecodist

722 Gosselin, T. (2020). R package radiator: RADseq data exploration, manipulation and visualization using
723   R. Retrieved from: https://thierrygosselin.github.io/radiator/

724 Goudet, J., Jombart, T., Kamvar, Z. N., Archer, E., & Hardy, O. (2017). hierfstat: estimation and tests of
725   hierarchical F-statistics. Retrieved from: https://cran.r-project.org/package=hierfstat

726 Guillot, G., Leblois, R., Coulon, A., & Frantz, A. C. (2009). Statistical methods in spatial genetics.
727   *Molecular Ecology*, *18*(23), 4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x

728 Hansen, M. M., Limborg, M. T., Ferchaud, A. L., & Pujolar, J. M. (2014). The effects of medieval dams
729   on genetic divergence and demographic history in brown trout populations. *BMC Evolutionary*
730   *Biology*, *14*(1), 1–14. https://doi.org/10.1186/1471-2148-14-122

731 Hanson, J. O., Marques, A., Veríssimo, A., Camacho☐Sanchez, M., Velo☐Antón, G., Martínez☐Solano,
732   Í., & Carvalho, S. B. (2020). Conservation planning for adaptive and neutral evolutionary
733   processes. *Journal of Applied Ecology*, *57*(11), 2159-2169. https://doi.org/10.1111/1365-
734   2664.13718

735 Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize
736   evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*(9), 1008–
737   1025. https://doi.org/10.1111/eva.12149

738 Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for
739   landscape genetics. *Landscape Ecology*, *21*(6), 797-807. https://doi.org/10.1007/s10980-005-5245-9

740 Hopken, M. W., Douglas, M. R., & Douglas, M. E. (2013). Stream hierarchy defines riverscape genetics
741   of a North American desert fish. *Molecular Ecology*, *22*(4), 956–971.
742   https://doi.org/10.1111/mec.12156

743 Hotaling, S., Giersch, J. J., Finn, D. S., Tronstad, L. M., Jordan, S., Serpa, L. E., ... & Weisrock, D. W.
744   (2019). Congruent population genetic structure but differing depths of divergence for three alpine
745   stoneflies with similar ecology and geographic distributions. *Freshwater Biology*, *64*(2), 335-347.
746   https://doi.org/10.1111/fwb.13223

747 Huber, P. R., Greco, S. E., & Thorne, J. H. (2010). Spatial scale effects on conservation network design:
748   Trade-offs and omissions in regional versus local scale planning. *Landscape Ecology*, *25*(5), 683–
749   695. https://doi.org/10.1007/s10980-010-9447-4

750 Hubert, M., Reynkens, T., Schmitt, E., & Verdonck, T. (2016). Sparse PCA for high-dimensional data
751   with outliers. *Technometrics*, *58*(4), 424–434. https://doi.org/10.1080/00401706.2015.1093962

752 Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust
753   principal component analysis. *Technometrics*, *47*(1), 64–79.

754　　　https://doi.org/10.1198/004017004000000563

755　Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in streams: using DNA to understand the
756　　　movement of freshwater fauna and their riverine habitat. *BioScience*, *59*(7), 573–583.
757　　　https://doi.org/10.1525/bio.2009.59.7.8

758　Husemann, M., Ray, J. W., King, R. S., Hooser, E. A., & Danley, P. D. (2012). Comparative
759　　　biogeography reveals differences in population genetic structure of five species of stream fishes.
760　　　*Biological Journal of the Linnean Society*, *107*(4), 867–885. https://doi.org/10.1111/j.1095-
761　　　8312.2012.01973.x

762　Hutchings, J. A., Myers, R. A., García, V. B., Lucifora, L. O., & Kuparinen, A. (2012). Life□history
763　　　correlates of extinction risk and recovery potential. *Ecological Applications*, *22*(4), 1061-1067.
764　　　https://doi.org/10.1890/11-1313.1

765　Hutchison, D. W., & Templeton, A. R. (1999). Correlation of pairwise genetic and geographic distance
766　　　measures: inferring the relative influences of gene flow and drift on the distribution of genetic
767　　　variability. *Evolution*, *53*(6), 1898-1914. https://doi.org/10.1111/j.1558-5646.1999.tb04571.x

768　Jones, O. R., & Wang, J. (2012). A comparison of four methods for detecting weak genetic structure from
769　　　marker data. *Ecology and Evolution*, *2*(5), 1048-1055. https://doi.org/10.1002/ece3.237

770　Jorgensen, D. J. (1993) Paleohydrology of the Central United States. *U.S. Geological Survey Bulletin*,
771　　　1989-D, 1-32. https://doi.org/10.3133/b1989D

772　Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015–
773　　　4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

774　Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide
775　　　population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*, 208.
776　　　https://doi.org/10.3389/fgene.2015.00208

777　Labonne, J., Ravigné, V., Parisi, B., & Gaucherel, C. (2008). Linking dendritic network structures to
778　　　population demogenetics: the downside of connectivity. *Oikos*, *117*(10), 1479-1490.
779　　　https://doi.org/10.1111/j.0030-1299.2008.16976.x

780　Laikre, L. (2010). Genetic diversity is overlooked in international conservation policy implementation.
781　　　*Conservation Genetics*, *11*(2), 349–354. https://doi.org/10.1007/s10592-009-0037-4

782　Lawson, D. J., Davies, N. M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., Hemani, G., Smith, G.
783　　　D., & Timpson, N. J. (2020). Is population structure in the genetic biobank era irrelevant, a
784　　　challenge, or an opportunity? *Human Genetics*, 139, 23-41. https://doi.org/10.1007/s00439-019-
785　　　02014-8

786　Lee, S. R., Jo, Y. S., Park, C. H., Friedman, J. M., & Olson, M. S. (2018). Population genomic analysis
787　　　suggests strong influence of river network on spatial distribution of genetic variation in invasive
788　　　saltcedar across the southwestern United States. *Molecular Ecology*, *27*(3), 636–646.
789　　　https://doi.org/10.1111/mec.14468

790　Legendre, P., & Fortin, M. J. (2010). Comparison of the Mantel test and alternative approaches for
791　　　detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular
792　　　Ecology Resources*, *10*(5), 831-844. https://doi.org/10.1111/j.1755-0998.2010.02866.x

793　Leonard, P. B., Baldwin, R. F., & Hanks, R. D. (2017). Landscape-scale conservation design across biotic
794　　　realms: Sequential integration of aquatic and terrestrial landscapes. *Scientific Reports*, *7*(1), 1–12.
795　　　https://doi.org/10.1038/s41598-017-15304-w

796　Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine,
797　　　V., Maxwell, S., Moidu, H., Tan, F., & Thieme, M. (2019). Global hydro-environmental sub-basin
798　　　and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 1–15.
799　　　https://doi.org/10.1038/s41597-019-0300-6

800　Lowe, W. H., Likens, G. E., McPeek, M. A., & Buso, D. C. (2006). Linking direct and indirect data on
801　　　dispersal: Isolation by slope in a headwater stream salamander. *Ecology*, *87*(2), 334–339.
802　　　https://doi.org/10.1890/05-0232

803　Mantel, N., & Valand, R. S. (1970). A technique of nonparametric multivariate analysis. *Biometrics*,
804　　　*26*(3), 547. https://doi.org/10.2307/2529108

