**Title:** Gene flow influences the genomic architecture of local adaptation in six riverine fish species

**Running Title:** Gene flow influences adaptation architecture of local adaptation

**Authors:** Yue Shi[1,2,*], Kristen L. Bouska[3], Garrett J. McKinney[4], William Dokai[1,2], Andrew Bartels[5], Megan V. McPhee[1], Wesley A. Larson[6,7]

**Contact information:**

[1] College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Road, Juneau, AK 99801, USA.

[2] Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, USA.

[3] U.S. Geological Survey, Upper Midwest Environmental Sciences Center, 2630 Fanta Reed Road, La Crosse, WI 54603, USA.

[4] NRC Research Associateship Program, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, WA 98112, USA.

[5] Long Term Resource Monitoring Program, Wisconsin Department of Natural Resources, 2630 Fanta Reed Road, La Crosse, WI 54603, USA.

18 [6] National Oceanographic and Atmospheric Administration, National Marine Fisheries Service,
19 Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau,
20 AK 99801, USA

21 [7] U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural
22 Resources, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481,
23 USA

24 *Corresponding author. Email address: yshi8@alaska.edu

**Abstract**

Understanding how gene flow influences adaptive divergence is important for predicting adaptive responses. Theoretical studies suggest that when gene flow is high, clustering of adaptive genes in fewer genomic regions would protect adaptive alleles from recombination and thus be selected for, but few studies have tested it with empirical data. Here, we used RADseq to generate genomic data for six fish species with contrasting life histories from six reaches of the Upper Mississippi River System, USA. We used four differentiation-based outlier tests and three GEA analyses to define neutral SNPs and outlier SNPs that were putatively under selection. We then examined the distribution of outlier SNPs along the genomes and investigated whether these SNPs were found in genomic islands of differentiation and inversions. We found that gene flow varied among species, and outlier SNPs were clustered more tightly in species with higher gene flow. The two species with the highest overall $F_{ST}$ (0.0303 - 0.0720) and therefore lowest gene flow showed little evidence of clusters of outlier SNPs, with outlier SNPs in these species spread uniformly across the genome. In contrast, nearly all outlier SNPs in the species with the lowest $F_{ST}$ (0.0003) were found in a single large putative inversion. Two other species with intermediate gene flow ($F_{ST}$ ~ 0.0025 - 0.0050) also showed clustered genomic architectures, with most islands of differentiation clustered on a few chromosomes. Our results provide important empirical evidence to support the hypothesis that increasingly clustered architectures of local adaptation are associated with high gene flow.

**Keywords:** Freshwater Fishes, Local Adaptation, Gene Flow, Genomic Islands of Differentiation, Chromosomal Inversions, Mississippi River

## Introduction

Understanding the genomic basis of adaptation is a central goal of evolutionary biology. Research on this topic largely focuses on identifying genetic markers involved in adaptation and assessing the distribution of these markers across the genome (Narum & Hess 2011; Yeaman 2013; Lotterhos & Whitlock 2014; Hoban *et al.* 2016; Forester *et al.* 2018). Substantial efforts have focused on this area of research for decades (Smith & Haigh 1974; Rieseberg 2001; Noor *et al.* 2001). However, results have been highly variable across taxa and systems and are influenced by variable demographic histories (Ravinet *et al.* 2017; Gagnaire 2020), making it difficult to gain a mechanistic understanding of the evolutionary processes that influence the genomic landscape of adaptation. For example, many studies have found that alleles contributing to local adaptation tend to be clustered together in genomic islands of differentiation, while other studies have found little or no evidence of adaptive alleles clustering within genomic islands (Nosil *et al.* 2009; Strasburg *et al.* 2012; Roda *et al.* 2017; Johannesson *et al.* 2020; Thompson *et al.* 2020). This mixed evidence across different study systems with differing demographic histories raises an important evolutionary question: when are loci affecting adaptive divergence expected to be tightly clustered?

Interpreting results from genome scans in the context of gene flow may aid in the understanding of genomic landscapes of adaptation (Marques *et al.* 2016). Gene flow can be beneficial for maintaining genetic diversity by introducing novel genetic variation but it can also impede local adaptation by introducing maladaptive foreign alleles into locally adapted populations (Bolnick & Nosil 2007). One potential evolutionary 'solution' that may minimize maladaptive effects of gene flow is for selection to favor clustered architectures of adaptation, where adaptive alleles

are tightly linked and locally favorable combinations of alleles are protected from disruption via low recombination (Yeaman 2013; Roesti 2018).

Several mechanisms have been proposed to explain the observations of clustered genomic architectures of adaptive alleles when gene flow is high, including divergence hitchhiking and the utilization of genomic rearrangements to protect adaptive loci from recombination. Divergence hitchhiking occurs when gene exchange between diverging populations is reduced around a gene under strong divergent selection (Via 2012). This process can produce islands of differentiation spanning multiple megabases, as free recombination among populations is reduced due to assortative mating (Via 2012). Genomic rearrangements, such as chromosomal inversions, can also facilitate adaptation in the face of high gene flow and lead to genomic islands of differentiation (Hoffmann & Rieseberg 2008; Yeaman 2013; Tigano & Friesen 2016; Roesti 2018; Wellenreuther & Bernatchez 2018; Aguirre Liguori *et al.* 2019; Huang *et al.* 2020; Cayuela *et al.* 2020). Recombination between inverted and noninverted arrangements is rare as recombinant gametes are generally inviable (Huang & Rieseberg 2020). Therefore, if an inversion isolates multiple adaptive alleles, this architecture will likely be favored, because co-adapted genotypes will be protected from recombination and allowed to evolve independently even in high gene flow environments (Rogers *et al.* 2013; Yeaman 2013).

Although the theories described above posit that the rate of evolution towards clustered architectures of local adaptation should increase with gene flow, this hypothesis has largely been tested with simulations rather than empirical data. For example, Yeaman & Whitlock (2011) used simulations to demonstrate increasing migration rate, or *m,* leads to increasingly concentrated genomic architectures of adaptation. However, when *m* is too high, adaptive

91 divergence is unlikely because frequent migration prevents even a perfectly adapted mutation
92 from overcoming the homogenizing effects of gene flow. A subsequent simulation study
93 (Yeaman 2013) highlighted that genomic rearrangement may often be an important component
94 of local adaptation and when genomic rearrangements are present, tight clustering of adaptive
95 loci can readily evolve even with high *m*.

96 In this study, we investigate how gene flow influences the genomic architecture of adaptation
97 using genomic data from six riverine fish species that encompass a diverse suite of life histories
98 and dispersal potentials (Figure 1B). These fish were sampled from the same sites in the Upper
99 Mississippi River System (UMRS) in the midwestern United States. The UMRS is an
100 interconnected large river system that hosts a diversity of aquatic habitats in terms of
101 temperature, turbidity, productivity, and flow (Figure 1A & C). Our study system provides a
102 unique opportunity to compare the genomic architecture of local adaptation in a natural
103 environment for species with contrasting life histories and to assess the influence of gene flow on
104 genomic architecture. Specifically, we test the hypothesis that the genomic islands of
105 differentiation are less frequent but larger for species with relatively high gene flow, whereas
106 genomic islands are more numerous and dispersed throughout the genome for species with low
107 degrees of gene flow. Our multi-species approach investigating six species inhabiting the same
108 environments is unique, as most previous studies have focused on closely related species pairs or
109 ecotypes (Nadeau *et al.* 2012; Renaut *et al.* 2012) rather than divergent species inhabiting the
110 same environments.

111 **Materials and Methods**

112 ***Study Design and Genotyping***

We collected genetic samples from six fish species found in the UMRS which are native to and commonly found in the region and have not been extensively stocked: Bullhead Minnow (*Pimephales vigilax*), Bluegill (*Lepomis macrochirus*), Freshwater Drum (*Aplodinotus grunniens*), Channel Catfish (*Ictalurus punctatus*), Gizzard Shad (*Dorosoma cepedianum*), and Emerald Shiner (*Notropis atherinoides*). The UMRS is congressionally defined as the commercially navigable portions of the Mississippi River main stem north of Cairo, Illinois and commercially navigable tributaries, including the entire Illinois River (Water Resources Development Act of 1986, 33 U.S.C. §§ 652). Fin-clip samples were collected from adult fish in summer 2018 and 2019 across six river reaches (Figure 1A). These reaches are stretches of the embanked floodplain along the river with commercial navigation channels. Five of the study reaches are navigation pools, named for their downstream lock and dam, and the other study reach, Open River Reach, is an unobstructed, channelized reach. We targeted a sample size of at least 48 samples per species per reach. Samples were genotyped at thousands of SNPs using restriction site-associated DNA (RAD) sequencing (see Supplementary Methods). Data on life history traits for each species, including exploitation status, feeding guild, habitat guild, reproductive guild, spawning migration, and total length were summarized in Table S1. We also obtained data for 20 environmental variables across the six river reaches (Table S2). All analyses were performed in parallel within each species and the results were compared among species. No analyses were performed between species.

