# Riverscape community genomics: A comparative analytical framework to identify common drivers of spatial structure

Running title: Riverscape Community Genomics

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

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

[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

[*]Corresponding Author: zdzbinde@uark.edu

## ABSTRACT

Genetic diversity is a key component of population persistence. However, most genetic investigations of natural populations focus on a single species, overlooking opportunities for multispecies conservation plans to benefit entire communities in an ecosystem. We developed a framework to evaluate genomic diversity within and among many species and demonstrate how this *riverscape community genomics* approach can be applied to identify common drivers of genetic structure. Our study evaluated genomic diversity in 31 co-distributed native stream fishes sampled from 75 sites across the White River Basin (Ozark Mountains, USA) using SNP genotyping (ddRAD). Despite variance in genetic divergence, general spatial patterns were identified corresponding to river network architecture. Most species ($N$=24) were partitioned into discrete sub-populations ($K$=2–7). We used partial redundancy analysis to compare species-specific genomic diversity across four models of genetic structure: Isolation by distance (IBD), isolation by barrier (IBB), isolation by stream hierarchy (IBH), and isolation by environment (IBE). A significant proportion of intraspecific genetic variation was explained by IBH ($x$□=62%), with the remaining models generally redundant. Our results indicated that gene flow is higher within rather than between hierarchical units (i.e., catchments, watersheds, basins), supporting the Stream Hierarchy Model and its generality. We discuss our conclusions regarding conservation and management and identify the 8-digit Hydrologic Unit (HUC) as the most relevant spatial scale for managing genetic diversity across riverine networks.

**KEYWORDS:**

Comparative population genetics; Fish; Hydrologic units; Population structure; RADseq; Stream hierarchy model

## 1 | INTRODUCTION

Genetic diversity is a quantitative metric applied across spatial and temporal scales (Huber et al., 2010; Leonard et al., 2017) tied to the evolutionary trajectories of species (Shelley et al., 2021). It also serves as a barometer for population-level persistence in accurately reflecting demographic history, connectivity, and adaptive potential (Davis et al., 2018; DeWoody et al., 2021; Paz-Vinas et al., 2018). Surprisingly, and despite its many accolades, genetic diversity is often underutilized in conservation planning (Laikre, 2010; Paz-Vinas et al., 2018), in part due to a suite of affiliated necessities (i.e., specialized equipment, technical expertise, and required externalities such as genomics centers), all of which expand its bottom line (Blanchet et al., 2020). Moreover, when assessment does occur, it is most often limited to populations within a single species or a small cadre of entities within a species-group, thus minimizing the potential for much-needed generalizations (Anthonysamy et al., 2018).

When the concept of genetic diversity is applied in a comparative sense across co-distributed species, it provides a solid framework from which community-wide management and policy can be defined. For example, multispecies assessments can reveal common dispersal barriers (Pilger et al., 2017; Roberts et al., 2013), congruent distributions of genetic diversity (Hotaling et al., 2019; Ruzich et al., 2019), relevant spatial scales for management (Blanchet et al., 2020), and associations among species characteristics and genetic diversity (Bohonak, 1999; Pearson et al., 2014). Despite its potential complexity, a comprehensive management strategy can emerge, one more appropriately aligned towards managing numerous species, with long-term conservation goals beneficial to an entire community (Blanchet et al., 2017). In addition, it also tacitly encourages support by stakeholders for an overarching management plan, one representing a consensus across multiple species and ecosystems (Douglas et al., 2020).

The spatial structure of genetic variation is primarily dictated by gene flow and genetic drift within a species (Holderegger et al., 2006), with the uniformity of its distribution (i.e., panmixia; Rosenberg et al., 2005) serving as an implicit null hypothesis. The *de facto* alternative is that genetic variation is spatially autocorrelated (i.e., isolation by distance, IBD; Wright, 1943). For most species, a

significant relationship between genetic dissimilarity and geographic distance is the expectation (Meirmans, 2012), yet the strength of this association will vary (Bohonak, 1999; Singhal et al., 2018). For example, genetic divergence may be promoted by environmental dissimilarities (i.e., isolation by environment, IBE; Wang & Bradburd, 2014) or by physical barriers to dispersal (i.e., isolation by barrier, IBB; Cushman et al., 2006; Ruiz-Gonzalez et al., 2015).

For aquatic biodiversity, patterns of genetic divergence will also be governed by the structure and architecture of the riverine network (in contemporary and historic representations). Organisms within such dendritic networks are demonstrably impacted by the physical structure of the habitat (Peterson et al., 2013; White et al., 2020), with genetic relatedness as a surrogate for the underlying structural hierarchy (Hughes et al., 2009). While this is most apparent within the contemporary structure of river networks, their historic structure, i.e., paleohydrology, also serves to bookmark genetic diversity (Mayden, 1988; Strange & Burr, 1997). Moreover, the hierarchical complexity of these networks will likewise dictate population processes, as reflected within genetic diversities and divergences (Chiu et al., 2020; Hopken et al., 2013; Thomaz et al., 2016). Thus, spatial genetic structuring within such biodiversity should reflect isolation by stream hierarchy (IBH; *sensu* Stream Hierarchy Model (SHM); Meffe & Vrijenhoek, 1988). The initial genesis for the SHM was narrowly defined within desert stream fishes of the American West (Meffe and Vrijenhoek, 1988). An assessment of its generality, as compared to alternative isolating regimes, was thus imperative (Brauer et al., 2018; Hopken et al., 2013).

The factors that cause genetic structure can be confounding on the one hand (Perez et al., 2018) but also correlated on the other (Meirmans, 2012; Wang & Bradburd, 2014). Different mechanisms can mask the occurrence of major drivers by promoting those more ancillary with regard to single-species assessments. The emerging results are twofold: Potentially erroneous conclusions, which in turn beget ineffective management strategies. These issues can be effectively mitigated using replicated multispecies assessments to allow influential major processes to surface, thus effectively categorizing both 'signal and noise' components with the former driving patterns of regional biodiversity (Roberts et al., 2013).

Hypotheses relating to genetic structure are best contrasted by partitioning available genetic variation via partial redundancy analysis (Borcard et al., 1992; Chan & Brown, 2020), thus allowing the contrast of multiple alternative models. In turn, the best-performing model should be substantially correlated with other (more redundant) models but also provide the best explanation for residual variation once competing models adequately explain antecedent variability (Cushman et al., 2006). If alternative models explain significant amounts of genetic variation, then the null hypothesis of panmixia would be rejected. The main drivers of genetic diversity should then emerge as comparisons are made across the community's many species. This approach also allows the appropriate scale to be defined at which genetic and conservation perspectives can be integrated to optimize benefits across species.

Our objective was to establish a framework from which the generality of the SHM could be tested across constituents of a riverscape fish community. This framework would allow key drivers to be identified, with a concurrent expectation of common processes re-emerging within these ecological networks as the analysis was processed. We accomplish this by comparing patterns of genetic diversity across 31 fish species within the White River Basin of the Ozark Mountains (AR/MO, USA). For each, we contrasted four alternative models (Cushman et al., 2013) representing major drivers of genetic structure: Isolation by distance (IBD), isolation by stream hierarchy (IBH), isolation by barrier (IBB), and isolation by environment (IBE). Our data represent thousands of SNPs (single nucleotide polymorphisms), as derived via recent advances in high-throughput sequencing (Peterson et al., 2012). This has, in turn, allowed thousands of individuals to be genotyped as a financially and logistically practical research endeavor across multiple non-model species (da Fonseca et al., 2016). We offer our approach as a potential blueprint for developing more comprehensive genetic management plans at the community level.

## 2 | MATERIALS AND METHODS

### 2.1 | Study system

Our study system, the White River Basin, is located within the Western Interior Highlands of North America, a previous component of the more extensive pre-Pleistocene Central Highlands extending north and east but subsequently subsumed by numerous glacial advances into two disjunct sub-components: Western Interior Highlands (i.e., Ozark Plateau, Ouachita Mountains), and Eastern Highlands (i.e., Appalachian Plateau, Blue Ridge, Appalachian Highlands) (Mayden, 1985). The Ozark Plateau remained an unglaciated refugium with elevated endemism and diversity (Warren et al., 2000). The White River Basin was established by at least Late Pliocene (>3 MYA; Jorgensen, 1993), but its eastern tributaries were captured by the Mississippi River when it bisected the basin during the Pleistocene (Mayden, 1988; Strange & Burr, 1997). This paleohydrologic signature may remain in contemporary patterns of population divergence in the White River Basin, as manifested by replicated patterns of genetic structure between eastern and western populations.

### 2.2 | Sampling

The sampling region for our study is composed of the White River and St. Francis River basins (AR/MO) (Figure 1). Both are tributaries to the Mississippi River, draining 71,911 $km^2$ and 19,600 $km^2$, respectively. Five sub-basins are apparent: St. Francis, Upper White, Black, Lower White, and Little Red rivers (Figure 1). These are further subdivided into the following hierarchical Hydrologic Units (HUC) (USGS & USDA-NRCS, 2013; USGS, 2021) representing different spatial scales: HUC-4 Subregions (*N*=2); HUC-6 Basins (*N*=3); HUC-8 Subbasins (*N*=19); HUC-10 Watersheds (*N*=129) (Figure 1).