805 Margules C.R., & Pressey, R.L. (2000) Systematic conservation planning. *Nature, 405*, 243–253.
806   https://doi.org/10.1038/35012251
807 Matthews, W.J. (1998). *Patterns in Freshwater Fish Ecology*. New York: Chapman and Hall.
808 Matthews, W. J., & Robison, H. W. (1988). The distribution of the fishes of Arkansas: A multivariate
809   analysis. *Copeia*, *1988*(2), 358-374. https://doi.org/10.2307/1445876
810 Matthews, W. J., & Robison, H. W. (1998). Influence of drainage connectivity, drainage area and regional
811   species richness on fishes of the interior highlands in Arkansas. *American Midland Naturalist*,
812   *139*(1), 1–19. https://doi.org/10.1674/0003-0031(1998)139[0001:IODCDA]2.0.CO;2
813 Mayden, R. L. (1985). Biogeography of Ouachita Highland fishes. *The Southwestern Naturalist*, *30*(2),
814   195–211. https://doi.org/10.2307/3670734
815 Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater
816   fishes. *Systematic Biology*, *37*(4), 329–355. https://doi.org/10.1093/sysbio/37.4.329
817 McRae, B.H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-
818   3820.2006.tb00500.x
819 Meffe, G.K., & Vrijenhoek, R. C. (1988). Conservation genetics in the management of desert fishes.
820   *Conservation Biology*, *2*(2), 157–169. https://doi.org/10.1111/j.1523-1739.1988.tb00167.x
821 Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846.
822   https://doi.org/10.1111/j.1365-294X.2012.05578.x
823 Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures.
824   *Molecular Ecology Resources*, *11*(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x
825 Naimi, B. (2013). R package usdm: Uncertainty analysis for species distribution models. Retrieved from:
826   https://cran.r-project.org/package=usdm
827 Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for
828   population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology
829   Resources*, *17*(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654
830 Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National
831   Academy of Sciences*. https://doi.org/10.1073/pnas.70.12.3321
832 Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, *98*(3),
833   625–640. https://doi.org/10.1093/genetics/98.3.625
834 Oaks, J. R. (2019). Full Bayesian comparative phylogeography from genomic data. *Systematic Biology*,
835   68(3), 371-395. https://doi.org/10.1093/sysbio/syy063
836 Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R.,
837   O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2020). R
838   package vegan: community ecology package. Retrieved from https://cran.r-
839   project.org/package=vegan
840 Olden, J. D., Poff, N. L., & Bestgen, K. R. (2008). Trait synergisms and the rarity, extirpation, and
841   extinction risk of desert fishes. *Ecology*, *89*(3), 847-856. https://doi.org/10.1890/06-1864.1
842 Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population
843   genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and
844   isolation by colonization. *Molecular ecology*, *22*(24), 5983-5999. https://doi.org/10.1111/mec.12561
845 Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach.
846   *Bioinformatics*, 26(3), 419-420. https://doi.org/10.1093/bioinformatics/btp696
847 Pavlova, A., Beheregaray, L. B., Coleman, R., Gilligan, D., Harrisson, K. A., Ingram, B. A., Kearns, J.,
848   Lamb, A. M., Lintermans, M., Lyon, J., Nguyen, T. T. T., Sasaki, M., Tonkin, Z., Yen, J. D. L., &
849   Sunnucks, P. (2017). Severe consequences of habitat fragmentation on genetic diversity of an
850   endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications*,
851   *10*(6), 531–550. https://doi.org/10.1111/eva.12484
852 Paz⬜Vinas, I., & Blanchet, S. (2015). Dendritic connectivity shapes spatial patterns of genetic diversity:
853   A simulation⬜based study. *Journal of Evolutionary Biology*, *28*(4), 986-994.
854   https://doi.org/10.1111/jeb.12626
855 Paz-Vinas, I., Loot, G., Hermoso, V., Veyssière, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018).

Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877). https://doi.org/10.1098/rspb.2017.2746

Paz⬚Vinas, I., Loot, G., Stevens, V. M., & Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, *24*(18), 4586-4604. https://doi.org/10.1111/mec.13345

Pearson, R. G., Stanton, J. C., Shoemaker, K. T., Aiello-Lammens, M. E., Ersts, P. J., Horning, N., Fordham, D. A., Raxworthy, C. J., Ryu, H. Y., McNees, J., & Akçakaya, H. R. (2014). Life history and spatial traits predict extinction risk due to climate change. *Nature Climate Change*, *4*(3), 217-221. https://doi.org/10.1038/nclimate2113

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*(4), 974-997. https://doi.org/10.1016/j.csda.2004.06.015

Peres-Neto, P.R., Legendre, P., Dray, S., Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, *87*(10), 2614-2625. https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2

Perez, M. F., Franco, F. F., Bombonato, J. R., Bonatelli, I. A. S., Khan, G., Romeiro-Brito, M., Fegies, A. C., Ribeiro, P. M., Silva, G. A. R., & Moraes, E. M. (2018). Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*, 24(12), 1883-1889. https://doi.org/10.1111/ddi.12816

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Peterson, E. E., Ver Hoef, J. M., Isaak, D. J., Falke, J. A., Fortin, M. J., Jordan, C. E., McNyset, K., Monestiez, P., Ruesch, A. S., Sengupta, A., Som, N., Steel, E. A., Theobald, D. M., Torgersen, C. E., & Wenger, S. J. (2013). Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters*, *16*(5), 707–719. https://doi.org/10.1111/ele.12084

Pilger, T. J., Gido, K. B., Propst, D. L., Whitney, J. E., & Turner, T. F. (2017). River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. *Molecular Ecology*, *26*(10), 2687–2697. https://doi.org/10.1111/mec.14079

Prunier, J. G., Poesy, C., Dubut, V., Veyssière, C., Loot, G., Poulet, N., & Blanchet, S. (2020). Quantifying the individual impact of artificial barriers in freshwaters: A standardized and absolute genetic index of fragmentation. *Evolutionary Applications*, *13*(10), 2566-2581. https://doi.org/10.1111/eva.13044

Radinger, J., & Wolter, C. (2015). Disentangling the effects of habitat suitability, dispersal, and fragmentation on the distribution of river fishes. *Ecological Applications*, *25*(4), 914–927. https://doi.org/10.1890/14-0422.1

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358. https://www.jstor.org/stable/25049339

Riginos, C., Buckley, Y. M., Blomberg, S. P., & Treml, E. A. (2014). Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *The American Naturalist*, *184*(1), 52–64. https://doi.org/10.1086/676505

Roberts, J. H., Angermeier, P. L., & Hallerman, E. M. (2013). Distance, dams and drift: What structures populations of an endangered, benthic stream fish? *Freshwater Biology*, *58*(10), 2050–2064. https://doi.org/10.1111/fwb.12190

Robison, H. W., & Buchanan, T. M. (2020). *Fishes of Arkansas* (2nd ed.). Fayetteville, AR: University of Arkansas Press.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, *1*(6), 0660–0671. https://doi.org/10.1371 /journal.pgen.0010070

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1006/bijl.1996.0127

Ruiz-Gonzalez, A., Cushman, S. A., Madeira, M. J., Randi, E., & Gómez-Moliner, B. J. (2015). Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, *24*(20), 5110–5129. https://doi.org/10.1111/mec.13392

Ruzich, J., Turnquist, K., Nye, N., Rowe, D., & Larson, W. A. (2019). Isolation by a hydroelectric dam induces minimal impacts on genetic diversity and population structure in six fish species. *Conservation Genetics*, *20*(6), 1421–1436. https://doi.org/10.1007/s10592-019-01220-1

Schmidt, B. V., & Schaefer, J. (2018). Ecological and landscape effects on genetic distance in an assemblage of headwater fishes. *Ecology of Freshwater Fish*, *27*(2), 617–631. https://doi.org/10.1111/eff.12375

Shelley, J. J., Holland, O. J., Swearer, S. E., Dempster, T., Le, M. C., Craig, F., & Miller, A. D. (2021). Landscape context and dispersal ability as determinants of population genetic structure in freshwater fishes. *Freshwater Biology*, 67(2), 338-352. https://doi.org/10.1111/fwb.13844

Singhal, S., Huang, H., Grundler, M. R., Marchán-Rivadeneira, M. R., Holmes, I., Title, P. O., Donnellan, S. C., & Rabosky, D. L. (2018). Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *The American Naturalist*, *192*(4), 432–447. https://doi.org/10.1086/699515

Strange, R. M., & Burr, B. M. (1997). Intraspecific phylogeography of North American highland fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, *51*(3), 885–897. https://doi.org/10.1111/j.1558-5646.1997.tb03670.x

Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution*, *70*(3), 731–739. https://doi.org/10.1111/evo.12883

Turner, T. F., & Robison, H. W. (2006). Genetic diversity of the Caddo Madtom, *Noturus taylori*, with comments on factors that promote genetic divergence in fishes endemic to the Ouachita Highlands. *Southwestern Naturalist*, *51*(3), 338–345. https://doi.org/10.1894/0038-4909(2006)51[338:GDOTCM]2.0.CO;2

Tyers, M. (2017). Riverdist: River network distance computation and applications. Retrieved from: https://cran.r-project.org/package=riverdist

USGS (2021). Watershed Boundary Dataset. Retrieved from: https://www.usgs.gov/national-hydrography/watershed-boundary-dataset. Accessed: 1 December 2021.