***Identification of Outlier SNPs and Neutral SNPs***

There are two types of outlier tests, differentiation-based methods and genotype-environment association (GEA) methods. Differentiation-based methods identify loci with high $F_{ST}$ values,

which are expected for loci involved in hard selective sweeps with large changes in allele frequencies (Brauer *et al.* 2016; Forester *et al.* 2018). By comparison, GEA analyses identify genetic variants associated with particular environmental factors and can identify loci under polygenic and "soft" selective sweeps with relatively small changes in allele frequencies (Eckert *et al.* 2010; Brauer *et al.* 2016; Forester *et al.* 2018). We ran four different differentiation-based outlier tests on each species (Bayescan, Arlequin, OutFLANK, and *pcadapt*; see Supplementary Methods for details) and defined "$F_{ST}$ outliers" as SNPs identified by at least two differentiation-based methods. In addition, we conducted three GEA analyses: redundancy analysis, latent factor mixed models, and a Bayesian method (Bayenv2). Details of these methods can be found in the Supplementary Methods. Prior to all three GEA analyses, we conducted principal component analysis (PCA) on 20 standardized environmental variables. Based on Kaiser-Guttmann criterion and the broken stick model, we retained the first two significant PCs as environmental composite variables in order to remove collinearity among variables (Figure S1A). Variables related to temperature, turbidity, pH, and dissolved oxygen had high loadings on environmental PC1 (Figure S1B), whereas productivity and flow-related variables contributed significantly to environmental PC2 (Figure S1C). We defined "GEA outliers" as SNPs identified by at least two GEA methods. To determine which environmental PC each GEA outlier was most strongly correlated with, we compared correlation coefficients between each environmental PC and genotype for each outlier using R function *cor* and assessed which environmental PC had the highest correlation coefficient.

We combined results from differentiation-based outlier tests and GEA analyses and defined (1) "outlier SNPs" as the union of the two sets, "$F_{ST}$ outliers" and "GEA outliers"; (2) and "neutral SNPs" as those that were not identified as outliers by any of the aforementioned seven methods.

158 Because elevated levels of linkage disequilibrium (LD) may have confounding effects when
159 assessing population structure as genomic regions with high LD will be overrepresented
160 (Abdellaoui *et al.* 2019), we conducted LD thinning on the neutral SNPs datasets using the
161 function *snp_ autoSVD* (max.iter=10, roll.size=0) in the R packages *bigsnpr* (Prive *et al.* 2018),
162 which uses sliding windows to remove SNPs correlated with the SNP with the highest MAF in
163 that window (R2 > 0.2) and removes regions with putative long-range LD. The thinned neutral
164 SNPs were used as the final sets of "neutral SNPs".

165 ***Neutral Genetic Differentiation***

166 We used three methods to estimate neutral population structure for each species using their
167 thinned neutral datasets. First, we calculated global Fstp ($F_{ST}$ corrected for sampling bias) using
168 the function *basic.stats* in *hierfstat* v.0.04-22 (Goudet 2005). Next, we calculated $F_{ST}$ between all
169 pairs of river reaches using *genet.dist* function (method="WC84") in *hierfstat*. Significance was
170 assessed by calculating 95% confidence interval of pairwise $F_{ST}$ values using *boot.ppfst* function
171 (nboot=1000) in *hierfstat*. A pairwise $F_{ST}$ value was considered significant if its confidence
172 interval did not include zero. Lastly, we conducted PCA using the R package *adegenet* v2.1.2
173 (Jombart 2008) to investigate neutral genetic differentiation among individuals.

174 To test for isolation by distance (IBD){Wright:1943wy}, we conducted a linear regression of
175 neutral genetic distance to the river distance separating the study reaches. We calculated Nei's
176 genetic distance using *dist.genpop* function (method=1) in *adegenet*. River miles (Table S3) were
177 converted to river kilometers as river distance. The statistical significance of IBD was evaluated
178 using Mantel test implemented in the *mantel.randtest* function (999 permutations) in *ade4*
179 {Dray:2007vs}.

***Genome Scans for Genomic Islands of Differentiation***

We aligned SNPs to reference genomes and conducted genome scans to investigate the genomic landscape of adaptive divergence for each species. Channel Catfish is the only species with a reference genome available in our study. For the other five species, we used the reference genomes from closely related species (Table S4). The reference genomes we used were all chromosome-level assemblies with full genome representation and high contiguity. The scaffold N50 length ranged from 7.4 Mb to 37.4 Mb and the contig N50 length ranged from 77.2 Kb to 4.3 Mb (Table S4). Sequences of filtered RAD loci were mapped to reference genomes with BWA-MEM v 0.7.17 using default settings (Li 2013). We retained sequences with mapping quality > 20 and removed sequences with "SA:Z" (chimeric alignment) and "XA:Z" tags (alternative hits) using *SAMtools* v1.10 (Li *et al.* 2009).

To identify genomic islands of differentiation, we first calculated per-SNP Fstp using the *basic.stats* function in *hierfstat* for all aligned SNPs across genomes. We then used a Hidden Markov Model (HMM) approach implemented in the R package *HiddenMarkov* v.1.8-11 (Hofer *et al.* 2012) to assign each SNP to one of three underlying states, "genomic background", "regions of high differentiation", and "regions of low differentiation" based on their Fstp values, following the methods detailed in Marques *et al.* (2016). The state status was further confirmed by comparing the mean Fstp values among the three states. Regions of high differentiation had the highest mean Fstp values and were the focus of the study. These regions can consist of one or many consecutive SNPs depending on the landscape of differentiation.

The HMM approach identified a large number of highly differentiated regions or "HMM islands", but many did not show especially high levels of differentiation and may be false

positives. Therefore, we only retained the HMM islands that contained at least one $F_{ST}$ outlier (defined previously). We excluded outliers identified only by *pcaadpt* because we discovered this method identified a much higher number of outliers compared to other methods, which could potentially increase false positive rate for island detection. We removed HMM islands located on unplaced scaffolds. We also removed HMM islands in situations where a chromosome only had one island and this island had only one SNP. Since the HMM islands were identified based on Fstp values and $F_{ST}$ outliers, we further examined how many total outlier SNPs (union of $F_{ST}$ outliers and GEA outliers) were found within these HMM islands.

### *Identification and Analyses of Putative Inversions*

To identify putative inversions in each species, we conducted a sliding window analysis of population structure across genomes using the R package *lostruct* (Li & Ralph 2019) following the methods described in Huang *et al.* (2020). We replaced missing genotypes with the most frequent genotype and divided each genome into nonoverlapping windows of either 20 or 50 SNPs depending on the total number of aligned SNPs for each species. We then used a multidimensional scaling (MDS) analysis (40 dimensions) to measure the differences in population structure patterns among windows, and we defined outlier windows as those with absolute values of loadings greater than 4 standard deviations above the mean averaged across all windows in the genome (Huang *et al.* 2020). Outlier windows (single or consecutive) were candidate regions for putative inversions. We also conducted three additional analyses on putative inversion regions to provide additional evidence of inversions as suggested by Huang *et al.* (2020). First, because inversions only suppress recombination in heterozygotes, three distinct genotypic clusters (0, 1, 2) should be detected along PC1 using PCA, with the outside clusters

(cluster 0 and 2) representing two homozygous groups for alternative orientations and the middle cluster (cluster 1) representing the heterozygous group between inversion haplotypes (Lotterhos 2019; McKinney *et al.* 2020). The discreteness of the clustering was calculated as the proportion of the between-cluster sum of squares over the total using the R function *kmeans* in *adegenet*. Second, we compared heterozygosity (the proportion of heterozygotes) among three clusters identified by PCA using Wilcoxon tests ($\alpha = 0.05$) to further confirm the middle group had significantly higher heterozygosity. Finally, we calculated LD ($r^2$) using PLINK v1.9 (Purcell *et al.* 2007; Chang *et al.* 2015) for SNPs with MAF > 0.01 on chromosomes with outlier windows and compared $r^2$ with all samples to $r^2$ calculated only from samples that were homozygous for the most common orientation. This comparison can distinguish inversions from other regions of reduced recombination (Bradley *et al.* 2011; Roesti *et al.* 2012; 2013), because inversions are expected to only suppress recombination in heterokaryotypes and recombination in homokaryotypes should be unaffected, while other mechanisms of recombination suppression are expected to affect all groups of individuals.