Sampling was approved by the University of Arkansas Institutional Animal Care and Use Committee (IACUC: #17077), with collecting permits as follows: Arkansas Game & Fish Commission (#020120191); Missouri Department of Wildlife Conservation (#18136); US National Parks Service (Buffalo River Permit; BUFF-2017-SCI-0013). Fishes were sampled using seine nets in wadable streams during low flow between June 2017 and September 2018. Time spent sampling a site ranged from 30–60

mins, with a target of 5-10 individuals/species encountered. Individuals were euthanized by immersion in tricaine methanesulfonate (MS-222) at a concentration of 500 mg/L, buffered to pH=7 with subsequent preservation in 95% ethanol. Formal species diagnosis occurred in the laboratory, and the right pectoral fin was removed from each specimen and stored in 95% ethanol at -20 °C prior to subsequent DNA extraction. Specimens are housed at the Arkansas Conservation and Molecular Ecology Lab, University of Arkansas, Fayetteville.

### 2.2 | Genomic data collection and filtering

Genomic DNA was isolated (Qiagen Fast kits; Qiagen Inc.) and quantified by fluorometry (Qubit; Thermo-Fisher Scientific). Individuals were genotyped using double-digest restriction site-associated DNA (ddRAD) sequencing (Peterson et al., 2012), with procedures modified appropriately (Chafin et al., 2019). Standardized DNA amounts (1,000 ng) were digested at 37°C with high-fidelity restriction enzymes *Msp*I (5'-CCGG-3') and *Pst*I (5'-CTGCAG-3') (New England Biosciences), bead-purified (Ampure XP; Beckman-Coulter Inc.), standardized to 100 ng, and then ligated with custom adapters containing in-line identifying barcodes (T4 Ligase; New England Biosciences). Samples were pooled in sets of 48 and size-selected from 326-426 bp, including adapter length (Pippin Prep; Sage Sciences). Illumina adapters and i7 index were added via 12-cycle PCR with Phusion high-fidelity DNA polymerase (New England Biosciences). Three libraries (3x48=144 individuals/lane) were pooled per lane and single-end sequenced on the Illumina HiSeq 4000 platform (1x100bp; Genomics & Cell Characterization Core Facility; University of Oregon, Eugene). Quality control checks, including fragment analysis and quantitative real-time PCR, were performed at the core facility before sequencing.

Raw Illumina reads were demultiplexed, clustered, filtered, and aligned in IPYRAD v.0.9.62 (Eaton & Overcast, 2020). Reads were first demultiplexed, allowing up to one barcode mismatch, yielding individual FASTQ files containing raw reads ($N$=3,060 individual files). Individuals averaged >2 million reads, with those extremely low removed ( $< x\square - 2s$) to reduce errors from poor quality sequencing. Individuals were screened for putative hybrids (Zbinden, Douglas, et al., 2022), and those

with admixed ancestry were removed. Raw sequence reads were partitioned by species (*N*=31) and aligned *de novo* in IPYRAD (Eaton & Overcast, 2020). Adapters/primers were removed, and reads with >5 bases having Phred quality <20 or read length <35 bases (after trimming) were discarded. Clusters of homologous loci were assembled using an 85% identity threshold. Putative homologs were removed if any of the following were met: <20x and >500x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of nucleotides polymorphic; >8 indels present; or presence in <15% of individuals. Paralogs were identified (and subsequently removed) as those clusters exhibiting either >2 alleles per site in consensus sequence or excessive heterozygosity (>5% of consensus bases or >50% heterozygosity/site among individuals).

Biallelic SNP panels for each species were then visualized and filtered with the R package RADIATOR (Gosselin, 2020). To ensure high data quality, loci were removed if: Monomorphic; minor allele frequency <3%; Mean coverage <20 or >200; Missing data >30%; SNP position on read >91; and if HWE lacking in one or more sampling sites ($\alpha = 0.0001$). To reduce linkage disequilibrium, only one SNP per locus was retained (that which maximized minor allele count). Finally, singleton individuals/species at a sampling site and those with >75% missing data in the filtered panel were removed.

### 2.3 | Genetic structure

Genetic structure was assessed using the resultant SNP genotypes. For each species (*N*=31), pairwise $F_{ST}$ (Weir & Cockerham, 1984) was calculated among sites (HIERFSTAT; Goudet et al., 2017). Jost's *D* was also quantified among sites and globally, as it is based on the effective number of alleles rather than heterozygosity and hence less biased by sampling differences (Jost, 2008). Additional global intraspecific $F_{ST}$ analogs were also quantified for comparison: Multi-allelic $G_{ST}$ (Nei, 1973) and unbiased $G''_{ST}$ (Meirmans & Hedrick, 2011) (MMOD; Winter, 2012). We tested for isolation by distance (IBD) using both linearized $F_{ST}$ and Jost's *D*. Their relationships with river distance (log-transformed) were assessed using

the Mantel test (Mantel & Valand, 1970) (ECODIST; Goslee & Urban, 2020), then visualized using linear regression (Rousset, 1997).

Admixture analysis of population structure and ancestry coefficients were estimated using sparse non-negative matrix factorization (sNMF) (Frichot et al., 2014). We ran sNMF for each species, with 20 repetitions per *K* value (1 to *N* sites or 20, whichever was smallest) and $\alpha$=100 (LEA; Frichot & François, 2015). The best *K* (i.e., number of distinct gene pools) from each sNMF run minimizes the cross-validation entropy criterion (Alexander & Lange, 2011). The best *K* was then used to impute missing data (*impute* function using method='mode' in LEA). The sNMF algorithm was then repeated (as above) using imputed genotypes. The resulting Q-matrices of ancestry coefficients were used to map population structure and served as the "IBB" (isolation by barrier) model below.

We further assessed among-site genetic variation between Hydrologic Units (HUCs) and discrete population clusters (determined via sNMF) using analysis of molecular variance (AMOVA) (Excoffier et al., 1992). AMOVA was performed for each species at four HUC levels (4-, 6-, 8-, and 10-digit) to compare the amount of genetic variation among HUCs, among all sites, and among sites within HUCs. The Watershed Boundary Dataset (USGS, 2021) assigned HUC classifications to each site. AMOVA was then performed for each species with genetic clusters $K>1$ to compare the amount of genetic variation among populations, among all sites, and among sites within populations. The variance components were used to estimate $\Phi$-statistics (analogous to *F*-statistics): $\Phi_{CT}$ = the genetic variation among groups (either HUCs or discrete populations); $\Phi_{ST}$ = the genetic variation among sites across all groups; and $\Phi_{SC}$ = the genetic variation among sites within groups. The wrapper R package POPPR (Kamvar et al., 2015) was used to implement the PEGAS (Paradis, 2010) version of AMOVA with default settings.

### 2.4 | Modelling genetic structure

We employed a variation partitioning framework (Capblancq & Forester, 2021; Chan & Brown, 2020) to compare four models of genetic structure for each species based on: IBD, IBB, IBH, and IBE. Individual genetic variation within each species was reduced to major axes of variation using principal components

analysis (PCA) on each SNP panel. The appropriate number of PCs retained for each species was based on observed eigenvalues, Rnd-Lambda (Peres-Neto et al., 2005), implemented in the R package PCDIMENSION (Coombes & Wang, 2019). Individual scores on the retained PCs represented individual genetic variation.

The first model (IBD) relied on river network distance measured between individuals (RIVERDIST; Tyers, 2017). The resulting distance matrix was then decomposed into positively correlated spatial eigenvectors using distance-based Moran's eigenvector maps (Chan & Brown, 2020) within the R package ADESPATIAL (Dray et al., 2020).

The second model (IBB) was based on individual population coefficients, i.e., population structure, from the Q-matrix generated above using sNMF. The assumption was that population structure indicates a reduction of gene flow between discrete populations due to a barrier (or high resistance) to dispersal. Note: This model could not be incorporated for species in which population structure was not apparent ($K$=1), and these species were thus tested using only three models.

The third model (IBH) was constructed using four levels of HUCs (4-, 6-, 8-, and 10-digit) that characterized an individual's position within the stream hierarchy, i.e., hydrologic unit (USGS, 2021). We transformed the data matrix of individuals by HUC so that each unique HUC was represented at each corresponding level as a binary 'dummy' variable.

The fourth model (IBE) relied on contrasting environmental variation across sites that harbored individuals. Environmental variables were taken from a compendium of 281 factors related to five major classes: hydrology/physiography, climate, land cover, geology/soil composition, and anthropogenic impact (HYDRORIVERSv.1.0; Linke et al., 2019). Variables for each site were extracted prior to being separated into the five major classes, with invariant factors and those exhibiting collinearity being removed in a stepwise manner (USDM; Naimi, 2013) until each had a variation inflation factor (VIF) <10. Standardization occurred by subtracting means and dividing by standard deviations. Variables within each class were selected for subsequent analyses using forward selection (Blanchet et al., 2008).

In summary: Variables were first tested for a relationship with the response data (individual genetic variation) using redundancy analysis (RDA). If the relationship was significant ($\alpha < 0.05$), a stepwise forward procedure was carried out such that variables were selected if the adjusted $R^2$ of the model increased significantly ($\alpha < 0.05$) and the adjusted $R^2$ did not exceed that of the overall model. This procedure was employed using the *ordiR2step* function in the R package VEGAN (Oksanen et al., 2020). The selected variables from each of the five classes were combined into a single matrix, then reduced to a set of PCs using robust principal components analysis (ROBPCA; Hubert et al., 2005). The number of PCs retained for each category was determined following Hubert and coworkers (2005), as implemented in the R package ROSPCA (Hubert et al., 2016).