USGS, & USDA-NRCS. (2013). Federal standards and procedures for the national Watershed Boundary Dataset (WBD). In *U.S. Geological Survey, Techniques and Methods* (4th ed.).

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649-5662. https://doi.org/10.1111/mec.12938

Waples, R. S. (2022). What is N_e, anyway? *Journal of Heredity*, *113*(4), 371-379. https://doi.org/10.1093/jhered/esac023

Warren, M. L., Burr, B. M., Walsh, S. J., Bart, H. L., Cashner, R. C., Etnier, D. A., Freeman, B. J., Kuhajda, B. R., Mayden, R. L., Robison, H. W., Ross, S. T., & Starnes, W. C. (2000). Diversity, distribution, and conservation status of the native freshwater fishes of the Southern United States. *Fisheries*, *25*(10), 7–31. https://doi.org/10.1577/1548-8446(2000)025<0007:ddacso>2.0.co;2

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

White, S. L., Hanks, E. M., & Wagner, T. (2020). A novel quantitative framework for riverscape genetics. *Ecological Applications*, *30*(7), 1–17. https://doi.org/10.1002/eap.2147

Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by Fst do not necessarily require large sample sizes when using many snp markers. *PLoS ONE*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Winter, D. J. (2012). MMOD: An R library for the calculation of population differentiation statistics.

958   *Molecular Ecology Resources*, *12*(6), 1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x

959   Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138.

960   Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J.
961   (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure,
962   as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus*
963   *californicus*). *Molecular Ecology*, *27*(10), 2347-2364. https://doi.org/10.1111/mec.14589

964   Xuereb, A., D'Aloia, C.C., Andrello, M., Bernatchez, L., & Fortin, M.-J. (2021) Incorporating putatively
965   neutral and adaptive genomic data into marine conservation planning. *Conservation Biology*, *35*(3),
966   909–920. https://doi.org/10.1111/cobi.13609

967   [dataset] Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022a).Ozark Fish
968   Community Genomics: ddRAD seq. *NCBI sequence read archive*. BioProject: PRJNA809538.

969   Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022b). A community genomics
970   approach to natural hybridization. *bioRxiv*. https://doi.org/10.1101/2022.10.26.513841

971   [dataset] Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022c). White River Fish
972   Genetic Structure. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/R8YTU

973   Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022d). Trait-based predictors of
974   genomic diversity within freshwater fish communities of the Ozarks. *bioRxiv*.
975   https://doi.org/10.1101/2022.10.26.513843

976   Zbinden, Z. D., Geheber, A. D., Lehrter, R. J., & Matthews, W. J. (2022). Multifaceted assessment of
977   stream fish alpha and beta diversity using spatial models. *Hydrobiologia*, *849*, 1795–1820.
978   https://doi.org/10.1007/s10750-022-04824-w

979   Zbinden, Z. D., Geheber, A. D., Matthews, W. J., & Marsh-Matthews, E. (2022). Fish communities,
980   species of greatest conservation need, and protected areas in southeastern Oklahoma, 2014-2016.
981   *Proceedings of the Oklahoma Academy of Science*, *101*, 14–32.

982   Zbinden, Z. D., & Matthews, W. J. (2017). Beta diversity of stream fish assemblages: partitioning
983   variation between spatial and environmental factors. *Freshwater Biology*, *62*(8), 1460–1471.
984   https://doi.org/10.1111/fwb.12960

CONFLICT OF INTEREST

986   The authors declare that they have no competing interests.

988   **AUTHOR CONTRIBUTIONS**

ZDZ conceived the research with input from all authors. Specimen collection was done by ZDZ & TKC.

990   ZDZ did laboratory work, bioinformatics, data analysis, and manuscript drafting. All authors contributed to interpretation of results, formulating conclusions, and critically revising the manuscript. MRD and

992   MED administered funding through their University of Arkansas Endowments.

994   **DATA AVAILABILITY STATEMENT**

Raw sequence reads are deposited in the SRA (BioProject PRJNA809538). Species-level SNP panel

996   alignments, metadata, and R code are available on Open Science Framework (White River Fish Genetic Structure; DOI 10.17605/OSF.IO/R8YTU).

998

1000   **BENEFIT-SHARING STATEMENT**

No benefits to report.

TABLE 1 Fish species (*N*=31) were collected at 75 sampling locations across the White River Basin of the Ozark Mountains, U.S.A. Summary data are tabulated for *N*=2,861 individuals across seven families genotyped and analyzed for genetic structure. Family=fish family; Species=species name; NI=number of individuals analyzed after filtering; NS=number of sites at which filtered individuals occurred; NI/S=mean number of individuals per site; Reads=mean number of raw reads recovered by Illumina HiSeq; Loci=mean number of loci recovered by iPyrad; Depth=mean coverage of loci; Ho=mean observed heterozygosity; SNPs=number of single nucleotide polymorphisms in the analyzed data panel; Miss=mean missing data; and PCs=number of principal components used to characterize neutral genetic variation and PCvar=the original genetic variance explained by the set of PCs.

Family	Species	NI	NS	NI/S	Reads	Loci	Depth	Ho	SNPs	Miss	PCs	PCvar
Atherinopsidae	*Labidesthes sicculus*	99	18	5.5	2401513	19532	83	0.0013	2956	0.11	17	40.2
Centrarchidae	*Lepomis macrochirus*	63	17	3.7	2369445	26142	61	0.0028	5873	0.14	19	45.5
	Lepomis megalotis	242	44	5.5	2330434	25126	59	0.0036	4841	0.13	48	45.2
	Micropterus dolomieu	56	15	3.7	2014858	21420	58	0.0018	2813	0.11	11	32.6
	Micropterus salmoides	15	7	2.1	2338155	22827	65	0.0018	2825	0.06	7	59.4
Cottidae	*Cottus carolinae*	24	9	2.7	2973760	27523	74	0.0012	5798	0.12	5	61.6
	Cottus hypselurus	40	8	5.0	3226846	28108	76	0.0015	7116	0.11	5	75.1
Fundulidae	*Fundulus catenatus*	112	23	4.9	2757508	30509	52	0.0014	3378	0.13	18	46.0
	Fundulus olivaceus	131	24	5.5	2647685	27631	51	0.0025	3111	0.14	22	42.6
Leuciscidae	*Campostoma anomalum*	93	20	4.7	2226556	16753	77	0.0036	3187	0.13	10	36.7
	Campostoma oligolepis	119	31	3.8	2038589	16107	76	0.0030	3121	0.12	40	44.7
	Chrosomus erythrogaster	53	7	7.6	2180045	16508	73	0.0033	3440	0.14	6	55.8
	Cyprinella galactura	72	10	7.2	1648530	14839	72	0.0029	3322	0.11	27	52.1
	Cyprinella whipplei	29	6	4.8	1870427	14599	84	0.0033	2847	0.12	8	39.5
	Luxilus chrysocephalus	57	13	4.4	1677176	15089	68	0.0025	2168	0.14	17	47.2
	Luxilus pilsbryi	244	31	7.9	2028625	16063	81	0.0033	4922	0.14	93	52.1
	Luxilus zonatus	98	16	6.1	2273167	16964	89	0.0030	5496	0.12	12	24.7
	Lythrurus umbratilis	23	5	4.6	1970516	16465	68	0.0032	2491	0.12	6	40.3
	Notropis boops	233	28	8.3	2355581	15684	104	0.0040	6161	0.11	71	43.8
	Notropis nubilus	191	32	6.0	2087695	15544	81	0.0040	4018	0.14	65	46.3
	Notropis percobromus	62	10	6.2	2082050	17852	74	0.0047	4393	0.13	36	65.6
	Notropis telescopus	81	13	6.2	2092015	16154	85	0.0024	4741	0.11	12	31.2
	Pimephales notatus	47	13	3.6	2106907	15271	92	0.0029	4022	0.13	11	49.3
	Semotilus atromaculatus	30	9	3.3	2216336	15406	84	0.0020	2644	0.15	2	63.6
Percidae	*Etheostoma blennioides*	52	14	3.7	2491915	21416	71	0.0024	5124	0.11	2	36.4
	Etheostoma caeruleum	358	50	7.2	2170268	21900	62	0.0044	3511	0.13	20	28.7
	Etheostoma flabellare	22	6	3.7	2288120	21041	62	0.0015	9927	0.08	4	88.7
	Etheostoma juliae	57	10	5.7	2513876	20652	84	0.0014	5473	0.1	7	39.5
	Etheostoma spectabile	49	10	4.9	2565769	23873	64	0.0051	5519	0.15	6	33.6
	Etheostoma zonale	74	15	4.9	2364158	21514	74	0.0033	10033	0.13	5	24.9
Poeciliidae	*Gambusia affinis*	35	8	4.4	2657603	24021	78	0.0021	3818	0.09	9	39.9
	MEAN	92.3	16.8	5.1	2289230.0	20081.7	73.6	0.0028	4486.7	0.12	20.0	46.2
	STDEV	80.8	11.2	1.5	341159.5	4697.6	12.0	0.0010	1931.1	0.02	22.1	14.3