We considered a region as a putative inversion only if all of the following criteria were met: (1) a distinct three-cluster PCA pattern with discreteness > 0.9; (2) significantly elevated heterozygosity in the middle PCA cluster compared to the other two clusters; and (3) elevated LD calculated with all samples, but not with homozygous samples. We visualized the genotypes of individuals inside the putative inversions using genotype heatmap, where individual genotypes were color coded with "homo1" and "homo2" representing alternate homozygous genotypes and "het" representing a heterozygous genotype. We assumed that the more derived inversion arrangement would have lower heterozygosity given its relatively recent origin compared to the ancestral state (Laayouni 2003; Twyford & Friedman 2015; Knief *et al.* 2016). Notably, when

examining our data, we found five additional regions with discreteness very close to 0.9 (0.893 - 0.898) that displayed distinct three-cluster PCR patterns, and we included these regions as candidates for putative inversions as well.

We calculated the haplotype frequency of each putative inversion for each reach using the formula $F = \frac{2C_{0\ or\ 2}+C_1}{2N}$ (Le Moan *et al.* 2021), where $C_0$ or $C_2$ is the number of individuals assigned to one of the homozygous clusters (cluster 0 or 2), $C_1$ is the number of individuals assigned to the heterozygous cluster (1) in the PCA, and N is the number of samples in the reach. Additionally, we conducted PCA analyses using all SNPs that were successfully aligned to genomes, SNPs within the identified inversions, and the remaining aligned SNPs after the SNPs in putative inversions were removed to compare patterns of genetic structure inferred from datasets with and without putative inversions.

***Genomic Properties of Clusters of Outlier SNPs***

Outlier SNPs can be found across many areas of the genome or can be clustered in only a few genomic islands with SNPs found much closer together than expected by chance. To determine whether outlier SNPs showed significant clustering, we investigated the distribution of outlier SNPs (union of $F_{ST}$ outliers and GEA outliers) across the genomes using the nearest neighbor distance (NND) metric (Samuk *et al.* 2017) and compared NND between outlier SNPs relative to NND between all SNPs using permutation tests. Specifically, for each species, we first partitioned the dataset by chromosomes and filtered out chromosomes without enough information (number of SNPs <30 or number of outlier SNPs < 3). We only focused on chromosomes with at least 20% of aligned outlier SNPs for each species, as a fixed value cutoff would provide a non-uniform threshold due to variations in the number of aligned outlier SNPs

269 across species. Secondly, for each remaining chromosome, we drew 10,000 samples of random
270 SNPs equal to the number of outlier SNPs on that chromosome and calculated the mean distance
271 between each SNP and its nearest neighbor in the random samples to generate the null
272 distribution of NND for that chromosome. Lastly, we examined whether outlier SNPs were
273 significantly clustered within the chromosomes by comparing the empirical average NND value
274 between outlier SNPs with the null distribution. We also calculated the difference between the
275 average NND between outlier SNPs and the average NND in the null distribution, in units of
276 standard deviations for each chromosome.

277 For each chromosome with significantly over-clustered outlier SNPs, we compare the following
278 genetic metrics between outlier SNPs and non-outlier SNPs (chromosomal background): Fstp,
279 heterozygosity ($H_O$), absolute differentiation ($D_{xy}$), and linkage disequilibrium (LD). Fstp and $H_O$
280 were calculated using the *basic.stats* function in *hierfstat* as described previously. Pairwise per-
281 site $D_{xy}$ was calculated as $p_1(1-p_2) + p_2(1-p_1)$, where $p_1$ is the frequency of a given allele
282 in the first population and $p_2$ is the frequency of that allele in the second population (Irwin *et al.*
283 2016). Allele frequency was estimated using *makefreq* function (missing = "mean") in *adegenet*.
284 Overall $D_{xy}$ was calculated as the mean of all pairwise $D_{xy}$ values. LD ($r^2$) was calculated using
285 PLINK v1.9 for SNPs with MAF > 0.01. We included $D_{xy}$, an absolute measure of genetic
286 differentiation, because defining islands of differentiation based solely on relative measures of
287 differentiation, such as Fstp, may identify regions resulting from variation of recombination rate
288 along the genome and background selection rather than divergent selection (Cruickshank &
289 Hahn 2014). We visualized the differences in the values of the above genetic metrics between
290 outlier SNPs and chromosomal background using boxplots and tested for significance using
291 permutation tests (10,000 permutations). The permutation test procedure was the same as the

permutation tests on NND. Ideally, $D_{xy}$ calculation should be conducted using a sliding window-based approach with all sites included (variant and invariant sites) to get unbiased estimation, such as *pixy* {Korunes:2021hd}. However, per-SNP based approach is most appropriate for our application due to the nature of RAD data. To ensure our $D_{xy}$ inference using per-SNP based approach is equivalent as the ideal window-based approach and because the window-based $D_{xy}$ calculation is highly dependent on SNP density, we examined variation of the number of SNPs across chromosomes using a non-overlapping sliding window analysis with a window size of 100K bp. We compared the number of SNPs in windows overlapping with HMM islands and windows not overlapping with HMM islands (chromosomal background) using wilcox test.

Lastly, we conducted Gene Ontology (GO) enrichment tests for functional enrichment of genes in the HMM islands on chromosomes with significantly over-clustered outlier SNPs. See Supplementary Materials for detailed methods about GO enrichment tests.

## Results

### *Summary of Sequencing, Outlier SNPs, and Neutral SNPs*

We RAD sequenced a total of 1,712 individuals, ranging from 275 - 288 individuals per species. RAD sequencing yielded an average of 5,780,907 retained reads per individual (range = 16,799 - 47,250,859). After filtering, 1,417 individuals (179 - 256 individuals per species) were retained and genotyped at 10,834 - 28,313 polymorphic SNPs depending on the species (Table S4). Out of these polymorphic SNPs, 0.05 % to 0.46% were identified as outlier SNPs (union of both $F_{ST}$ outliers and GEA outliers), and 95.8 % - 99.1% were identified as neutral SNPs (after thinning) in each species (Table S5). For most species, the majority of GEA outliers were found to be

strongly associated with environmental PC1 (temperature, turbidity, pH, and dissolved oxygen related). In contrast, GEA outliers in Freshwater Drum were strongly associated with environmental PC2 (productivity and flow related) (Table S5).

***Neutral Genetic Differentiation***

Patterns of population structure estimated from the thinned neutral datasets spanned a large gradient of genetic differentiation across species (Figure 2, Table S6). Bullhead Minnow had the highest global Fstp value of 0.0720 with pairwise $F_{ST}$ values ranging from 0.0041 to 0.1543, followed by Bluegill (global Fstp = 0.0303; pairwise $F_{ST}$ = 0.0014 - 0.0739), Freshwater Drum (global Fstp = 0.0050, pairwise $F_{ST}$ = -0.0003 - 0.0169), Channel Catfish (global Fstp = 0.0025, pairwise $F_{ST}$ = 0.0003 - 0.0048), and Gizzard Shad (global Fstp = 0.0024, pairwise $F_{ST}$ =0.0003 - 0.0051). Emerald Shiner had the lowest global Fstp value among all six species, 0.0003, with pairwise $F_{ST}$ values ranging from -0.0004 to 0.0014.

Results of the neutral PCAs (Figure 3) corroborated the patterns described above. In Bullhead Minnow, we detected five genetic clusters, with individuals from each river reach forming a single cluster except for Pool 8 and Pool 13, which were grouped together. In Bluegill, individuals from the three northern river reaches (Pool 4, Pool 8, and Pool 13) were genetically similar, Pool 26 and La Grange formed a second cluster, while the most southerly reach, Open River, formed its own cluster. In Freshwater Drum, individuals from La Grange clearly grouped separately from other populations along with some individuals from Pool 26 and Open River. In Channel Catfish, individuals from the Open River and La Grange were slightly separated from other reaches. In Gizzard Shad, individuals from the three northern river reaches were slightly

separated from those in the southern river reaches. Lastly, there was no apparent population structure in Emerald Shiner.