Individual genetic variation (a matrix of PCs for each species) was then partitioned among the four explanatory models of genetic structure (Partial redundancy analysis; Anderson & Legendre, 1999; Capblancq & Forester, 2021). This allowed an estimation of individual genetic variation explained by each model, all models combined, and then each "pure" model after partitioning out variability explained by the other three. This allows the correlation structure among competing models to be visualized as redundant relationships.

## 3 | RESULTS

### 3.1 | Sampling and data recovery summarized

Collections ($N$=75; Figure 1) yielded $N$=72 species and $N$=3,605 individuals. On average, we collected ~11 species/site, typical for streams sampled with seine nets in North America (Matthews, 1998) and similar highland streams within the Mississippi Basin (Zbinden, Geheber, Lehrter, & Matthews, 2022; Zbinden, Geheber, Matthews, Marsh-Matthews, 2022).

We genotyped $N$=3,060 individuals across $N$=31 species, with at least two individuals collected at ≥5 sampled sites. Simulations and empirical evaluations underscore the accuracy of $F_{ST}$ estimates when large numbers of SNPs (≥1,500) are employed across a minimum of two individuals (Nazareno et al., 2017; Willing et al., 2012). After removing samples with missing data >75% and those as singletons of

their species at a site, the remaining $N$=2,861 were analyzed for genetic structure (Table 1). The number of individuals analyzed per species ranged from 15–358 ($x\bar{}$=92.3; $s$=80.8), and the sites at which each species was collected ranged from 5–50 ($x\bar{}$=16.8; $s$=11.2). Number of individuals/species/site ranged from 2–15 ($x\bar{}$=5.1; $s$=1.5). Mean number of raw reads/individual/species spanned from 1.65 million to 3.22 million ($x\bar{}$=2,289,230.0; $s$=341,159.5). Mean N of loci/species recovered by IPYRAD ranged from 14,599–30,509 ($x\bar{}$=20,081.7; $s$=4,697.6) with a mean sequencing depth/locus of 73.6x ($s$=12.0x). After filtering loci and retaining one SNP per locus, the panels for each species contained 2,168–10,033 polymorphic sites ($x\bar{}$=4,486.7; $s$=1,931.1) with mean missing data/species at 12% ($s$=2%).

### 3.2 | Genetic structure

#### 3.2.1 | Among-site genetic divergence

Distributions of among-site $F_{ST}$ and $D$ varied widely among species (Figure 2), as did global indices of genetic divergence (Table 2). All three global indices of fixation or genetic divergence ($G_{ST}$, $G''_{ST}$, $D$) were negatively correlated with within-site heterozygosity ($H_S$), positively correlated with total heterozygosity ($H_T$), and highly, positively correlated with each other (Table 3).

Regarding IBD, a significant relationship was found between linearized among-site $F_{ST}$ and log-transformed among-site river network distance for 23 (74%) of the $N$=31 species (Figure 3). Mantel coefficients ranged from 0.11–0.88 ($x\bar{}$=0.51; $s$=0.19). The slope of the linear relationship between $F_{ST}$ and distance for each species ranged from 0.003–2.62 ($x\bar{}$=0.46; $s$=0.76). Results were largely similar when IBD was tested with Jost's $D$, again with the same 23 species showing a significant relationship, along with two additional taxa: Smallmouth Bass (*Micropterus dolomieu*; Lacepède, 1802) and Largemouth Bass [*Micropterus salmoides*; (Lacepède, 1802)]. Mantel correlation coefficients ranged from 0.15–0.92 ($x\bar{}$=0.51; $s$=0.19). The slope of the linear relationship between Jost's $D$ and log river network distance for each species ranged from 0.0007–0.28 ($x\bar{}$=0.04; $s$=0.06).

### 3.2.2 | Population structure

An apparent lack of discrete genetic structure emerged across seven species, suggesting continuous structuring at the spatial scale of our study (Figure 4). For the remaining 24 species, at least two and up to seven discrete sub-populations were identified (Figure 5). This structure corresponded at the broadest hierarchical level to the two major northern basins: Upper White and Black rivers, for all species sampled in both sub-basins ($N$=22). There was also evidence of fine-scale structure for five species within the Little Red River Basin. Smaller catchments with distinct gene pools across multiple species included: North Fork (4 spp.), Buffalo (3 spp.), Upper Black (4 spp.), Current (3 spp.), and Spring rivers (4 spp.).

### 3.2.3 | AMOVA

Discrete genetic structuring was also supported via AMOVA. Genetic variation among HUCs was significant for 24 species (Table 4). The genetic variance explained for these species by HUCs ranged from 1–70% ($\bar{x}$=25.0%; $s$=20.7%). For the other seven species, variation among HUCs was ≤ 1%, save for Ozark Sculpin (*Cottus hypselurus*; Robins & Robison, 1985) and Creek Chub [*Semotilus atromaculatus*; (Mitchill, 1818)]. HUC differences for these accounted for >80% of the genetic variance but were non-significant due to a lack of power. Southern Redbelly Dace [*Chrosomus erythrogaster*; (Rafinesque, 1820)] could not be tested due to a lack of repeated samples within HUC levels. Further evidence of genetic structure among HUCs was revealed in the pattern of $\Phi_{SC}$ (genetic divergence among sites within HUCs) < $\Phi_{ST}$ (divergence among all sites) found across 26 species. The 8-digit HUC level explained the greatest genetic variance across 21 species (Table 4).

Genetic variation among discrete population clusters (based on sNMF) was significant for 21 of the $N$=31 species (Table 4). Seven species were best described as single populations ($K$=1) and were therefore not tested further. For those exhibiting structure, the genetic variance among clusters ranged from 5–95% ($\bar{x}$=38.0%; $s$=26.5%). The three species without significant structure, despite K>1 via sNMF, could likely be explained by low power resulting from a small number of sample sites. Again, as

with HUCs, $\Phi_{SC} < \Phi_{ST}$ was observed. However, all tested species showed this pattern (i.e., sites within the same population were less differentiated than sites across all populations).

### 3.3 | Models of genetic structure

Variability in genetic diversity was partitioned across four models of genetic structure for the $N$=31 species. Principal components of SNP panel variation served as representatives of genetic variation. Across species, the number of genetic PCs ranged from 2–93 ($x\square$=20.0; $s$=20.1; Table 1). Cumulative genomic variance explained ranged from 24.7–88.7% ($x\square$=46.2%; $s$=14.3%; Table 1).

Combining the four models (IBD, IBB, IBH, IBE) accounted for between 3–100% of the genomic diversity across species ($x\square$=63.0%; $s$=35.3%; Figure 6). Isolation by stream hierarchy (IBH; $x\square$=62.0%; $s$=34.7%) and barrier (IBB; $x\square$=49.3%; $s$=30.0%) contributed most to the total variation explained, while distance (IBD; $x\square$=32.1%; $s$=25.1%) and environment (IBE; $x\square$=33.0%; $s$=21.4%) explained less (Figure 6). Variation explained by "pure" models, after accounting for that explained by the other three, was >0 only for stream hierarchy and barrier (Figure 6), suggesting that distance and environment are encapsulated by the former. Indeed, correlative structure among models revealed most genetic variance was explained by stream hierarchy, with the other models largely redundant (Figure 7).

## 4 | DISCUSSION

Genetic diversity is an essential metric for inferring evolutionary processes and guiding conservation. Single-species estimates of genetic diversity are standard given practical constraints, e.g., funding mandates for species of conservation concern. However, adopting a multispecies framework for analyzing genetic diversity could allow for more comprehensive management plans to be developed by focusing on commonalities (rather than differences) among species. The Stream Hierarchy Model (Meffe & Vrijenhoek, 1988) posits that the dispersal of stream-dwelling organisms is more limited between hierarchical units (basins, sub-basins, watersheds) than within. If this model was generalizable, it could determine relevant scales and regions for managing genetic diversity.

Our multispecies approach yielded two salient points: 1) From a macro-perspective, river network topology and complexity are manifested in common patterns of genetic structure across species; and 2) on a finer scale, the degree of intraspecific genetic divergence varies widely among co-distributed species. Most species showed significant IBD patterns but also discrete population sub-structure, as reflected most strongly by sub-basin delineations (e.g., HUC-8). These patterns were corroborated by AMOVA and variance partitioning and are generalized across species. Overall, stream fish genetic structure was indicative of dispersal limited primarily among *versus* within river catchments.

## 4.1 | Drivers of isolation at the basin-wide scale

### 4.1.1 *Isolation by Distance and river networks*

IBD is expected when a genetic study's spatial extent is greater than individuals' average dispersal distance, i.e., distance moved from natal habitat to breeding habitat. Indeed, significant IBD patterns were detected in 81% of the species in our study. However, the strength of the relationship was generally weak (Mantel $r$ =0.47 & 0.51 for linearized $F_{ST}$ and $D$, respectively).

While IBD may primarily explain genetic variation along a single stream or river, i.e., linear scale, it fails to incorporate the spatial structure of riverine networks (Thomaz et al., 2016). Therefore, IBD may not be an appropriate general model for fish genetic structure at the network scale (Hopken et al., 2013). IBD plots for many species (Figure 3) showed high genetic divergence even among relatively proximate localities, with apparent clusters indicating discrete rather than continuous structure (Guillot et al., 2009). This evidence suggests that — at the network scale — a more nuanced pattern occurs, with high residual variation resulting. The failure of IBD to account for large amounts of variation in genetic divergence reflects additional resistance to dispersal, as caused by longitudinal changes in habitat characteristics such as slope, depth, volume, and predator composition. For example, two river reaches of equal length can have very different habitat matrices, and these can be more influential on gene flow than space alone (Guillot et al., 2009; Lowe et al., 2006; Ruiz-Gonzalez et al., 2015).