1012 **TABLE 2** Summary of genetic structure observed for $N=31$ species of fish collected across the White
1013 River Basin, U.S.A. Classifications to family and species are provided for each, along with summaries of
1014 genetic structure: H_T=total heterozygosity; H_S=within-site heterozygosity; G_{ST}=Nei's fixation index; G''
1015 $_{ST}$=unbiased fixation index; D=Jost's genetic differentiation; IBD=significant tests of isolation by distance
1016 denoted "X"; Structure=whether the species could be subdivided into more than one population, denoted
1017 "X"; Model=the isolation model explaining the most individual genetic variance; and Model Var=the
1018 amount of variance explained by the best isolation model. Species are ordered by Jost's D.

1019

Family	Species	H_T	H_S	G_{ST}	G''_{ST}	D	IBD	Structure	Model	Model Var
Percidae	*Etheostoma flabellare*	0.35	0.02	0.93	0.96	0.40	-	X	stream hierarchy	99%
Leuciscidae	*Semotilus atromaculatus*	0.30	0.09	0.70	0.79	0.26	X	X	stream hierarchy	91%
Cottidae	*Cottus hypselurus*	0.24	0.07	0.73	0.81	0.22	-	X	stream hierarchy	99%
Leuciscidae	*Chrosomus erythrogaster*	0.27	0.11	0.59	0.71	0.21	X	X	stream hierarchy	98%
Cottidae	*Cottus carolinae*	0.26	0.11	0.58	0.69	0.19	X	X	stream hierarchy	93%
Leuciscidae	*Campostoma anomalum*	0.20	0.12	0.38	0.45	0.09	X	X	stream hierarchy	87%
Percidae	*Etheostoma blennioides*	0.21	0.13	0.35	0.43	0.09	X	X	stream hierarchy	98%
Leuciscidae	*Pimephales notatus*	0.25	0.18	0.28	0.36	0.09	X	X	stream hierarchy	98%
Percidae	*Etheostoma juliae*	0.23	0.16	0.29	0.37	0.09	X	X	stream hierarchy	97%
Leuciscidae	*Lythrurus umbratilis*	0.30	0.25	0.17	0.27	0.09	-	-	stream hierarchy	69%
Percidae	*Etheostoma spectabile*	0.20	0.14	0.31	0.38	0.08	X	X	stream hierarchy	99%
Fundulidae	*Fundulus olivaceus*	0.24	0.18	0.25	0.32	0.08	X	X	stream hierarchy	88%
Fundulidae	*Fundulus catenatus*	0.20	0.14	0.31	0.37	0.07	X	X	stream hierarchy	83%
Atherinopsidae	*Labidesthes sicculus*	0.18	0.14	0.24	0.29	0.05	X	X	stream hierarchy	84%
Leuciscidae	*Notropis telescopus*	0.20	0.16	0.20	0.25	0.05	X	X	stream hierarchy	60%
Percidae	*Etheostoma caeruleum*	0.14	0.10	0.27	0.30	0.04	X	X	stream hierarchy	90%
Percidae	*Etheostoma zonale*	0.16	0.13	0.20	0.25	0.04	X	X	stream hierarchy	98%
Leuciscidae	*Luxilus chrysocephalus*	0.26	0.23	0.11	0.15	0.04	X	X	stream hierarchy	38%
Centrarchidae	*Lepomis megalotis*	0.18	0.15	0.17	0.21	0.04	X	X	stream hierarchy	47%
Poeciliidae	*Gambusia affinis*	0.26	0.24	0.10	0.14	0.04	X	X	stream hierarchy	59%
Leuciscidae	*Cyprinella whipplei*	0.26	0.24	0.09	0.14	0.04	X	X	stream hierarchy	50%
Centrarchidae	*Micropterus salmoides*	0.30	0.28	0.06	0.10	0.03	X	-	stream hierarchy	12%
Leuciscidae	*Luxilus zonatus*	0.19	0.17	0.11	0.14	0.03	-	X	stream hierarchy	76%
Centrarchidae	*Lepomis macrochirus*	0.24	0.22	0.07	0.10	0.02	-	-	stream hierarchy	19%
Centrarchidae	*Micropterus dolomieu*	0.23	0.22	0.07	0.10	0.02	X	-	stream hierarchy	57%
Leuciscidae	*Notropis boops*	0.17	0.16	0.06	0.08	0.01	X	X	stream hierarchy	23%
Leuciscidae	*Notropis nubilus*	0.14	0.13	0.07	0.08	0.01	X	X	stream hierarchy	13%
Leuciscidae	*Campostoma oligolepis*	0.17	0.16	0.05	0.06	0.01	X	X	stream hierarchy	15%
Leuciscidae	*Cyprinella galactura*	0.18	0.18	0.04	0.05	0.01	-	-	stream hierarchy	12%
Leuciscidae	*Notropis percobromus*	0.18	0.18	0.03	0.04	0.01	X	-	stream hierarchy	3%
Leuciscidae	*Luxilus pilsbryi*	0.14	0.13	0.02	0.02	0.00	X	-	stream hierarchy	6%
	MEAN	0.22	0.16	0.25	0.30	0.08				63%
	STDEV	0.05	0.06	0.23	0.25	0.09				35%

1020 **TABLE 3** Summary of correlation among population genetic parameter estimates calculated for $N=31$
1021   fish species collected across the White River Basin, U.S.A. H_S=within-site heterozygosity; H_T=total
1022   heterozygosity; G_{ST}=Nei's fixation index; G''_{ST}=unbiased fixation index; and D=Jost's genetic
1023   differentiation. Pearson's product-moment correlation between each parameter estimate is shown in the
1024   table below. Only significant ($\alpha < 0.05$) correlations are shown.