Overall, the IBD patterns corroborated the above results on neutral population differentiation. Based on Mantel test results, Nei's genetic distance was significantly correlated with the river distance between study reaches in Bullhead Minnow, Bluegill, Channel Catfish, and Gizzard Shad ($p < 0.05$). The correlation was nearly significant in Freshwater Drum ($p = 0.078$). However, there was no such correlation in Emerald Shiner ($p = 0.328$). The correlation coefficient was highest in Bullhead Minnow (0.6428) and Bluegill (0.8455), intermediate in Freshwater Drum (0.3550), Channel Catfish (0.3641), and Gizzard Shad (0.5118), and lowest in Emerald Shiner (0.0757).

***Genome Scan for Genomic Islands of Differentiation***

We aligned SNPs to reference genomes and conducted genome scans to investigate the genomic landscape of adaptive divergence. A total of 3,348 - 16,620 loci were aligned to the corresponding reference genomes with alignment rate varying from 26.4% to 97.5% depending on genetic divergence from the reference species (Table S4). Correspondingly, a total of 3 - 43 outlier SNPs were aligned with alignment rate per species varying from 25.0% to 100% (Table S5).

Genome scan results revealed highly variable genomic landscapes of population differentiation among the six species (Figure 2). In general, outlier SNPs and HMM islands in species with lower neutral differentiation were more tightly clustered and found on fewer chromosomes, whereas those in species with higher neutral differentiation were spread out across the genomes.

Bullhead Minnow, the species with highest neutral population structure, displayed a high level of baseline differentiation without obvious peaks of highly differentiated loci. We only detected 3 HMM islands on 2 chromosomes, which contained a total of 4 outlier SNPs. In Bluegill, the species with the second highest neutral population structure, we identified 83 islands that were dispersed across nearly all chromosomes (22 out of 24 chromosomes). Additionally, 36 outlier SNPs were located in 31 islands across 14 chromosomes. Freshwater Drum had an intermediate level of population differentiation and displayed a more clustered architecture of genomic islands of differentiation compared to Bullhead Minnow and Bluegill. In total, 14 islands were detected across 6 chromosomes. Of these islands, 3 islands (21%) were on chromosome 7 and 7 islands (50%) were on chromosome 17. Additionally, over half of the outlier SNPs were found on these islands on chromosome 7 and 17 (17% and 37.5%, respectively). Channel Catfish had a relatively low level of differentiation and displayed highly clustered architectures of genomic islands of differentiation. We identified 15 islands across 10 chromosomes. Almost half of the outlier SNPs were found on two islands, with 6 outlier SNPs (21%) found on an island on chromosome 20 and 4 outlier SNPs (20%) on another island on chromosome 28. Gizzard Shad had a similar level of neutral global Fstp as Channel Catfish, but we did not detect any islands of high differentiation, possibly due to its relatively low genome alignment rate (26.4%). Lastly, Emerald Shiner, the species with lowest overall neutral population differentiation, displayed the strongest signal of clustered architecture of local adaptation. In Emerald Shiner, 15 islands were detected across 4 chromosomes with 11 islands (73%) clustered on chromosome 9. Furthermore, 18 out of 22 aligned outlier SNPs (82%) were found within these islands on chromosome 9.

***Identification and Analyses of Putative Inversions***

Using local PCA in *lostruct*, we identified 21 candidate regions for putative inversions where individuals clustered into three distinct groups on PC1 and with the middle PCA cluster displaying significantly higher heterozygosity than the other two clusters (Table S7). Of all candidate regions, only the ones on chromosome 14 in Channel Catfish and chromosome 6, 9, and 19 in Emerald Shiner were characterized by elevated LD blocks extending over several Mb, while LD decayed very quickly on other chromosomes (Figure S2). However, we detected recombination suppression in both heterozygous and homozygous groups in the outlier region on chromosome 14 in Channel Catfish (Figure S3). This pattern is likely due to the effect of chromosome centers, which reduce recombination in all individuals, and is inconsistent with inversions, which should only suppress recombination in heterokaryotypes. We therefore decided to exclude this region from inversion analysis. Only the candidate regions on chromosome 6 (Figure S4), 9 (Figure 4), and 19 (Figure S5) in Emerald Shiner passed our stringent criteria and were considered as putative inversions. These three putative inversions spanned large genomic regions, 18.0, 42.7, and 25.6 Mbp, respectively (Table S7). The heterokaryotype had significantly higher heterozygosity than the two homokaryotypes (Figure 4B, S4B, S5B, and S6). Between the two homokaryotypes of all three putative inversions, there were also significant differences in heterozygosity (Figure 4B, S4B, and S5B). We assumed that the arrangement with lower heterozygosity was the derived inverted type. The putative inversion on chromosome 9 (cluster 0) was only detected in the three southern river reaches (Figure 4C), whereas the other two inversions occurred at similar frequency across all six river reaches, ranging in frequencies from 0.15 to 0.39 for the inversion on chromosome 6 (cluster 2; Figure S4C), and from 0.19 to 0.28 for the inversion on chromosome 19 (cluster 0; Figure S5C). Moreover, outlier SNPs and HMM islands were consistently associated with the putative inversion on chromosome 9 in

Emerald Shiner (Figure 2). Specifically, the average Fstp within the putative inversion is 0.02 with many Fstp values > 0.1, whereas the average Fstp outside of the inversion is 0.001. Contrastingly, no outlier SNPs or HMM islands were found within the inversions on the chromosome 6 and 19.

Analyzing datasets with and without putative inversions in Emerald Shiner produced substantially different patterns of genetic structure (Figure 5). Both PCA analyses based on all aligned SNPs and SNPs within the three identified inversions showed a similar genetic structure pattern, with six well-separated clusters. This illustrates that the clustering inferred from the full dataset is driven by these three inversions. After the SNPs in these inversions were removed, the remaining aligned loci demonstrated a lack of clustering, with panmictic population structure.

***Genomic Properties of Clusters of Outlier SNPs***

The following five chromosomes contained at least 20% of aligned outlier SNPs within a given species: (1) chromosome 2 in Bullhead Minnow; (2) chromosome 7 and 17 in Freshwater Drum; (3) chromosome 20 in Channel Catfish; and (4) chromosome 9 in Emerald Shiner (Table S8). The aligned outlier SNPs on these five chromosomes were all found within the HMM islands. Outlier SNPs in these five chromosomes were closer together in the genome than expected (Table S8, permutation test: two-sided $p < 0.001$). In particular, outlier SNPs on chromosome 9 in Emerald Shiner demonstrated extremely high clustering. These outlier SNPs were about 3.8 standard deviations closer together than the null average and at least 1 standard deviation closer compared to other chromosomes in other species.

In the above five chromosomes that we identified, we found, as expected, significantly higher Fstp values at outlier SNPs (Figure 6; Table S8). However, comparisons of $H_O$ and $D_{xy}$ between outlier SNPs and chromosomal background (non-outlier SNPs) using permutation tests showed three different patterns (Figure 6; Table S8): (1) outlier SNPs on chromosome 2 in Bullhead Minnow had similar $H_O$ and $D_{xy}$; (2) outlier SNPs on chromosome 7 and 17 in Freshwater Drum had similar $H_O$, but higher $D_{xy}$ (note that the difference in $D_{xy}$ on chromosome 7 is close to significant, two-sided $p = 0.0646$); (3) outlier SNPs on chromosome 20 in Channel Catfish and chromosome 9 in Emerald Shiner had significantly lower values of $H_O$ and $D_{xy}$. Because there was no significant difference in SNP density in windows overlapping with HMM islands and windows in the chromosomal backgrounds on all five chromosomes (Figure S7), the per-SNP $D_{xy}$ measure was considered to be equally biased between SNPs within the HMM islands and SNPs in the background and should achieve equivalent inference as the window-based approach (e.g. *pixy*). We also found significantly elevated LD within outlier SNPs in all chromosomes (Figure 6; Table S8). Taken together, these results indicate that the clusters of outlier SNPs on these five chromosomes have higher relative divergence (i.e. Fstp) than their chromosomal backgrounds; the clusters of outlier SNPs on chromosome 7 and 17 in Freshwater Drum also demonstrated higher absolute divergence ($D_{xy}$).

A total of 2, 9, and 2 GO terms were significantly enriched ($p < 0.05$) in the HMM islands on chromosome 2 in Bullhead Minnow, chromosome 17 in Freshwater Drum, and the putative inversion on chromosome 9 in Emerald Shiner, respectively (Table S9). Enriched GO terms included membrane organization, regulation of cellular component size, cell communication, and regulation of ion transmembrane transport. There were no annotated genes found within the HMM islands on chromosome 7 in Freshwater Drum and chromosome 20 in Channel Catfish.