366 4.1.2 *Stream Hierarchy Model*

367 Our results show that individual genetic variation is best explained by the Stream Hierarchy Model
368 (Brauer et al., 2018; Hopken et al., 2013; Meffe & Vrijenhoek, 1988). In other words, the majority of
369 variation explained by IBD, IBE, and IBB could be accounted for by IBH alone. This was corroborated
370 via variation partitioning, in which IBD, IBE, and IBB models were redundant with IBH. A concordance
371 of population structure with stream hierarchy yielded a similar percentage of among-site genetic variation,
372 as explained by among-HUC and among-population groupings. In short, variance explained by distance
373 and environment was due to differences among HUC drainages. These results highlight the necessity of
374 accounting for population structure prior to exploring the relationship between genotypes and
375 environmental heterogeneity, e.g., within genotype by environment frameworks (Lawson et al., 2020).
376
377 4.1.3 *Disentangling cumulative effects*
378 Our analyses also revealed complex spatial patterns of genetic diversity. We evaluated competing
379 isolation models using a framework that identified distance and barriers as putative drivers, with strong
380 genetic divergence identified even across short geographical distances (Chan & Brown, 2020; Ruiz-
381 Gonzalez et al., 2015). This interaction can confound analyses that incorporate either alone. For example,
382 if sampling is clustered, discrete genetic groups can be spuriously inferred along an otherwise continuous
383 gradient of genetic variation (Frantz et al., 2009). Furthermore, a continuous pattern can be erroneously
384 extrapolated when the underlying reality is described by distinct clusters separated by geographic distance
385 (Meirmans, 2012). Here we echo the importance of testing various hypotheses concerning genetic
386 structure (Perez et al., 2018). Idiosyncrasies and complex interactions cannot be discerned by testing
387 single models in isolation (e.g., discrete structure or IBD).
388
389 **4.2 | Drivers of variation within and among species**
390 The species assayed herein display marked differences concerning dispersal capability (Shelley et al.,
391 2021). Given this, we expected genetic structure to widely vary among species across our study region

(Comte & Olden, 2018; Husemann et al., 2012; Pilger et al., 2017). Dispersal-related traits drive gene flow among localities and determine the spatial scale at which patterns of genetic structure emerge (Bohonak, 1999; Riginos et al., 2014). The physical structure of the river network then further modulates these patterns by dictating dispersal pathways of metapopulations and their colonization and extinction probabilities (Falke et al., 2012; Labonne et al., 2008; Fagan, 2002). These superimposed processes promote genetic divergence among distal populations (Thomaz et al., 2016; Chiu et al., 2020). Similar patterns emerge when analyzing community diversity via species composition. Headwater streams tend to have very different communities due to dispersal limitations (Finn et al., 2011; Zbinden & Matthews, 2017; Zbinden, Geheber, Lehrter, & Matthews, 2022). Hence the interaction between traits and environment is an overarching influence that unites ecology and evolution.

Many species studied herein are small-bodied with aggregate distributions in upland and headwater streams (Robison & Buchanan, 2020). Thus, species-specific dispersal limitations, as imposed by unsuitable riverine habitats (Radinger & Wolter, 2015; Schmidt & Schaefer, 2018), explain considerable variation in genetic structuring within the White River. Large rivers are hypothesized as inhospitable habitats to upland fishes (e.g., resources, depth, turbidity, substrates) and impose resistance to successful migration (e.g., higher discharge, greater density of large-bodied predators). These characteristics constrain migration and limit gene flow amongst basins that drain into large rivers (Fluker et al., 2014; Schmidt & Schaefer, 2018; Turner & Robison, 2006). The results are asymmetric gene flow and source-sink metapopulation dynamics, with susceptible species, those smaller and less tolerant, diverging most rapidly (Campbell Grant et al., 2007).

Other life-history traits may also play a role as well. For example, those that directly influence effective population size (Nei & Tajima, 1981) may generate differences among species regarding the rate at which genetic differences arise (Blanchet et al., 2020). Species with 'slow' life histories, characterized by longer generations and delayed maturity, show an increased probability of local extirpation, inflating genetic drift concomitant with global extinction risk (Hutchings et al., 2012; Pearson et al., 2014; Chafin et al., 2019). Similar contingencies exist for other ecological traits, such as highly specialized trophic

418 adaptations, narrow environmental tolerances, or those that follow the same general mechanism by
419 predisposing species to fragmented population structures (Olden et al., 2008). Ecological traits are
420 mirrored by morphology (Douglas & Matthews, 1992), underscoring an interaction of trait effects that are
421 difficult to disentangle. Ultimately, intraspecific genetic divergence is driven by a combination of factors
422 that influence population size, demographic history, and connectivity. Clearly, these complex interactions
423 among drivers require more comparative multispecies assessments as they shape genetic diversity and
424 structure within and among species (microevolutionary scale) and thus ultimately lead to speciation and
425 extinction (macroevolutionary scale). Our analytical framework outlined herein provides a template for
426 such community-genomics studies.
427
428 **4.3 | Disentangling historic and contemporary drivers**
429 4.3.1 *Paleohydrology in the White River system*
430 In this study, discrete population structure coincides with major topological divides within the White
431 River stream network, such as a consistent east/west divide between Upper White and Black rivers,
432 mirroring prior community composition studies (Matthews & Robison, 1988; 1998). Similar patterns
433 were observed at smaller scales among drainages within the study region, as reported for White River
434 crayfish (Fetzner & DiStefano, 2008). While the Lower White and Black rivers are certainly
435 contemporary large-river habitats, both would have been much larger pre-Pleistocene when together they
436 represented the main channel of the Old Mississippi River (Mayden, 1988; Strange & Burr, 1997). This
437 large-river habitat would have separated the eastern and western highland tributaries, with inhospitable
438 habitat for upland species. Pronounced limitations regarding historic dispersal induced by the Old
439 Mississippi could explain the greater isolation of the Little Red River and Black River tributary
440 populations compared to those in the Upper White River. Here, additional work should incorporate
441 coalescent perspectives (e.g., Oaks, 2019) that test the role of past geomorphic events in driving co-
442 divergence and co-demographic patterns, such as the Pleistocene incursion by the Old Mississippi into the
443 modern Black River channel.

444

445 4.3.2 *Contemporary drivers*

446 Spatial discontinuities in genetic structure can also reveal contemporary barriers to migration/gene flow
447 (Lee et al., 2018; Ruiz-Gonzalez et al., 2015). The Upper White River dams (e.g., Norfork, Bull Shoals,
448 Table Rock, and Beaver dams) represent the most apparent anthropogenic barriers to gene flow. Dams
449 elsewhere have demonstrated discrete population structures above and below the structure (Roberts et al.,
450 2013). However, impacts can be limited due to the relatively short period these dams have been in place
451 (Ruzich et al., 2019). Those on the White River were constructed between 1912 (Taneycomo Powersite
452 Dam) and 1966 (Beaver Dam).

453 Our study was not explicitly designed to assess impoundment effects on diversity, nor were they
454 directly tested. Nevertheless, evidence of discrete population structure has emerged, corresponding to the
455 location of such dams. Four species showed discrete populations within the North Fork River above the
456 Norfork Dam: Southern Redbelly Dace [*Chrosomus erythrogaster*; (Rafinesque, 1820)]; Yoke Darter
457 (*Etheostoma juliae*; Meek, 1891); Northern Studfish [*Fundulus catenatus*; (Storer, 1846)]; and
458 Blackspotted Topminnow [*Fundulus olivaceus*; (Storer, 1845)] (sites colored magenta; Figure 5). One
459 species, Orangethroat Darter [*Etheostoma spectabile*; (Agassiz, 1854)], showed a distinct population in
460 the James River above Table Rock Dam (sites colored gold; Figure 5). However, both North Fork and
461 James rivers drain eight-digit HUC watersheds, which explains high amounts of genetic variation across
462 the study region, regardless of dams. This highlights the importance of understanding 'natural' network-
463 wide patterns of genetic structure prior to deriving conclusions regarding anthropogenic barriers,
464 particularly when they coincide with stream hierarchy. Differentiating dams as barriers *versus* stream
465 hierarchy could be accomplished through divergence time estimates (Hansen et al., 2014). That aspect, as
466 it now stands, is beyond the scope of our study.

467

468

469

## 5| CONCLUSIONS

The multispecies comparative approach employed here revealed general patterns that could not have been discerned from a singular study of any one species. Additionally, the variability in intraspecific genetic structure among species provides a specific, all-encompassing dimension that single-species studies cannot. While meta-analytic frameworks have some potential, they are limited by confounding effects that stem from differences between studies, such as markers, sample sizes, environmental exigencies, and historic context. This necessitates a community-level approach within a study region. Further work aimed at modeling variables can lead to greater insight, ultimately improving our hypotheses regarding genetic diversity for which contemporary data are unavailable.