1025

	H_S	H_T	G_{ST}	G''_{ST}
H_T	ns	-		
G_{ST}	-0.75	0.52	-	
G''_{ST}	-0.71	0.55	0.99	-
D	-0.65	0.67	0.97	0.96

TABLE 4 Genetic variation of fish species (*N*=31) sampled across the White River Basin (Ozark Mountains, U.S.A.), was tested using analysis of molecular variance (AMOVA) to determine the proportion of genetic variation differing among distinct hydrologic units (HUCs) and among discrete population clusters. HUC tests were performed at four HUC-levels (4-, 6-, 8-, and 10-digit HUCs) and the level depicting the most genetic variance is shown. Var=percent genetic variance explained; sig=the significant of the test (* for <0.05 and ns for >0.05); Φ_{ST} = genetic variation among sites across all groups; Φ_{SC} = genetic variation among sites within a group.

| Family | Species | HUC-level | Hydrologic Units | | | | | | Population Clusters | | | | | |
| | | | Among HUCs | | Among Sites | | | | Among Pops | | Among Sites | | | |
			%var	sig.	%var	Φ_{ST}	sig.	Φ_{SC}	%var	sig.	%var	Φ_{ST}	sig.	Φ_{SC}
Atherinopsidae	*Labidesthes sicculus*	HUC-8	21%	*	19%	0.40	*	0.24	25%	*	18%	0.436	*	0.243
Centrarchidae	*Lepomis macrochirus*	-	0%	ns	7%	0.07	*	0.07	-	-	-	-	-	-
	Lepomis megalotis	HUC-4	70%	*	7%	0.77	*	0.23	37%	*	6%	0.428	*	0.098
	Micropterus dolomieu	HUC-8	5%	*	7%	0.12	*	0.07	-	-	-	-	-	-
	Micropterus salmoides	HUC-4	3%	*	0%	0.02	ns	0.00	-	-	-	-	-	-
Cottidae	*Cottus carolinae*	HUC-8	66%	*	9%	0.74	*	0.26	62%	*	15%	0.772	*	0.402
	Cottus hypselurus	HUC-8	84%	ns	5%	0.89	ns	0.31	85%	ns	7%	0.917	*	0.442
Fundulidae	*Fundulus catenatus*	HUC-8	36%	*	15%	0.51	*	0.23	36%	*	16%	0.516	*	0.244
	Fundulus olivaceus	HUC-8	18%	*	18%	0.36	*	0.22	16%	*	21%	0.372	*	0.252
Leuciscidae	*Campostoma anomalum*	HUC-8	53%	*	2%	0.55	*	0.05	61%	*	7%	0.680	*	0.175
	Campostoma oligolepis	HUC-8	6%	*	1%	0.07	ns	0.01	5%	*	3%	0.081	*	0.036
	Chrosomus erythrogaster	-	-	-	-	-	-	-	62%	*	21%	0.829	*	0.548
	Cyprinella galactura	HUC-8	7%	*	0%	0.07	ns	0.00	-	-	-	-	-	-
	Cyprinella whipplei	HUC-8	14%	*	4%	0.18	*	0.05	14%	ns	7%	0.202	*	0.078
	Luxilus chrysocephalus	HUC-8	14%	*	7%	0.21	*	0.08	17%	*	10%	0.266	*	0.120
	Luxilus pilsbryi	HUC-10	1%	ns	1%	0.02	*	0.01	-	-	-	-	-	-
	Luxilus zonatus	HUC-10	15%	*	3%	0.18	*	0.03	9%	*	10%	0.199	*	0.115
	Lythrurus umbratilis	-	0%	ns	22%	0.20	*	0.22	-	-	-	-	-	-
	Notropis boops	HUC-8	6%	*	3%	0.09	*	0.03	6%	*	6%	0.113	*	0.059
	Notropis nubilus	HUC-4	10%	*	7%	0.17	*	0.08	16%	*	1%	0.172	*	0.015
	Notropis percobromus	HUC-8	1%	*	1%	0.01	ns	0.01	-	-	-	-	-	-
	Notropis telescopus	HUC-8	33%	*	1%	0.34	*	0.01	41%	*	3%	0.436	*	0.046
	Pimephales notatus	HUC-8	17%	*	26%	0.44	*	0.32	13%	*	32%	0.453	*	0.372
	Semotilus atromaculatus	HUC-8	87%	ns	1%	0.88	*	0.08	92%	*	2%	0.934	*	0.194
Percidae	*Etheostoma blennioides*	HUC-8	61%	*	2%	0.62	*	0.04	67%	*	2%	0.686	*	0.053
	Etheostoma caeruleum	HUC-8	40%	*	3%	0.44	*	0.06	45%	*	5%	0.497	*	0.093
	Etheostoma flabellare	-	0%	ns	99%	0.98	*	0.98	95%	*	3%	0.977	ns	0.580
	Etheostoma juliae	HUC-8	34%	*	11%	0.45	*	0.16	36%	*	12%	0.478	*	0.182
	Etheostoma spectabile	HUC-8	29%	*	10%	0.38	*	0.14	26%	*	13%	0.394	*	0.181
	Etheostoma zonale	HUC-8	32%	*	2%	0.34	*	0.02	38%	*	5%	0.422	*	0.074
Poeciliidae	*Gambusia affinis*	HUC-4	7%	*	13%	0.20	*	0.14	13%	ns	11%	0.239	*	0.123