## Discussion

### *Neutral Population Structure Reflects Differences in Life History Strategies Among Species*

We found highly variable neutral population structure among our six riverine fish species that generally reflected differences in life history strategies. For example, both Bullhead Minnow and Bluegill, which had the highest levels of genetic differentiation, are nest spawners whose eggs and larvae are not transported by currents, limiting gene flow. In contrast, Gizzard Shad and Emerald Shiner, which had the lowest levels of structure in our study, are both broadcast spawners, allowing their eggs to be carried freely by the currents, facilitating gene flow. Genetic studies on similar fish species have generally corroborated the patterns we observed, with nest spawning species such as smallmouth bass (*Micropterus dolomieu*) exhibiting high levels of genetic structure in open systems compared to broadcast spawning species such as walleye (*Sander vitreus)* (Ruzich *et al.* 2019; Euclide *et al.* 2020; 2021)

An exception to the pattern described above was Freshwater Drum, as they are migratory broadcast spawners but displayed an intermediate level of population structure, with individuals from La Grange along with some individuals from southern populations in Pool 26 and Open River forming a distinct group. One possible explanation for this pattern is limited movement of Freshwater Drum between the Illinois River, where La Grange is located, and the mainstem Mississippi River. Unfortunately, movement data for this species are generally lacking, making it difficult to corroborate this hypothesis without additional research. Channel Catfish also deviated from the expected patterns of population structure based on life history, as they are nest spawners but displayed relatively low levels of differentiation with individuals from the Open River and La Grange were slightly separated from other reaches. It is possible that the highly migratory

nature of this species mixed with potentially low spawning fidelity (Pellett *et al.* 1998) could explain the low to intermediate levels of population differentiation we observed.

### *GEA Outliers Reflect Adaptive Divergence in Response to Habitat Heterogeneity*

Most of the GEA outliers that we found were associated with environmental PC1, which had the highest loadings for temperature and turbidity. It is likely that these GEA outliers reflect adaptive divergence driven by the large latitudinal gradient that we sampled. Our study system spans two major Köppen climate zones, with pools 4, 8, and 13 in a humid continental climate characterized by warm summers and very cold winters (below 0 °C) , and Pool 26, Open River, and La Grange in a humid subtropical climate characterized by very warm and humid summers and mild winters (above 0 °C). Although we could not disentangle the effects of temperature and turbidity because they co-varied, we suspect that temperature is likely a major selective force shaping adaptive divergence in our study system given its pervasive effects across all levels of biological processes, from the biochemistry of metabolism (Deutsch *et al.* 2015) to reproduction (Pankhurst & Munday 2011) and the fact that most fish are ectotherms. Multiple studies have illustrated strong signals of adaptive divergence across temperature gradients in continuously distributed marine species, even when differentiation at neutral markers is low (Limborg *et al.* 2012; Stanley *et al.* 2018; Wilder *et al.* 2020). However, few studies have investigated temperature-mediated adaptive divergence in continuously distributed freshwater fish. Our study suggests riverine fish display patterns of adaptive divergence driven by temperature that are similar to those found in marine systems, highlighting that populations of continuously distributed riverine species may display the potential for local adaptation across their range, although experimental evidence is necessary for further validation.

While GEA outliers for most species in our study were generally associated with environmental PC1, outliers in Freshwater Drum were associated with environmental PC2, which displayed high loadings for measures of productivity including chlorophyll and nitrogen, and to a lesser extent, flow. This result suggests that the environmental variables influencing adaptive divergence in Freshwater Drum may differ from our other study species. Specifically, it is possible that Freshwater Drum is more affected by eutrophication caused by agricultural runoff compared to our other study species. Numerous studies have demonstrated that fish species respond differently to eutrophication depending on their life histories (Tammi *et al.* 1999; Hondorp *et al.* 2010; Jacobson *et al.* 2017). Alternatively, Freshwater Drum might have evolved in response to other variables that co-vary with agricultural inputs, which is outside of our datasets to address. A combination of DNA, RNA, and functional methodologies with field experiments will be needed to clarify the genes and mechanisms shaping adaptation in nature {PardoDiaz:2015em}

***Gene Flow Influences the Genomic Architecture of Local Adaptation***

Theoretical studies and genetic simulations predict that increased gene flow will lead to increasingly concentrated genomic architecture of adaptation (Yeaman & Whitlock 2011; Via 2012; Yeaman 2013). However, few empirical studies have tested this hypothesis in natural populations, and the results of these empirical studies have not necessarily supported theoretical work (Burri *et al.* 2015; Renaut *et al.* 2019). Our study included six fish species spanning a wide gradient of genetic differentiation (overall $F_{ST}$ from 0.0004 – 0.07), indicating highly variable levels of gene flow. Gene flow appeared to be correlated with the landscape of adaptive divergence, as species with high gene flow (Emerald Shiner, Channel Catfish and Freshwater

509 Drum) displayed more clustered architecture of adaptation than low gene flow species (Bullhead
510 Minnow and Bluegill). Our results are somewhat similar to a recent study which examined
511 adaptive divergence of four flatfish species across a strong salinity gradient in the Baltic Sea (Le
512 Moan *et al.* 2019). Specifically, Le Moan *et al.* (2019) found more evidence of clustered
513 architectures of adaption in species displaying low genetic differentiation compared to those
514 displaying higher differentiation. In general, they concluded that genome-wide pattern of
515 divergence was mostly shaped by the complex demographic history in addition to gene flow and
516 selection. However, Le Moan *et al.* (2019) studied species with similar life history traits (all
517 pelagic spawners with long larval dispersal) and sampled a much smaller gradient of genetic
518 differentiation (overall $F_{ST}$ from 0.005 – 0.02) than our study. Furthermore, examining the
519 effects of gene flow on landscapes of adaptive differentiation was not a central goal of their
520 study.

521 Though our finding that clustered genomic architectures of adaptation (i.e., genomic islands of
522 differentiation) increase with gene flow is in line with theoretical expectations and the results
523 from Le Moan *et al.* (2019), this finding is inconsistent with other studies positing that islands of
524 differentiation are the result of variation in intrinsic recombination rate rather than the
525 combination of gene flow and selection (Roesti *et al.* 2012; Renaut *et al.* 2019). In fact, there is
526 considerable debate over the mechanisms that lead to islands of differentiation, with past
527 research suggesting that these islands can be caused by variation in recombination rates (Roesti
528 *et al.* 2012; Renaut *et al.* 2019), linked selection (Cruickshank & Hahn 2014; Burri *et al.* 2015),
529 divergence hitchhiking (Via 2012), genomic rearrangements including chromosomal inversions
530 (Rogers *et al.* 2013; Yeaman 2013), and elevated linkage preserving locally adapted alleles
531 (Yeaman & Whitlock 2011). While the cluster of islands on chromosome 9 in Emerald Shiner

appears to be caused by an inversion (see following section), the mechanisms that created the islands on the other four chromosomes we identified are less clear.

To investigate the genomic mechanisms that created the islands (clusters of outlier SNPs) on the remaining four chromosomes, we calculated the following four metrics: $F_{ST}$, $H_O$, $D_{xy}$, and LD, and compared these metrics between outlier SNPs and chromosomal background. While islands on all four chromosomes displayed elevated $F_{ST}$ and LD as expected, we did observe differences in the remaining two metrics among the chromosomes. $H_O$ was similar to neutral regions on three out of four chromosomes, and $D_{xy}$ was elevated or similar to neutral regions on the same three chromosomes. While LD can be a useful metric for understanding genomic processes, we found that it did not help us differentiate the mechanisms responsible for creating islands in the current study and instead focused on $H_O$ and $D_{xy}$. Estimates of $H_O$ and $D_{xy}$ suggest that the islands on one chromosome with reduced diversity (islands on chromosome 20 in Channel Catfish) may have been created by linked selection (Cruickshank & Hahn 2014; Burri *et al.* 2015), while the islands on chromosome 2 in Bullhead Minnow and chromosomes 7 and 17 in Freshwater Drum may have arisen through divergent selection (Kulmuni & Westram 2017).