Importantly, our comparative framework supports the Stream Hierarchy Model as a general model for the genetic structure of lotic fish species and suggests that hydrologic units characterize regional genetic diversity quite well. Out of this result emerged the potential for HUC units to serve as a 'rule of thumb' for riverine biodiversity conservation. None of the species evaluated herein were panmictic. Genetic variation among HUCs was apparent despite limited evidence of discrete population or continuous structure. Across a suite of commonly occurring fishes representing seven families, we identified greater intraspecific gene flow within basins/sub-basins, rather than gene flow among them. Therefore, fish populations within separate HUCs at the 8-digit+ level (e.g., HUC6, HUC4, HUC2) should be considered isolated until proven otherwise (Shelley et al., 2021).

As previously recognized, independent populations warrant independent management (Hopken et al., 2013). When migration is low or non-existent, management of one population is unlikely to impact another. Genetic variation unique to hydrologic units could allow for adaptation to future environmental change, while on the other hand, isolation of populations could underscore elevated extirpation risks (Harrisson et al., 2014). Furthermore, efforts to propagate populations via stocking or translocation should carefully assess the genetic landscape of the species in question, particularly before co-mingling diversity from different sub-basins (Meffe & Vrijenhoek, 1988). Such uninformed mixing of genetic stocks could

promote outbreeding and the erosion of unique genetic diversity within river catchments. However, this must be weighed against the risks of local extirpation (Pavlova et al., 2017).

Given this study's general and comparative nature, we refrain from designating populations within species as potential management units (MUs). However, species showing high levels of genetic structure (Table 2) should be assessed individually for such designation, possibly requiring more fine-scaled, targeted sampling. Additional river/sub-basin-specific management efforts could also be justified, given the presence of unique populations across multiple species (Hopken et al., 2013). Here we specifically refer to: The Little Red, North Fork, Buffalo, Upper Black, Current, and Spring rivers. These may indeed represent evolutionarily significant catchments, and this insight underscores the potential for community-level genetic examination to elevate management to the ecosystem scale.

## ACKNOWLEDGEMENTS

We thank M. Flurry, M. George, T. Goodhart, K. Hollar, and M. Reed, who assisted with DNA extractions. The Arkansas High-Performance Computing Center provided analytical resources. Funding was provided by the University of Arkansas Distinguished Doctoral Fellowship and Harry and Jo Leggett Chancellor's Fellowship (ZDZ), the Bruker Professorship in Life Sciences (MRD), the Twenty-First Century Chair in Global Change Biology (MED), and by an NSF Postdoctoral Research Fellowship in Biology (TKC) [DBI: 2010774]. The findings, conclusions, and opinions expressed in this article represent those of the authors and do not necessarily represent the views of the NSF nor other affiliated or 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

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

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

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

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

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

565 da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L.,
566 Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation
567 biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*,
568 3–13. https://doi.org/10.1016/j.margen.2016.04.012
569 Davis, C. D., Epps, C. W., Flitcroft, R. L., & Banks, M. A. (2018). Refining and defining riverscape
570 genetics: How rivers influence population genetic structure. *Wiley Interdisciplinary Reviews:*
571 *Water*, *5*(2), e1269. https://doi.org/10.1002/wat2.1269
572 DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance
573 of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147-4154.
574 https://doi.org/10.1111/mec.16051
575 Douglas, M. R., Anthonysamy, W. J. B. Mussmann, S. M., Davis, M. A., Louis, W. and Douglas, M. E.
576 (2020). Multi-targeted management of upland game birds at the agroecosystem interface in
577 midwestern North America. *PLoS ONE* 15(4), e0230735.
578 https://doi.org/10.1371/journal.pone.0230735
579 Douglas, M. E., and Matthews, W. J. (1992). Does morphology predict ecology? Hypothesis testing
580 within a freshwater stream fish assemblage. *Oikos*, 65(2), 213-224. https://doi.org/10.2307/3545012
581 Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G.,
582 Legendre, P., & Wagner, H. (2020). R package "adespatial": Multivariate Multiscale Spatial
583 Analysis. Retrieved from: https://cran.r-project.org/package=adespatial
584 Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets.
585 *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966
586 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from
587 metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction
588 data. *Genetics*, *131*(2), 479–491. https://doi.org/10.1093/genetics/131.2.479
589 Fagan, W. F. (2002). Connectivity, fragmentation, and extinction risk in dendritic
590 metapopulations. *Ecology*, *83*(12), 3243-3249. https://doi.org/10.1890/0012-
591 9658(2002)083[3243:CFAERI]2.0.CO;2
592 Falke, J. A., Bailey, L. L., Fausch, K. D., & Bestgen, K. R. (2012). Colonization and extinction in
593 dynamic habitats: an occupancy approach for a Great Plains stream fish assemblage. *Ecology*, *93*(4),
594 858-867. https://doi.org/10.1890/11-1515.1
595 Fetzner Jr, J. W., & DiStefano, R. J. (2008). Population genetics of an imperiled crayfish from the White
596 River drainage of Missouri, USA. *Freshwater Crayfish*, *16*, 131-146.
597 Finn, D. S., Bonada, N., Múrria, C., & Hughes, J. M. (2011). Small but mighty: Headwaters are vital to
598 stream network biodiversity at two levels of organization. *Journal of the North American*
599 *Benthological Society*, *30*(4), 963–980. https://doi.org/10.1899/11-012.1
600 Fluker, B. L., Kuhajda, B. R., & Harris, P. M. (2014). The influence of life-history strategy on genetic
601 differentiation and lineage divergence in darters (Percidae: Etheostomatinae). *Evolution*, *68*(11),
602 3199–3216. https://doi.org/10.1111/evo.12505
603 Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to
604 determine the genetic structure of a continuously distributed population: Clusters or isolation by
605 distance? *Journal of Applied Ecology*, *46*(2), 493–505. https://doi.org/10.1111/j.1365-
606 2664.2008.01606.x
607 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies.
608 *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382
609 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation
610 of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.
611 https://doi.org/10.1534/genetics.113.160572
612 Goslee, S., & Urban, D. (2020). ecodist: Dissimilarity-based functions for ecological analysis. Retrieved
613 from: https://cran.r-project.org/package=ecodist
614 Gosselin, T. (2020). R package radiator: RADseq Data Exploration, Manipulation and Visualization using
615 R. Retrieved from: https://thierrygosselin.github.io/radiator/

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

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

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

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

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

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

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

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

Hubert, M., Reynkens, T., Schmitt, E., & Verdonck, T. (2016). Sparse PCA for High-Dimensional Data With Outliers. *Technometrics*, *58*(4), 424–434. https://doi.org/10.1080/00401706.2015.1093962

Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, *47*(1), 64–79. https://doi.org/10.1198/004017004000000563

Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. *BioScience*, *59*(7), 573–583. https://doi.org/10.1525/bio.2009.59.7.8

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

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

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

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

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

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

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

Lawson, D. J., Davies, N. M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., Hemani, G., Smith, G. D., & Timpson, N. J. (2020). Is population structure in the genetic biobank era irrelevant, a

challenge, or an opportunity? *Human Genetics*, 139, 23-41. https://doi.org/10.1007/s00439-019-02014-8

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

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

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

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

Mantel, N., & Valand, R. S. (1970). A Technique of Nonparametric Multivariate Analysis. *Biometrics*, *26*(3), 547. https://doi.org/10.2307/2529108

Matthews, W.J. (1998). *Patterns in Freshwater Fish Ecology*. New York: Chapman and Hall.

Matthews, W. J., & Robison, H. W. (1988). The Distribution of the Fishes of Arkansas: A Multivariate Analysis. *Copeia*, *1988*(2), 358. https://doi.org/10.2307/1445876

Matthews, W. J., & Robison, H. W. (1998). Influence of drainage connectivity, drainage area and regional species richness on fishes of the interior highlands in Arkansas. *American Midland Naturalist*, *139*(1), 1–19. https://doi.org/10.1674/0003-0031(1998)139[0001:IODCDA]2.0.CO;2

Mayden, R. L. (1985). Biogeography of Ouachita Highland Fishes. *The Southwestern Naturalist*, *30*(2), 195–211. https://doi.org/10.2307/3670734

Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Biology*, *37*(4), 329–355. https://doi.org/10.1093/sysbio/37.4.329

Meffe, G.K., & Vrijenhoek, R. C. (1988). Conservation Genetics in the Management of Desert Fishes. *Conservation Biology*, *2*(2), 157–169. https://doi.org/10.1111/j.1523-1739.1988.tb00167.x

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, *11*(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Naimi, B. (2013). R package usdm: Uncertainty analysis for species distribution models. Retrieved from: https://cran.r-project.org/package=usdm

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654

Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.70.12.3321

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, *98*(3), 625–640. https://doi.org/10.1093/genetics/98.3.625

Oaks, J. R. (2019). Full bayesian comparative phylogeography from genomic data. *Systematic Biology*, 68(3), 371-395. https://doi.org/10.1093/sysbio/syy063

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2020). R package vegan: community ecology package. Retrieved from https://cran.r-project.org/package=vegan

Olden, J. D., Poff, N. L., & Bestgen, K. R. (2008). Trait synergisms and the rarity, extirpation, and extinction risk of desert fishes. *Ecology*, *89*(3), 847-856. https://doi.org/10.1890/06-1864.1