FIGURE 1 Fish were sampled at *N*=75 locations across the White River Basin (Ozark Mountains, U.S.A.). The study basin is contained within the larger Mississippi River Basin, and is a direct tributary to the mainstem Mississippi. The study region is subdivided into five subbasins: Upper White, Lower White, Black, Little Red, and the St. Francis. Beyond these basins, USGS Hydrologic Unit Codes (HUCs) were also used to characterize the stream hierarchy position of sampling locations (4-, 6-, 8-, and 10-digit HUCs).
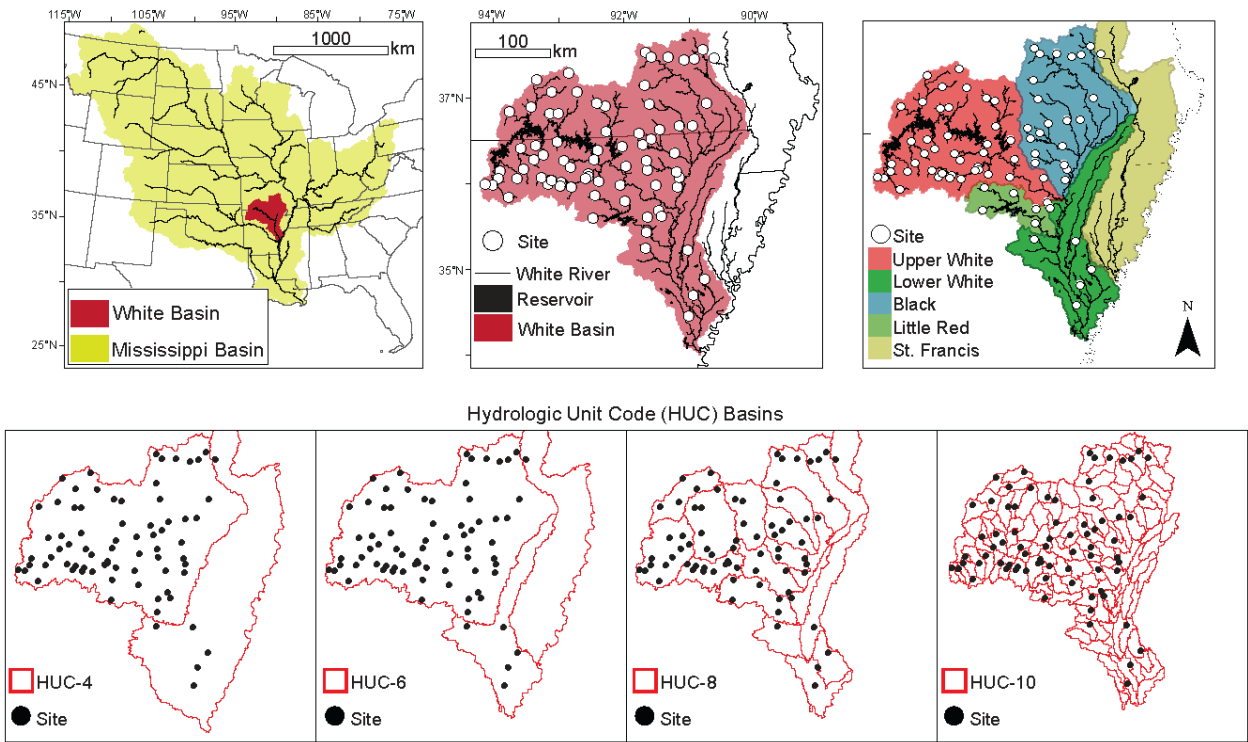
FIGURE 2 The analytic approach to partitioning individual genetic variation across four spatio-environmental predictor matrices. The approach was applied separately to 31 freshwater fish species collected across the White River Basin of the Ozarks, USA. The full redundancy analysis model is shown at the top of the figure, where genetic diversity is explained by geographic distance (IBD), discrete population structure (IBB), stream hierarchical position (IBH), and environmental variation among habitats (IBE). The initial data matrices representing genetic β-diversity (i.e., response variable) and the four explanatory variables sets are depicted at the top. Each matrix is labeled to show rows, columns, and values (e.g., individuals, single nucleotide polymorphisms, and alleles). These matrices each pass through analyses and/or transformations (gray ellipses) to yield the matrices used for modeling at the bottom of the figure. PCA = principal component analysis; PCs = principal components; dbMEM = distance-based Moran's eigenvector maps; sNMF = sparse non-negative matrix factorization; dummy transform = transforming categorical variable into separate binary variables; VIF = variance inflation factor; ROBPCA = robust principal components analysis. Note that environmental factors were standardized (z-score).
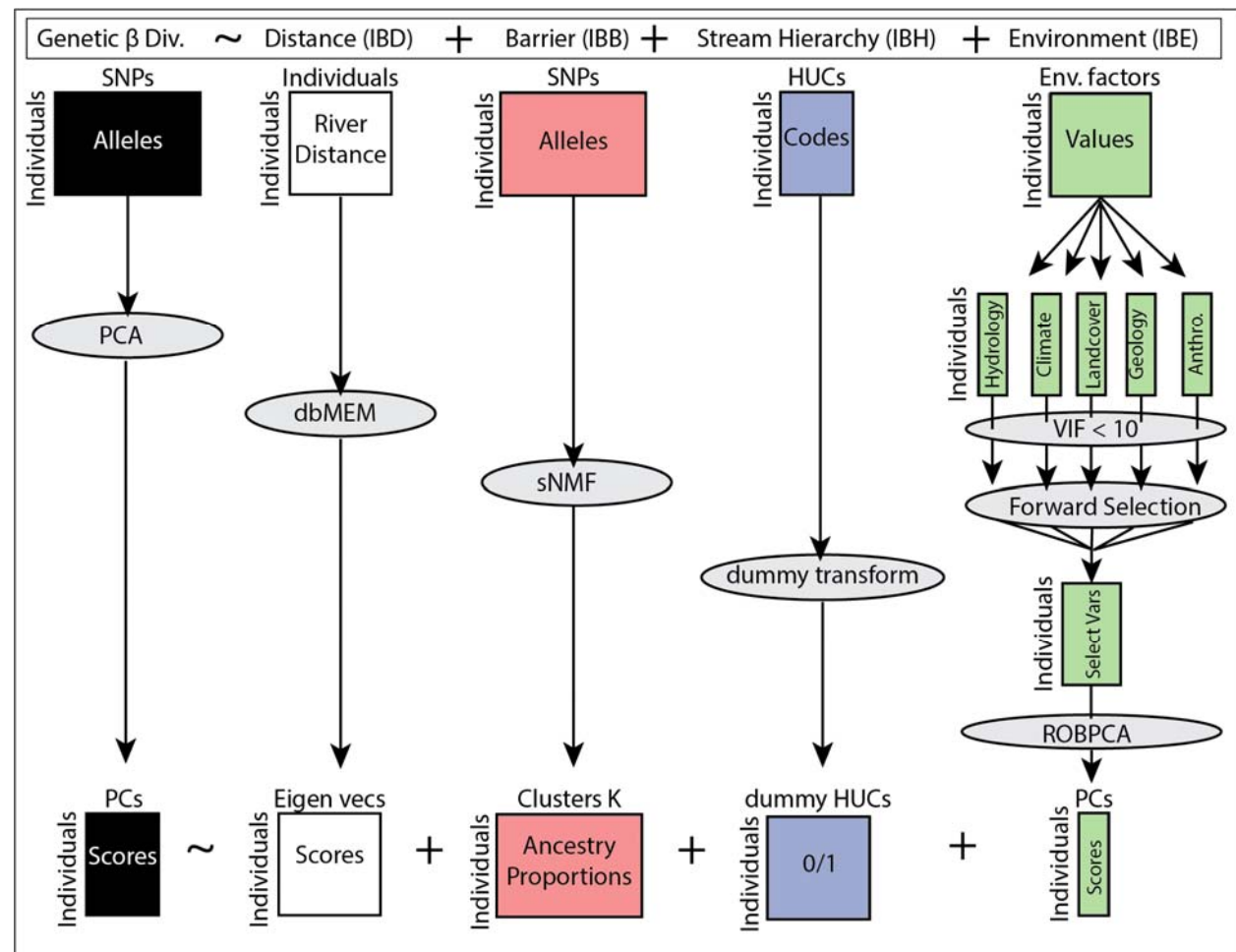
FIGURE 3 Genetic structure of *N*=31 fish species collected across the White River Basin (Ozark Mountains, U.S.A.) as summarized by among-site F_{ST} (Weir and Cockerham's θ) and Jost's *D*. Boxplots show the distributions of both pairwise estimates among sampling sites for each species. Inner quantiles are colored to indicate species in the same family (*N*=7).
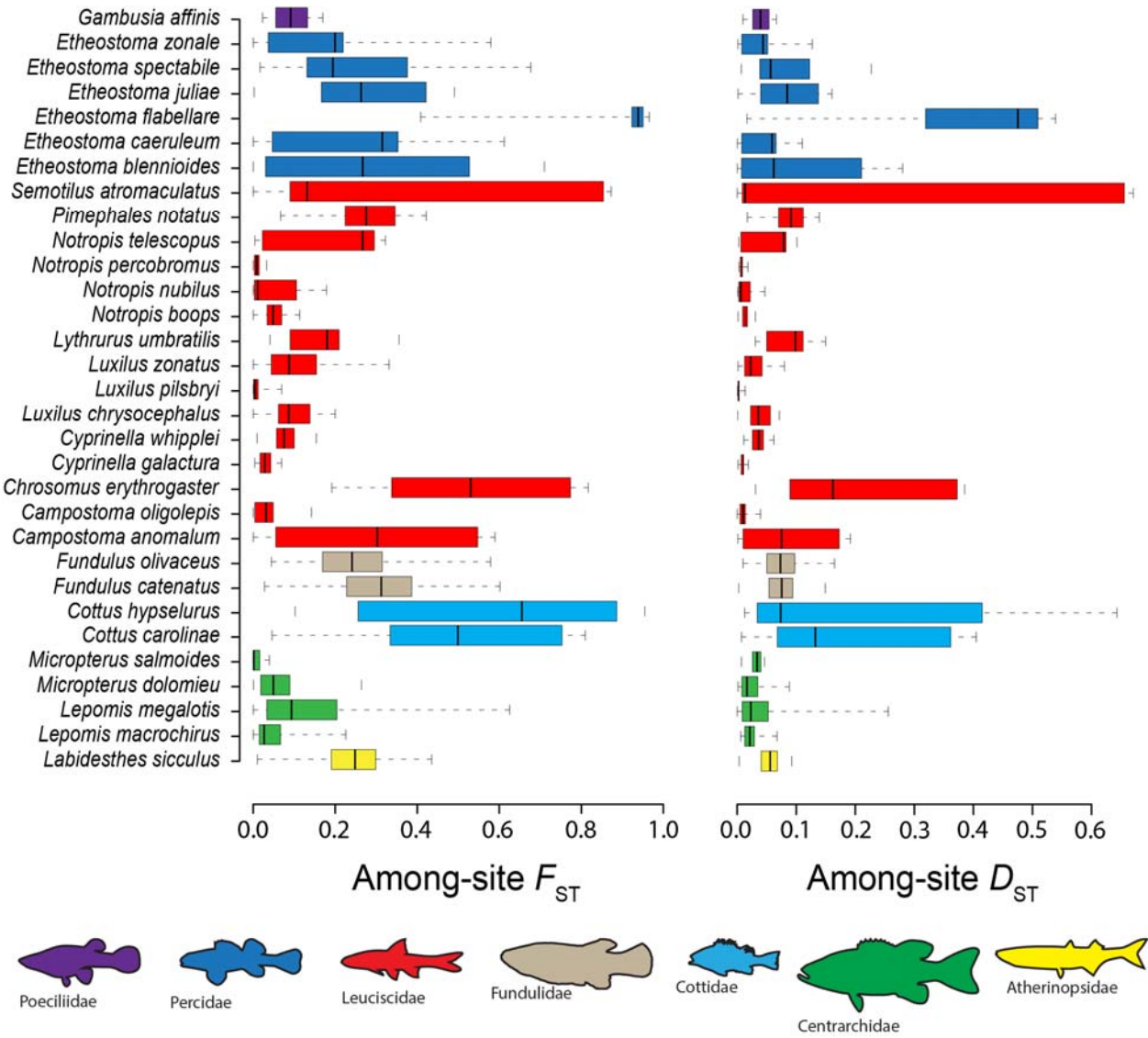
1061 **FIGURE 4** Isolation by distance plots for N=31 fish species collected across the White River Basin (
1062 Ozark Mountains, U.S.A.). Each depicts the relationship between among-site F_{ST} (linearized) and log river
1063 distance among sites. The following are represented below each species name: m=slope of the linear
1064 regression model (dashed red line) and r= the Mantel coefficient indicating the strength of the correlation
1065 between genetic structure and distance. Significant r-values denoted with a red asterisk ($\alpha \leq$ 0.05).