Islands created by divergent selection are hypothesized to have a major role in facilitating adaptive divergence with gene flow, whereas islands created by linked selection are likely a result of the underlying genomic landscape and do not necessarily reflect recent adaptive divergence (Cruickshank & Hahn 2014; Burri *et al.* 2015). Thus, it is extremely important to differentiate these two types of islands when investigating adaptive divergence. The most effective way distinguish between these island types is to compare measures of absolute diversity, as islands created by linked selection should show reduced absolute diversity while

554 islands created by divergent selection should not (Cruickshank & Hahn 2014; Irwin *et al.* 2016).
555 Applying this method to our data provided evidence that islands on three of the chromosomes in
556 our study were created by divergent selection and are likely involved in adaptive divergence with
557 gene flow, whereas the islands on the other chromosome were likely a result of ancient linked
558 selection that acted to reduce diversity in particular genomic regions but is not influencing
559 contemporary adaptive divergence. Other intrinsic genomic factors can lead to a heterogeneous
560 genomic landscape of differentiation as well, such as variations in recombination, mutation, and
561 gene density {Ravinet:2017dz}, however, investigation on these genomic factors is beyond the
562 scope of the current study. Since the past demographic history of the studied populations is
563 unknown, we also cannot rule out the role of genetic incompatibility between populations
564 {Ravinet:2017dz}{ Schumer:2018hc}. In ecological speciation, reproductive isolation evolves as
565 a result of divergent selection on habitat use {Via:2008bc}. However, the genetic changes
566 responsible for the initial barriers to gene flow between populations is often unknown because of
567 the confounding effects of genetic differences that accumulate over time {Via:2008bc}. A full
568 exploration of demographic history of the study populations could help us gain a mechanistic
569 understanding of the evolutionary processes that influence the genomic landscape of adaptation,
570 but, again, this was outside the scope of our study.

571 ***A Chromosomal Inversion Facilitates Local Adaptation with High Gene Flow in Emerald***
572 ***Shiner***

573 Our results and those of previous empirical and theoretical studies suggest that divergent
574 selection can result in clusters of adaptive loci through mechanisms such as divergence
575 hitchhiking when gene flow is relatively high (Yeaman & Whitlock 2011; Via 2012). However,

when gene flow is extremely high, it is likely that additional genomic mechanisms, such as structural polymorphisms, may be required to protect clusters of adaptive loci from among-population recombination caused by gene flow (Yeaman & Whitlock 2011; Rogers *et al.* 2013; Yeaman 2013; Tigano & Friesen 2016). The gradient of gene flow sampled in our study presents an excellent opportunity to test this hypothesis. In our study, clustered architectures of adaptation were common in species with relatively high gene flow, such as Channel Catfish and Freshwater Drum (average overall $F_{ST}$ = 0.004), but these clustered architectures did not appear to be associated with structural polymorphisms. In contrast, in Emerald Shiner, the species with highest gene flow (overall $F_{ST}$ = 0.0004), nearly all of the adaptive loci identified were found in a single genomic region that displayed strong evidence of a chromosomal inversion. Taken together, our results provide novel empirical evidence to support the theory that chromosomal inversions are important for facilitating adaptive divergence in systems with extremely high gene flow.

Our study also adds to the growing body of evidence that chromosomal inversions are important for facilitating adaptive divergence in continuously distributed fish species. Inversions putatively involved in adaptive divergence have been documented in many fishes including Atlantic cod (*Gadus morhua*) (Kirubakaran *et al.* 2016), lingcod (*Ophiodon elongatus*) (Longo *et al.* 2020), rainbow trout (*Oncorhynchus mykiss*) (Arostegui *et al.* 2019), Pacific herring (*Clupea pallasii*) (Petrou *et al.* 2021), Atlantic silverside (*Menidia menidia*) (Wilder *et al.* 2020), and European plaice (*Pleuronectes platessa*) (Le Moan *et al.* 2021). However, all of these studies were conducted on marine fish or salmonids, making our study the first to provide evidence of a putative adaptive inversion in a non-salmon freshwater fish. It is likely that the lack of previous evidence for adaptive inversions in freshwater fish is due to the generally higher genetic structure

observed in these species, making inversions less necessary for adaptation. However, our study illustrates that inversions are likely a larger component of adaptive divergence in freshwater fish than previously assumed, highlighting the importance of future studies aimed at characterizing them in additional species.

Although inferring the functional significance of the putatively adaptive inversion that we detected is difficult, it is possible to speculate on its role in facilitating adaptive divergence. The putatively derived variant of this inversion was only detected in the three southern river reaches in our study, which are substantially warmer and more turbid than northern reaches. This suggests that the derived inversion variant may have evolved and increased in frequency as Emerald Shiner adapted to warmer and/or more turbid environments in more southern regions. Inversions putatively linked to adaptive divergence across environmental and latitudinal gradients have also been identified in marine species such as lingcod (Longo *et al.* 2020) and Atlantic silverside (Wilder *et al.* 2020), but these studies faced similar difficulties when attempting to describe the functional significance of the adaptive inversions they identified. Future research combining whole genome resequencing with physiological challenge studies would be useful for assessing the functional role of these inversions in the process of adaptive divergence.

## Conclusions

Our data from six riverine fish species in the Upper Mississippi River System displaying a large gradient of life history strategies suggest that higher gene flow leads to increasingly concentrated genomic architectures of adaptation. Our results provide evidence that the mechanisms that create islands of differentiation can be highly variable across species, with both ancient linked

621 selection and more contemporary divergent selection playing important roles in creating genomic
622 islands of differentiation. Additionally, our study provides further evidence that chromosomal
623 inversions are important for facilitating adaptive divergence in continuously distributed species
624 with extremely high gene flow and also sheds light on the documented importance of inversions
625 in freshwater fish. Taken together, our findings represent a significant contribution towards
626 understanding the evolutionary processes that influence the genomic landscape of adaptation in
627 non-model organisms, though the generality of our findings is constrained due to the fact that we
628 only investigated 6 species. It is also important to note that we did not explicitly account for
629 variable demographic histories among species that could influence our results (Ravinet *et al.*
630 2017). A full exploration of demographic history using tools such as approximate Bayesian
631 computational analysis would help illuminate how demographic history could influence
632 landscapes of adaptive divergence and is a ripe area for future research, but this was outside the
633 scope of our study. Additionally, our study used RADseq, which does not assess the full suite of
634 adaptive divergence across the genome. Future studies should focus on whole genome
635 resequencing to better understand variation within genomic islands of differentiation and to
636 assess the functional role of these islands in promoting adaptive divergence.

637 **Acknowledgements**

638 This study was funded by the U.S. Army Corps of Engineers Upper Mississippi River
639 Restoration Program (grant number 96514790661277) and supported by the Turing High
640 Performance Computing cluster at Old Dominion University. We thank U.S. Geological Survey
641 Upper Midwest Environmental Sciences Center, Iowa Department of Natural Resources,
642 Minnesota Department of Natural Resources, Wisconsin Department of Natural Resources,
643 Illinois Natural History Survey, and Missouri Department of Conservation for their great efforts

in sample collection. We thank Kristen Gruenthal for assistance in the laboratory, and Kathi Jo Jankoswki for help with environmental data access. We are grateful to Uland Thomas for allowing us to use his fish profile photos of the six study species and Hyeon Jeong Kim for help with photo editing. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## References

Abdellaoui A, Hottenga J-J, de Knijff P *et al.* (2019) Population structure, migration, and diversifying selection in the Netherlands. *European Journal of Human Genetics*, **21**, 1277–1285.

Aguirre Liguori JA, Gaut BS, Jaramillo Correa JP *et al.* (2019) Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Molecular Ecology*, **24**, 2663–17.

Arostegui MC, Quinn TP, Seeb LW, Seeb JE, McKinney GJ (2019) Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. *Molecular Ecology*, **28**, 1412–1427.

Bolnick DI, Nosil P (2007) Natural selection in populations subject to a migration load. *Evolution*, **61**, 2229–2243.

Bradley KM, Breyer JP, Melville DB *et al.* (2011) An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3 Genes|Genomes|Genetics*, **1**, 3–9.

Brauer CJ, Hammer MP, Beheregaray LB (2016) Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, **25**, 5093–5113.

Burri R, Nater A, Kawakami T *et al.* (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, **25**, 1656–1665.

Cayuela H, Rougemont Q, Laporte M *et al.* (2020) Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Molecular Ecology*, **29**, 2379–2398.

Chang CC, Chow CC, Tellier LC *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 559–16.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Deutsch C, Ferrel A, Seibel B, Pörtner H-O, Huey RB (2015) Climate change tightens a metabolic constraint on marine habitats. *Science*, **348**, 1132–1135.

Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.

Euclide PT, MacDougall T, Robinson JM *et al.* (2021) Mixed-stock analysis using Rapture genotyping to evaluate stock-specific exploitation of a walleye population despite weak genetic structure. *Evolutionary Applications*.