718 Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach.
719 *Bioinformatics*, 26(3), 419-420. https://doi.org/10.1093/bioinformatics/btp696
720 Pavlova, A., Beheregaray, L. B., Coleman, R., Gilligan, D., Harrisson, K. A., Ingram, B. A., Kearns, J.,
721 Lamb, A. M., Lintermans, M., Lyon, J., Nguyen, T. T. T., Sasaki, M., Tonkin, Z., Yen, J. D. L., &
722 Sunnucks, P. (2017). Severe consequences of habitat fragmentation on genetic diversity of an
723 endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications*,
724 *10*(6), 531–550. https://doi.org/10.1111/eva.12484
725 Paz-Vinas, I., Loot, G., Hermoso, V., Veyssière, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018).
726 Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal*
727 *Society B: Biological Sciences*, *285*(1877). https://doi.org/10.1098/rspb.2017.2746
728 Pearson, R. G., Stanton, J. C., Shoemaker, K. T., Aiello-Lammens, M. E., Ersts, P. J., Horning, N.,
729 Fordham, D. A., Raxworthy, C. J., Ryu, H. Y., McNees, J., & Akçakaya, H. R. (2014). Life history
730 and spatial traits predict extinction risk due to climate change. *Nature Climate Change*, *4*(3), 217-
731 221. https://doi.org/10.1038/nclimate2113
732 Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping
733 rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data*
734 *Analysis*, *49*(4), 974-997. https://doi.org/10.1016/j.csda.2004.06.015
735 Perez, M. F., Franco, F. F., Bombonato, J. R., Bonatelli, I. A. S., Khan, G., Romeiro-Brito, M., Fegies, A.
736 C., Ribeiro, P. M., Silva, G. A. R., & Moraes, E. M. (2018). Assessing population structure in the
737 face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*, 24(12),
738 1883-1889. https://doi.org/10.1111/ddi.12816
739 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest
740 RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-
741 model species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135
742 Peterson, E. E., Ver Hoef, J. M., Isaak, D. J., Falke, J. A., Fortin, M. J., Jordan, C. E., McNyset, K.,
743 Monestiez, P., Ruesch, A. S., Sengupta, A., Som, N., Steel, E. A., Theobald, D. M., Torgersen, C.
744 E., & Wenger, S. J. (2013). Modelling dendritic ecological networks in space: An integrated
745 network perspective. *Ecology Letters*, *16*(5), 707–719. https://doi.org/10.1111/ele.12084
746 Pilger, T. J., Gido, K. B., Propst, D. L., Whitney, J. E., & Turner, T. F. (2017). River network
747 architecture, genetic effective size and distributional patterns predict differences in genetic structure
748 across species in a dryland stream fish community. *Molecular Ecology*, *26*(10), 2687–2697.
749 https://doi.org/10.1111/mec.14079
750 Radinger, J., & Wolter, C. (2015). Disentangling the effects of habitat suitability, dispersal, and
751 fragmentation on the distribution of river fishes. *Ecological Applications*, *25*(4), 914–927.
752 https://doi.org/10.1890/14-0422.1
753 Riginos, C., Buckley, Y. M., Blomberg, S. P., & Treml, E. A. (2014). Dispersal Capacity Predicts Both
754 Population Genetic Structure and Species Richness in Reef Fishes. *The American Naturalist*, *184*(1),
755 52–64. https://doi.org/10.1086/676505
756 Roberts, J. H., Angermeier, P. L., & Hallerman, E. M. (2013). Distance, dams and drift: What structures
757 populations of an endangered, benthic stream fish? *Freshwater Biology*, *58*(10), 2050–2064.
758 https://doi.org/10.1111/fwb.12190
759 Robison, H. W., & Buchanan, T. M. (2020). *Fishes of Arkansas* (2nd ed.). Fayetteville, AR: University of
760 Arkansas Press.
761 Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005).
762 Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS*
763 *Genetics*, *1*(6), 0660–0671. https://doi.org/10.1371 /journal.pgen.0010070
764 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation
765 by distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1006/bijl.1996.0127
766 Ruiz-Gonzalez, A., Cushman, S. A., Madeira, M. J., Randi, E., & Gómez-Moliner, B. J. (2015). Isolation
767 by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting
768 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

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

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

Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022). A community genomics approach to natural hybridization. Manuscript submitted for publication.

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

Zbinden, Z. D., Geheber, A. D., Matthews, W. J., & Marsh-Matthews, E. (2022). Fish communities, species of greatest conservation need, and protected areas in southeastern Oklahoma, 2014-2016. *Proceedings of the Oklahoma Academy of Science2*, *101*, 14–32.

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

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## AUTHOR CONTRIBUTIONS

ZDZ conceived the research with input from all authors. Specimen collection was done by ZDZ & TKC. 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 MED administered funding through their University of Arkansas Endowments.

**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 |

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

| 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% |

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

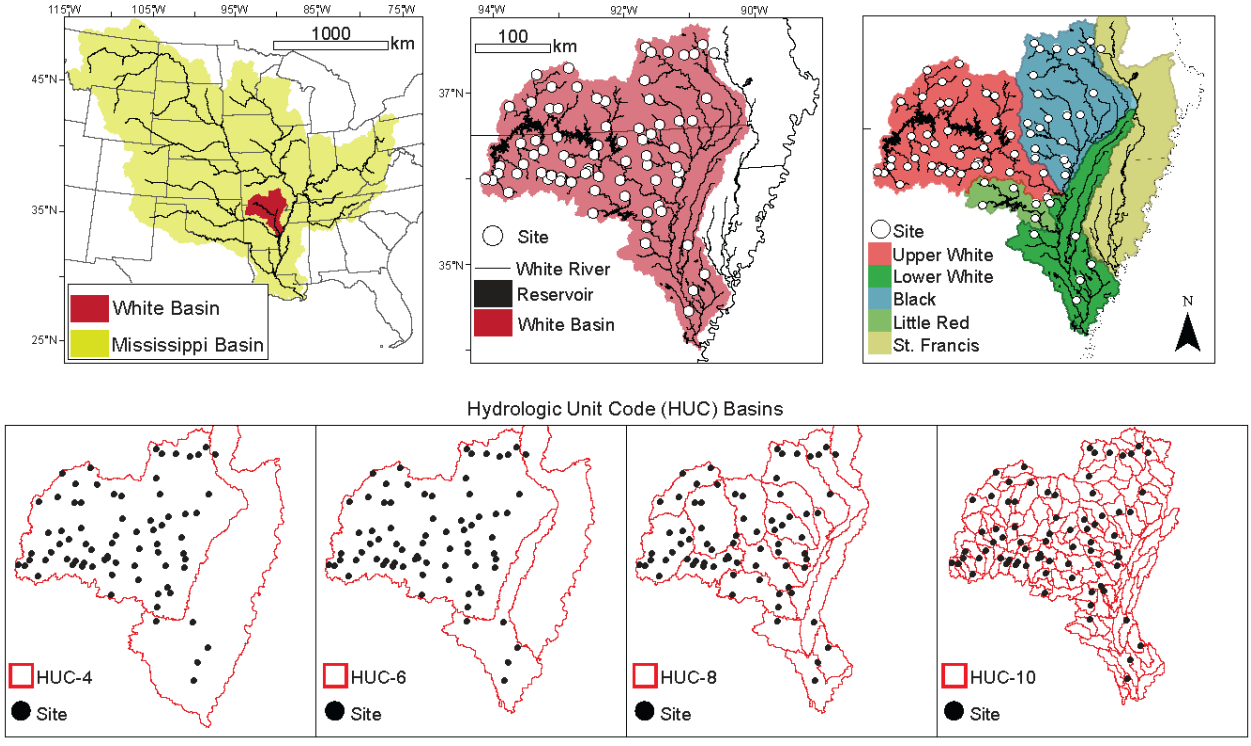
858

| | $H_S$ | $H_T$ | $G_{ST}$ | $G''_{ST}$ |
|---|---|---|---|---|
| $H_T$ | 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); $\Phi_{ST}$ = genetic variation among sites across all groups; $\Phi_{SC}$ = genetic variation among sites within a group.