1066

1067

41
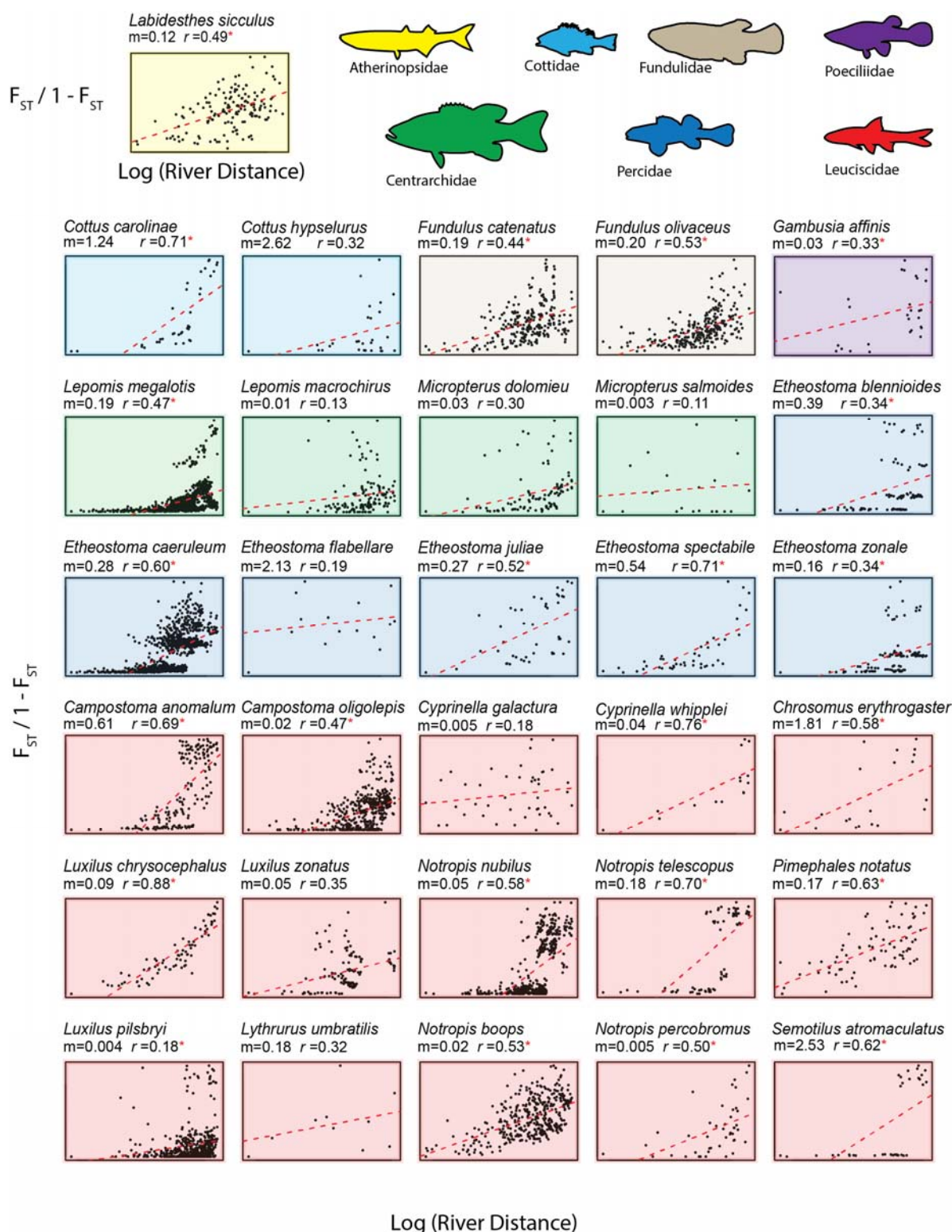
FIGURE 5 Sampling distribution maps of seven species which showed no evidence of discrete genetic population structure within the White River Basin (Ozark Mountains, U.S.A.). A total of *N*=31 species were sampled across 75 sites. The number of collection sites (red circles) for each species is denoted by N; K=the number of discrete genetic populations discerned from sparse non-negative matrix factorization.
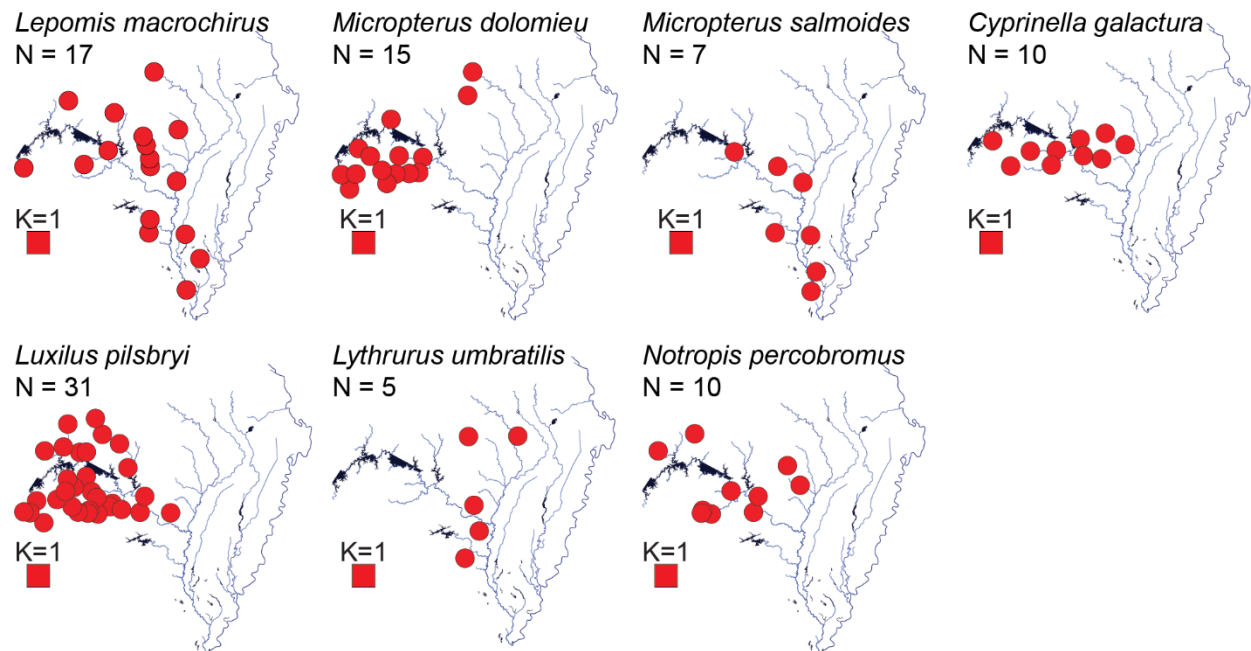
1074   **FIGURE 6** Sampling distribution maps of 24 species which showed evidence of genetic population structure
1075   within the White River Basin (Ozark Mountains, U.S.A.). *N*=31 species were sampled across 75 sites. K= the
1076   number of discrete genetic populations discerned from sparse non-negative matrix factorization. Sampling
1077   sites are denoted as pie charts representing the average population coefficients for each site. N= number of
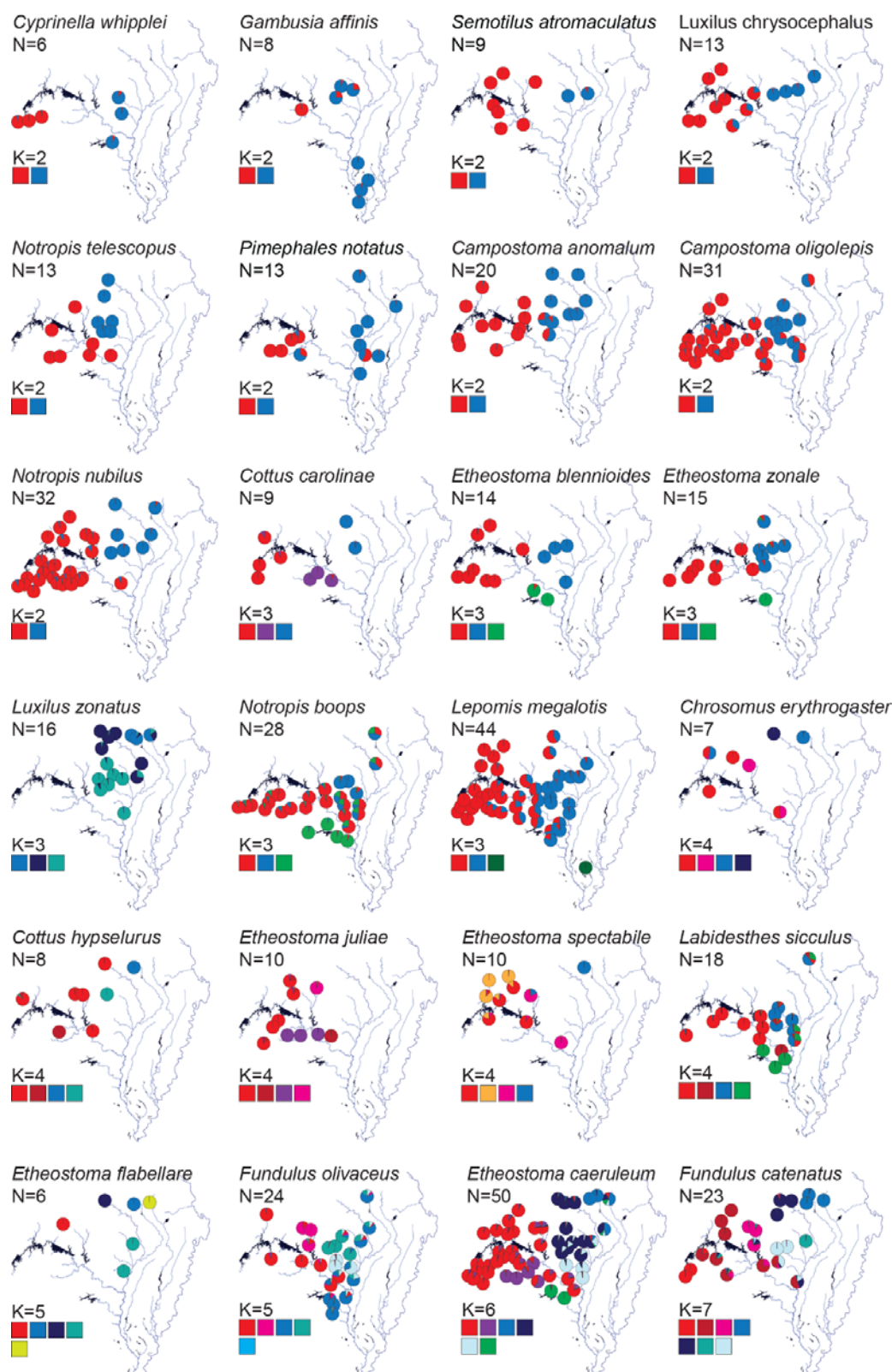1078   sites where each species was collected.