Euclide PT, Ruzich J, Hansen SP, Zorn TG, Larson WA (2020) Genetic structure of Smallmouth Bass in the Lake Michigan and Upper Mississippi River drainages relates to habitat, distance, and drainage boundaries. *Transactions of the American Fisheries Society*, 383–397.

Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, **27**, 2215–2233.

Gagnaire P-A (2020) Comparative genomics approach to evolutionary process connectivity. *Evolutionary Applications*, **13**, 1320–1334.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, **188**, 379–397.

Hofer T, Foll M, Excoffier L (2012) Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, **13**, 1–13.

Hoffmann AA, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics*, **39**, 21–42.

Hondorp DW, Breitburg DL, Davias LA (2010) Eutrophication and fisheries: separating the effects of nitrogen loads and hypoxia on the pelagic-to-demersal ratio and other measures of landings composition. *Marine and Coastal Fisheries*, **2**, 339–361.

Huang K, Rieseberg LH (2020) Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, **11**, 1–13.

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH (2020) Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, **29**, 2535–2549.

Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL (2016) Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, **25**, 4488–4507.

Jacobson PC, Hansen GJA, Bethke BJ, Cross TK (2017) Disentangling the effects of a century of eutrophication and climate warming on freshwater lake fish assemblages. *PLoS ONE*, **12**, e0182667.

Jacquemin SJ, Doll JC, Pyron M, Allen M, Owen DAS (2015) Effects of flow regime on growth rate in freshwater drum, *Aplodinotus grunniens*. *Environmental Biology of Fishes*, 993–1003.

Johannesson K, Le Moan A, Perini S, André C (2020) A darwinian laboratory of multiple contact zones. *Trends in Ecology & Evolution*, 1–16.

Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Kirubakaran TG, Grove H, Kent MP *et al.* (2016) Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 2130–2143.

Knief U, Hemmrich-Stanisak G, Wittig M *et al.* (2016) Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome biology*, 1–22.

Kulmuni J, Westram AM (2017) Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology*, 3093–3103.

Laayouni H (2003) The evolutionary history of *Drosophila buzzatii*. XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. *Molecular Biology and Evolution*, **20**, 931–944.

Le Moan A, Bekkevold D, Hemmer-Hansen J (2021) Evolution at two time frames: ancient structural variants involved in post-glacial divergence of the European plaice (Pleuronectes platessa). *Heredity*, **126**, 668–683.

Le Moan A, Gaggiotti O, Henriques R *et al.* (2019) Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, **26**, 4452–43.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, **1303.3887v2**, 1–3.

Li H, Ralph P (2019) Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics*, **211**, 289–304.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Limborg MT, Helyar SJ, de Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.

Longo GC, Lam L, Basnett B *et al.* (2020) Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. *Evolutionary Applications*, 2536–2554.

Lotterhos KE (2019) The Effect of Neutral Recombination Variation on Genome Scans for Selection. *G3 Genes|Genomes|Genetics*, **9**, 1851–1867.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of $F_{ST}$ outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Marques DA, Lucek K, Meier JI *et al.* (2016) Genomics of rapid incipient speciation in sympatric Threespine Stickleback. *PLoS Genetics*, **12**, e1005887–34.

McKinney GJ, Pascal CE, Templin WD *et al.* (2020) Dense SNP panels resolve closely related Chinook salmon populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **77**, 451–461.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353.

Narum SR, Hess JE (2011) Comparison of $F_{ST}$ outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, **98**, 12084–12088.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Pankhurst NW, Munday PL (2011) Effects of climate change on fish reproduction and early life history stages. *Marine and Freshwater Research*, **62**, 1015–1026.

772 Pellett TD, Van Dyck GJ, Adams JV (1998) Seasonal migration and homing of Channel Catfish
773 in the lower Wisconsin River, Wisconsin. *North American Journal of Fisheries*
774 *Management*, 85–95.
775 Petrou EL, Fuentes-Pardo AP, Rogers LA *et al.* (2021) Functional genetic diversity in an
776 exploited marine species and its relevance to fisheries management. *Proceedings of the*
777 *Royal Society B: Biological Sciences*, 20202398.
778 Prive F, Aschard H, Ziyatdinov A, Blum MGB (2018) Efficient analysis of large-scale genome-
779 wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*, **34**, 2781–2787.
780 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association
781 and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–
782 575.
783 Ravinet M, Faria R, Butlin RK *et al.* (2017) Interpreting the genomic landscape of speciation: a
784 road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, **30**, 1450–1477.
785 Renaut S, Grassa CJ, Yeaman S *et al.* (2019) Genomic islands of divergence are not affected by
786 geography of speciation in sunflowers. *Nature Communications*, 1–8.
787 Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during
788 speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B:*
789 *Biological Sciences*, **367**, 354–363.
790 Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology &*
791 *Evolution*, **16**, 351–358.
792 Roda F, Walter GM, Nipper R, Ortiz-Barrientos D (2017) Genomic clustering of adaptive loci
793 during parallel evolution of an Australian wildflower. *Molecular Ecology*, **26**, 3687–3699.
794 Roesti M (2018) Varied Genomic Responses to Maladaptive Gene Flow and Their Evidence.
795 *Genes*, **9**, 298–16.
796 Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary
797 diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular*
798 *Ecology*, **21**, 2852–2862.
799 Roesti M, Moser D, Berner D (2013) Recombination in the threespine stickleback genome -
800 patterns and consequences. *Molecular Ecology*, **22**, 3014–3027.
801 Rogers SM, Mee JA, Bowles E (2013) The consequences of genomic architecture on ecological
802 speciation in postglacial fishes. *Current Zoology*, **59**, 53–71.
803 Ruzich J, Turnquist K, Nye N, Rowe D, Larson WA (2019) Isolation by a hydroelectric dam
804 induces minimal impacts on genetic diversity and population structure in six fish species.
805 *Conservation Genetics*, **20**, 1421–1436.
806 Rypel AL, Bayne DR, Mitchell JB (2006) Growth of Freshwater Drum from lotic and lentic
807 habitats in Alabama. *Transactions of the American Fisheries Society*, 987–997.
808 Samuk K, Owens GL, Delmore KE *et al.* (2017) Gene flow and selection interact to promote
809 adaptive divergence in regions of low recombination. *Molecular Ecology*, **26**, 4378–4390.
810 Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**,
811 23–35.
812 Stanley RRE, DiBacco C, Ben Lowen *et al.* (2018) A climate-associated multispecies cryptic
813 cline in the northwest Atlantic. *Science Advances*, **4**, eaaq0929.
814 Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across
815 plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the*
816 *Royal Society B: Biological Sciences*, **367**, 364–373.

Tammi J, Lappalainen A, Mannio J, Rask M, Vuorenmaa J (1999) Effects of eutrophication on fish and fisheries in Finnish lakes: a survey based on random sampling. *Fisheries Management and Ecology*, **6**, 173–186.
Thompson NF, Anderson EC, Clemento AJ *et al.* (2020) A complex phenotype in salmon controlled by a simple change in migratory timing. *Science*, 609–613.
Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*, **25**, 2144–2164.
Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, **69**, 1476–1486.
Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 451–460.
Wellenreuther M, Bernatchez L (2018) Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, **33**, 427–440.
Wilder AP, Palumbi SR, Conover DO, Therkildsen NO (2020) Footprints of local adaptation span hundreds of linked genes in the Atlantic silverside genome. *Evolution Letters*, **4**, 430–443.
Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, **110**, E1743–E1751.
Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.

## Data Accessibility

Demultiplexed RAD sequencing data used in this study are archived in the NCBI Sequence Read Archive with a BioProject ID, PRJNA674918. Sample meta information along with the sequence accession numbers can be found in Table S10. The *vcf* files (post filtering) and genepop files (neutral SNPs after thinning) can be found on Dryad. Other data files and bioinformatic scripts supporting this article will be available on Github (https://github.com/melodysyue/MissR_geneflow).

## Author Contributions

YS, KLB, AB and WAL conceived of the study, designed the study, and coordinated the study. YS and WD carried out the molecular lab work. YS and WAL conducted data analyses and drafted the manuscript; GJM helped interpreted the results regarding putative inversions. MC

supervised the project. All authors commented the manuscript and gave final approval for publication.

## Conflict of Interest

The authors declare that they have no conflict of interests.

**Figure 1.** (A) Map of the six study reaches along the Upper Mississippi River System, (B) key reproduction-related life history traits of the six study species, and (C) positions of the six study reaches in the environmental space of 20 variables using PCA biplot. See Table S1 for details of life history traits and Table S2 for details of environmental data. Use of fish images is permitted by Uland Thomas.