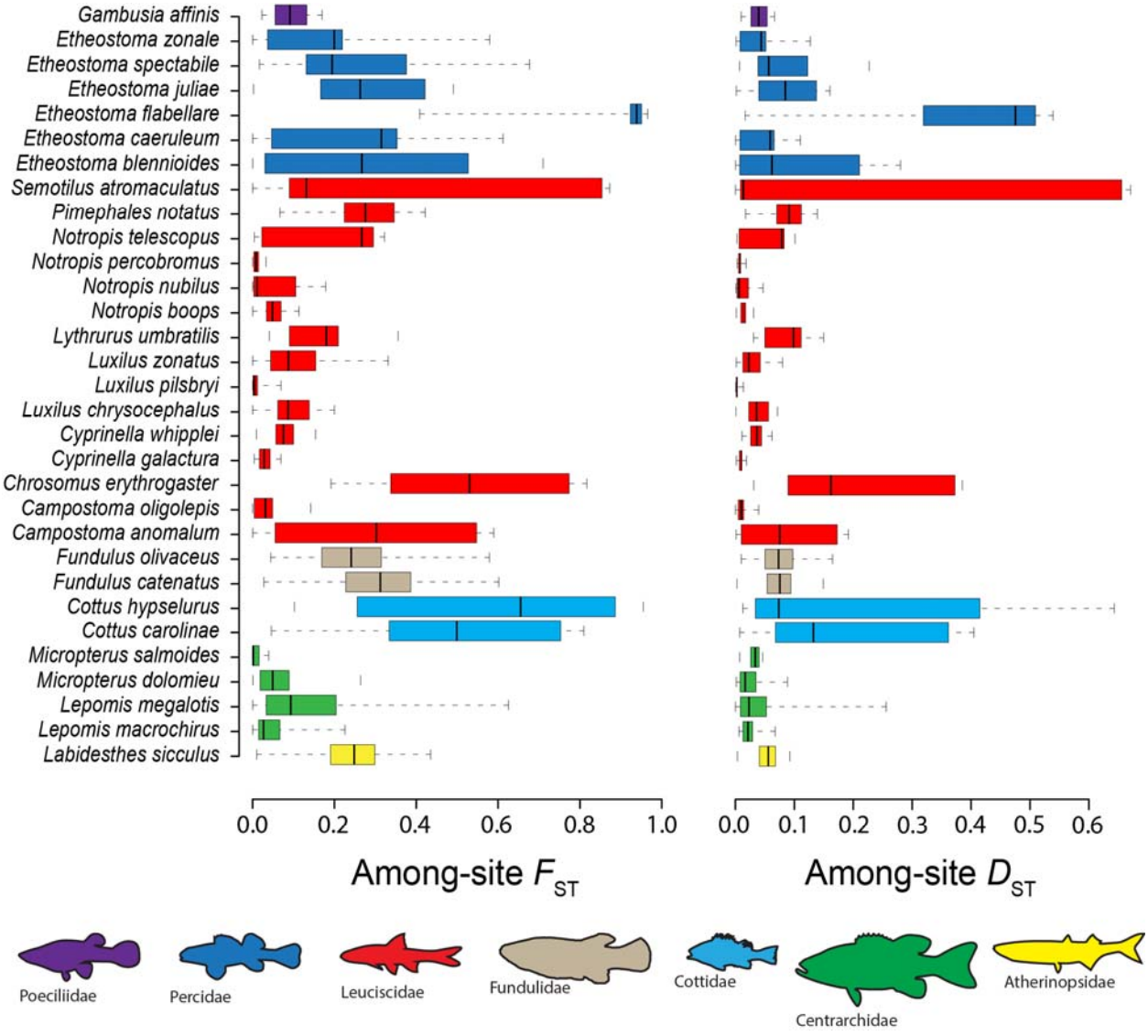
| Family | Species | Hydrologic Units | | | | | | | Population Clusters | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HUC-level | Among HUCs | | Among Sites | | | | Among Pops | | Among Sites | | | |
| | | | %var | sig. | %var | $\Phi_{ST}$ | sig. | $\Phi_{SC}$ | %var | sig. | %var | $\Phi_{ST}$ | sig. | $\Phi_{SC}$ |
| 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 |

865 **FIGURE 1** Fish were sampled at *N*=75 locations across the White River Basin (Ozark Mountains,
866 U.S.A.). The study basin is contained within the larger Mississippi River Basin, and is a direct tributary to
867 the mainstem Mississippi. The study region is subdivided into five subbasins: Upper White, Lower White,
868 Black, Little Red, and the St. Francis. Beyond these basins, USGS Hydrologic Unit Codes (HUCs) were
869 also used to characterize the stream hierarchy position of sampling locations (4-, 6-, 8-, and 10-digit
870 HUCs).
871




872

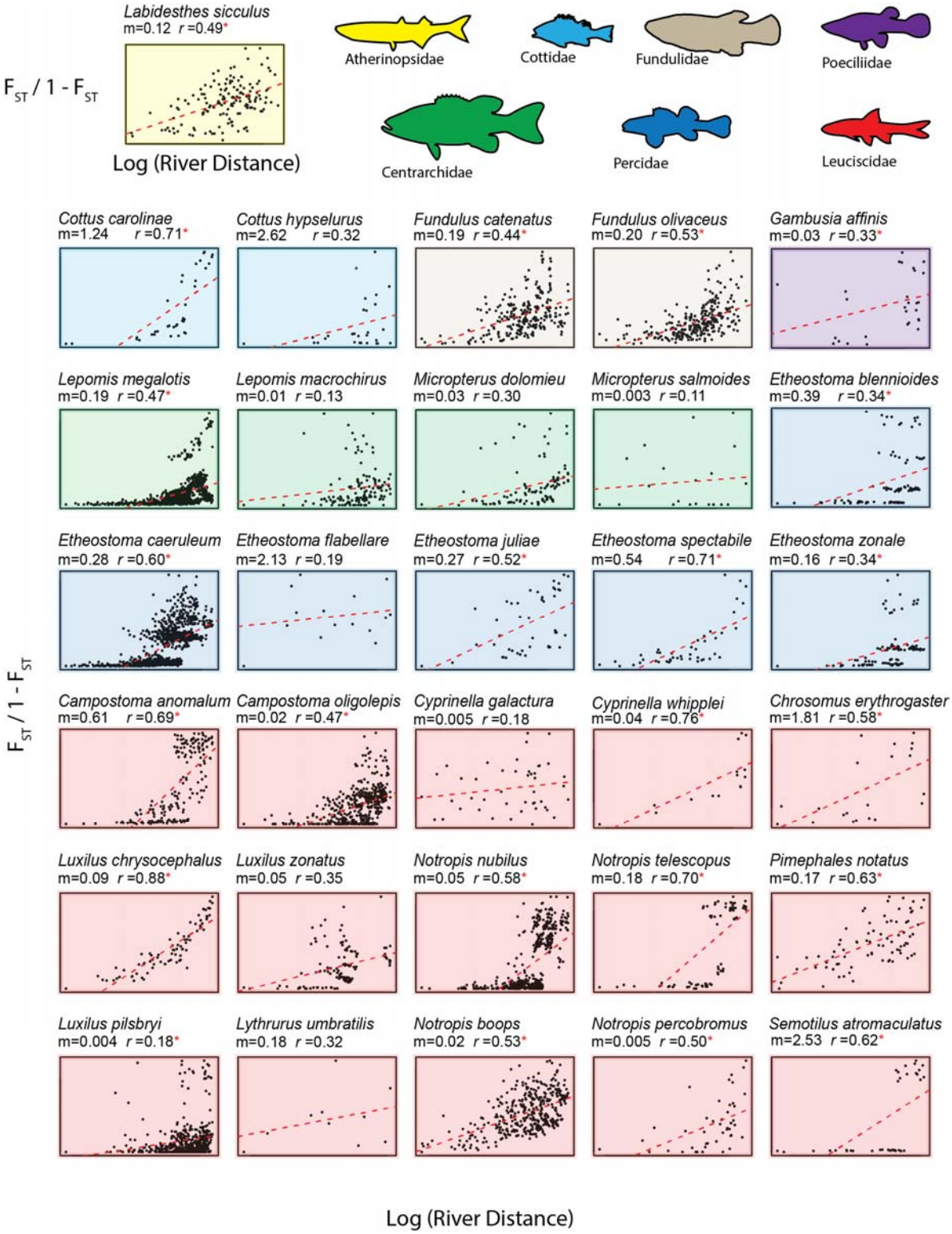
873 **FIGURE 2** Genetic structure of *N*=31 fish species collected across the White River Basin (Ozark
874 Mountains, U.S.A.) as summarized by among-site $F_{ST}$ (Weir and Cockerham's θ) and Jost's *D*. Boxplots
875 show the distributions of both pairwise estimates among sampling sites for each species. Inner quantiles
876 are colored to indicate species in the same family (*N*=7).
877



878

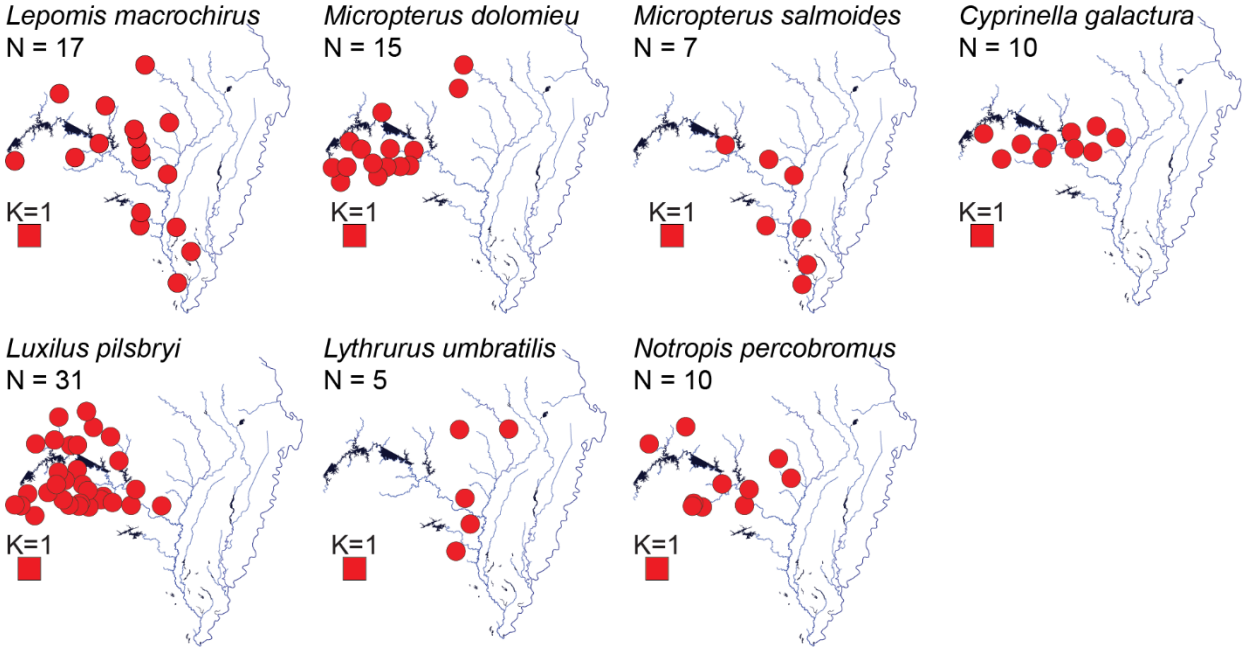
879 **FIGURE 3** Isolation by distance plots for $N$=31 fish species collected across the White River Basin (
880 Ozark Mountains, U.S.A.). Each depicts the relationship between among-site $F_{ST}$ (linearized) and log river
881 distance among sites. The following are represented below each species name: m=slope of the linear
882 regression model (dashed red line) and $r$= the Mantel coefficient indicating the strength of the correlation
883 between genetic structure and distance. Significant $r$-values denoted with a red asterisk ($\alpha \leq 0.05$).