1079

1080

1081 **FIGURE 7** Neutral genetic variation was partitioned between four explanatory models for *N*=31 fish
1082 species sampled across the White River Basin (Ozark Mountains, U.S.A.). Partitioning was conducted
1083 separately for each species. The four models represent: (i) isolation by *distance*, the river network
1084 distance among individuals represented by spatial eigenvectors; (ii) isolation by *barrier*, represented by
1085 population structure coefficients among individuals; (iii) isolation by *stream hierarchy*, based on the
1086 hydrologic units (at four different hierarchical levels) in which an individual was collected; and (iv)
1087 isolation by *environment*, characterized by the environmental heterogeneity across sampling sites where
1088 individuals were collected. Total = the genetic variation explained by all four models combined. The
1089 "Pure" models represent the variation explained by each model after partialling out the variation
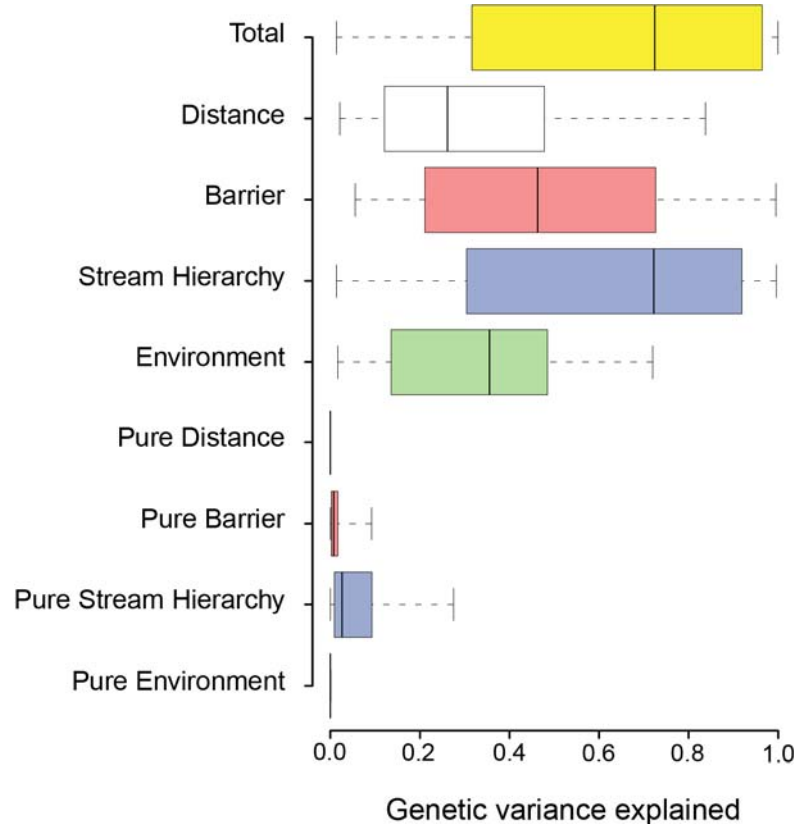1090 explained by the other three models.
1091

1092

1093 **FIGURE 8** Venn diagrams depict neutral genetic variation resulting from four models as applied to *N*=31
1094 fish species sampled from the White River Basin (Ozark Mountains, U.S.A.). Models were based on: (i)
1095 isolation by distance, isolation by barrier, isolation by stream hierarchy, and isolation by environment.
1096 Values in the Venn diagrams are percent of genetic variance explained (i.e., rounded adjusted R^2 values).
1097 Total variance explained is shown below each diagram. The bottom two rows show species that showed
1098 no discrete population structure (i.e., no isolation by barrier) and thus only three of the models were
1099 tested.

1100

45