884




885

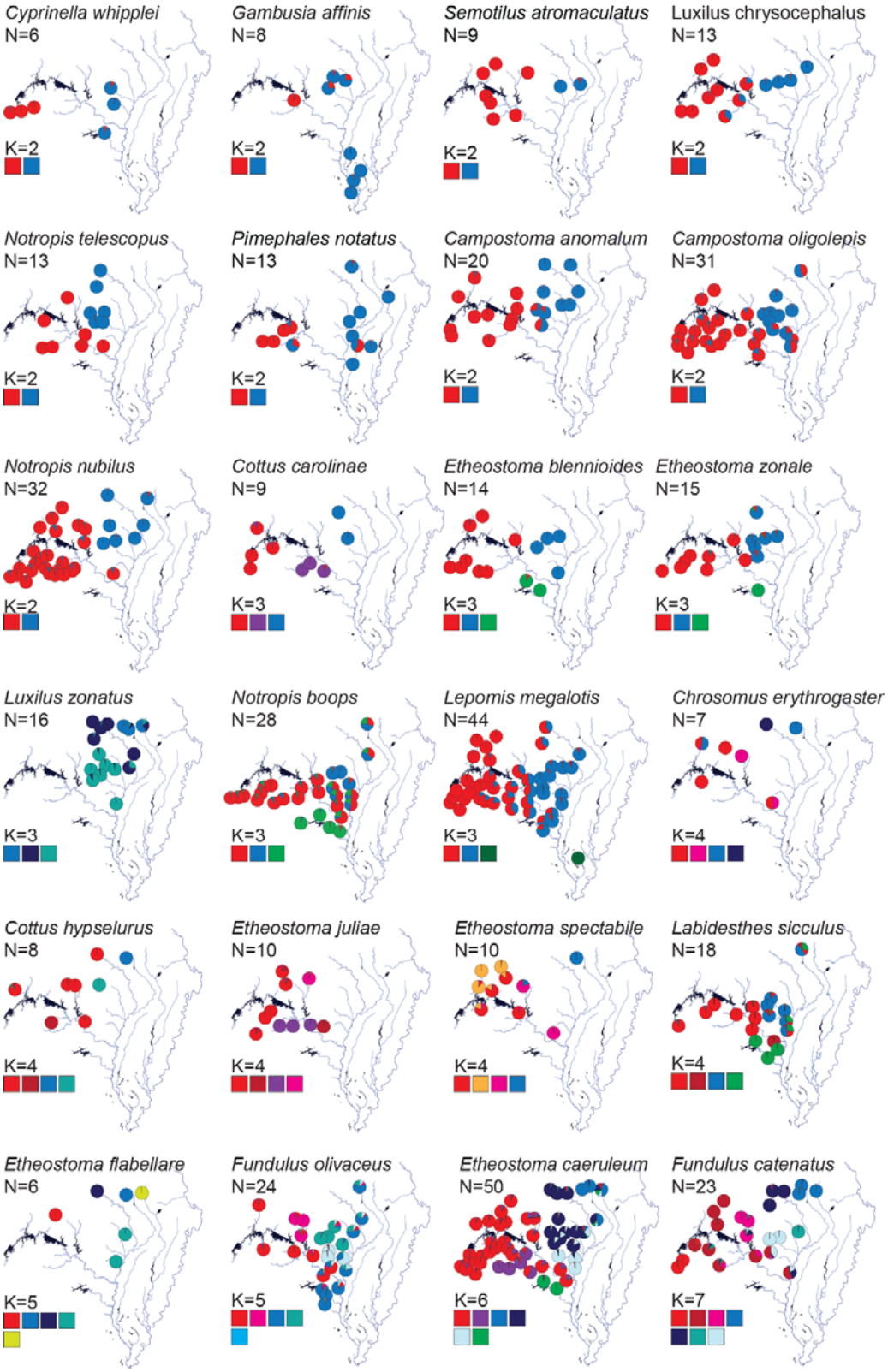
886 **FIGURE 4** Sampling distribution maps of seven species which showed no evidence of discrete genetic
887 population structure within the White River Basin ( Ozark Mountains, U.S.A.). A total of *N*=31 species
888 were sampled across 75 sites. The number of collection sites (red circles) for each species is denoted by
889 N; K=the number of discrete genetic populations discerned from sparse non-negative matrix factorization.
890




891

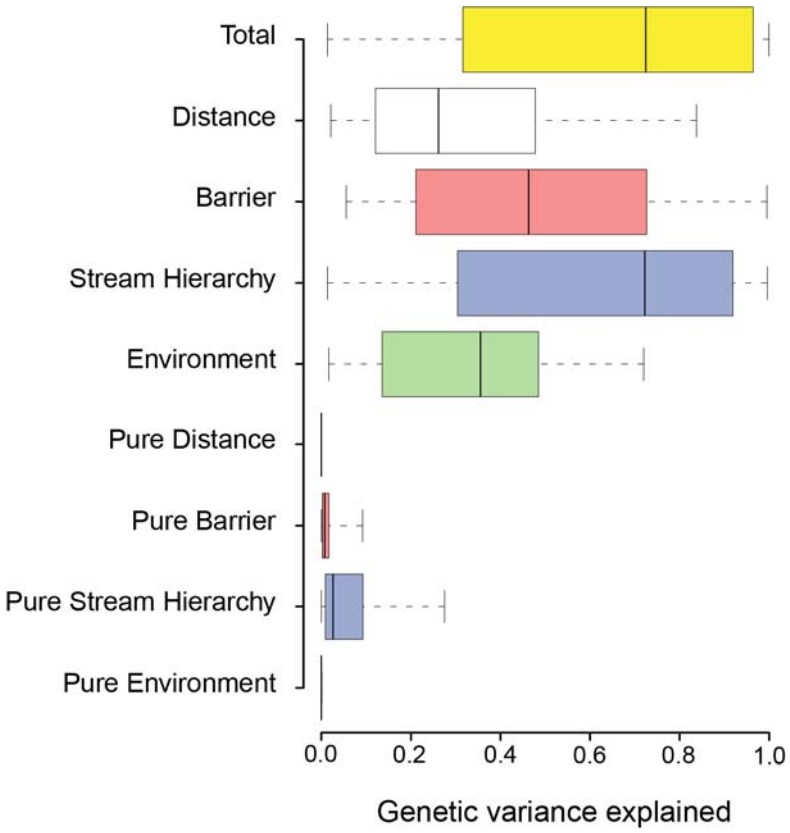
892 **FIGURE 5** Sampling distribution maps of 24 species which showed evidence of genetic population structure
893 within the White River Basin (Ozark Mountains, U.S.A.). *N*=31 species were sampled across 75 sites. K= the
894 number of discrete genetic populations discerned from sparse non-negative matrix factorization. Sampling
895 sites are denoted as pie charts representing the average population coefficients for each site. N= number of
896 sites where each species was collected.

897




898

899 **FIGURE 6** Neutral genetic variation was partitioned between four explanatory models for *N*=31 fish
900 species sampled across the White River Basin (Ozark Mountains, U.S.A.). Partitioning was conducted
901 separately for each species. The four models represent: (i) isolation by *distance*, the river network
902 distance among individuals represented by spatial eigenvectors; (ii) isolation by *barrier*, represented by
903 population structure coefficients among individuals; (iii) isolation by *stream hierarchy*, based on the
904 hydrologic units (at four different hierarchical levels) in which an individual was collected; and (iv)
905 isolation by *environment*, characterized by the environmental heterogeneity across sampling sites where
906 individuals were collected. Total = the genetic variation explained by all four models combined. The
907 "Pure" models represent the variation explained by each model after partialling out the variation
908 explained by the other three models.
909




910

**FIGURE 7** Venn diagrams depict neutral genetic variation resulting from four models as applied to *N*=31 fish species sampled from the White River Basin (Ozark Mountains, U.S.A.). Models were based on: (i) isolation by distance, isolation by barrier, isolation by stream hierarchy, and isolation by environment. Values in the Venn diagrams are percent of genetic variance explained (i.e., rounded adjusted $R^2$ values). Total variance explained is shown below each diagram. The bottom two rows show species that showed no discrete population structure (i.e., no isolation by barrier) and thus only three of the models were tested.