**Figure 2.** Manhattan plots depicting the genomic landscape of differentiation (Fstp, corrected $F_{ST}$) across the genomes for the six study species. Species are ordered based on neutral population differentiation, with neutral global Fstp values labeled next to the species name. Outlier SNPs (union of $F_{ST}$ outliers and GEA outliers) are in red, genomic islands of differentiation identified using HMM after filtering are in blue, and identified putative inversions are in purple. Reference genomes and alignment summary can be found in Table S4.

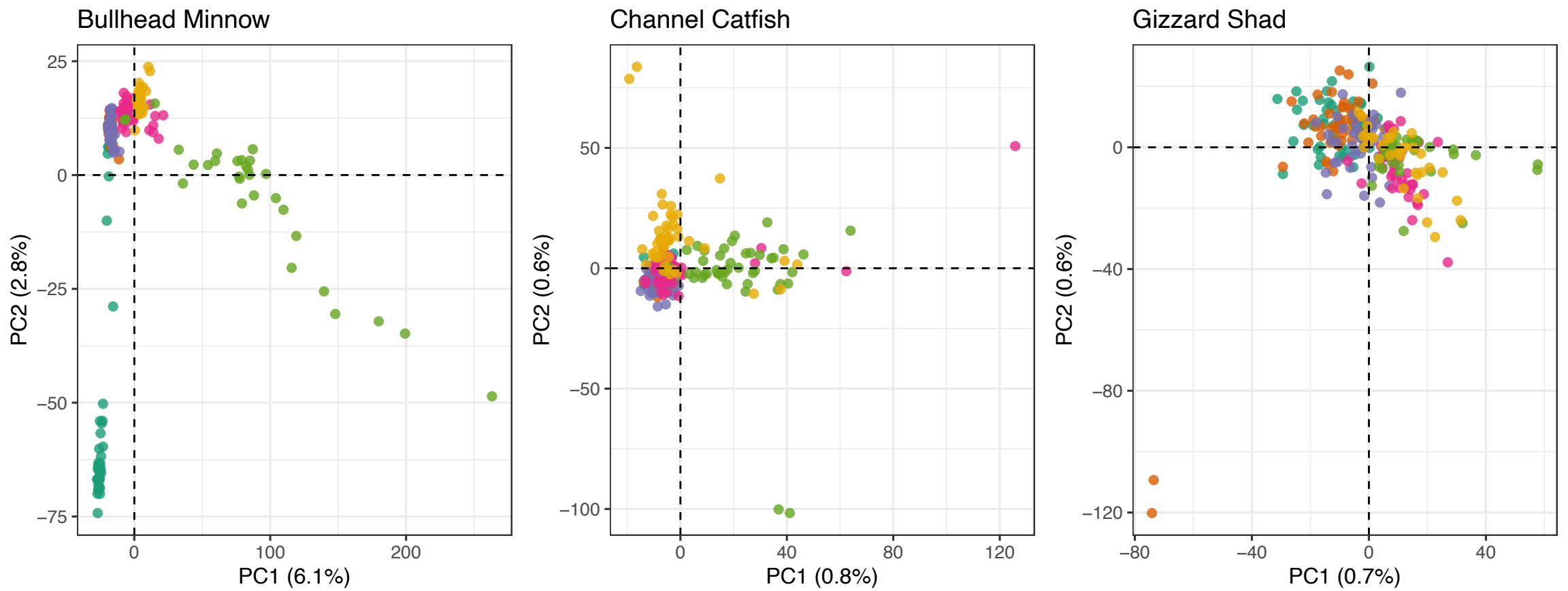
**Figure 3.** Principal component analyses using neutral SNPs (after thinning) for the six study species. The percentage of variance explained by each principal component (PC) is labeled on the *x*- and *y*- axes.

**Figure 4.** Characterization of putative inversion on chromosome 9 in Emerald Shiner. (A) PCA based on SNPs within the putative inversion region. Three clusters identified using k-means clustering correspond to two homozygote groups (blue and red) and a heterozygote group (purple). The discreteness of the clustering was calculated by the proportion of the between-cluster sum of squares over the total using the R function *kmeans* in *adegenet*. (B) Observed individual heterozygosity in each PCA cluster. Significance was assessed using Wilcoxon tests with alpha level of 0.05. Note: *** = 0.001. (C) Genotype frequency distribution for putative inversion across six study reaches. Bars represent the proportion of individuals belonging to a PCA cluster. (D) and (E) are LD heatmaps for chromosome 9 using all individuals (D) and only individuals homozygous for the more common orientation (E).

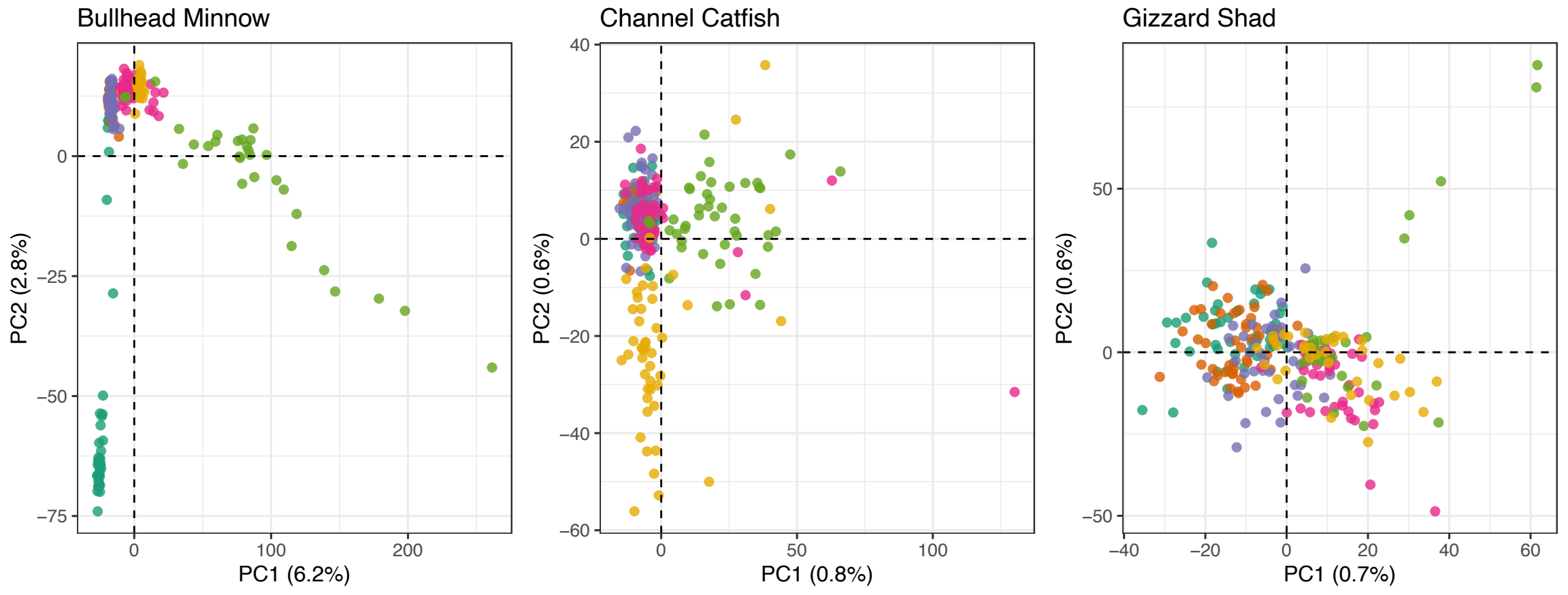
**Figure 5.** Principal component analyses for Emerald Shiner using different sets of loci: (A) All aligned SNPs (3,348 SNPs); (B) Putative inversions on chromosome 6, 9 and 19 (228 SNPs); (C) After the removal of three putative inversions (3,120 SNPs). The percentage of variance explained by each principal component (PC) is labeled on the *x*- and *y*- axes.

**Figure 6.** Comparisons of corrected $F_{ST}$ (Fstp), heterozygosity ($H_O$), absolute divergence ($D_{xy}$), and LD ($r^2$) between outlier SNPs (Outliers, red) and non-outlier SNPs(Background, gray) on the corresponding chromosomes for chromosomes clusters of outlier SNPs, including on chromosome 2 in Bullhead Minnow, chromosome 7 and 17 in Freshwater Drum, chromosome 20 in Channel Catfish, and chromosome 9 in Emerald Shiner. Significance was assessed using permutation tests (10,000 permutations) with alpha level of 0.05. Note: *** = 0.001, ** = 0.01, * = 0.05, n.s = not significant.

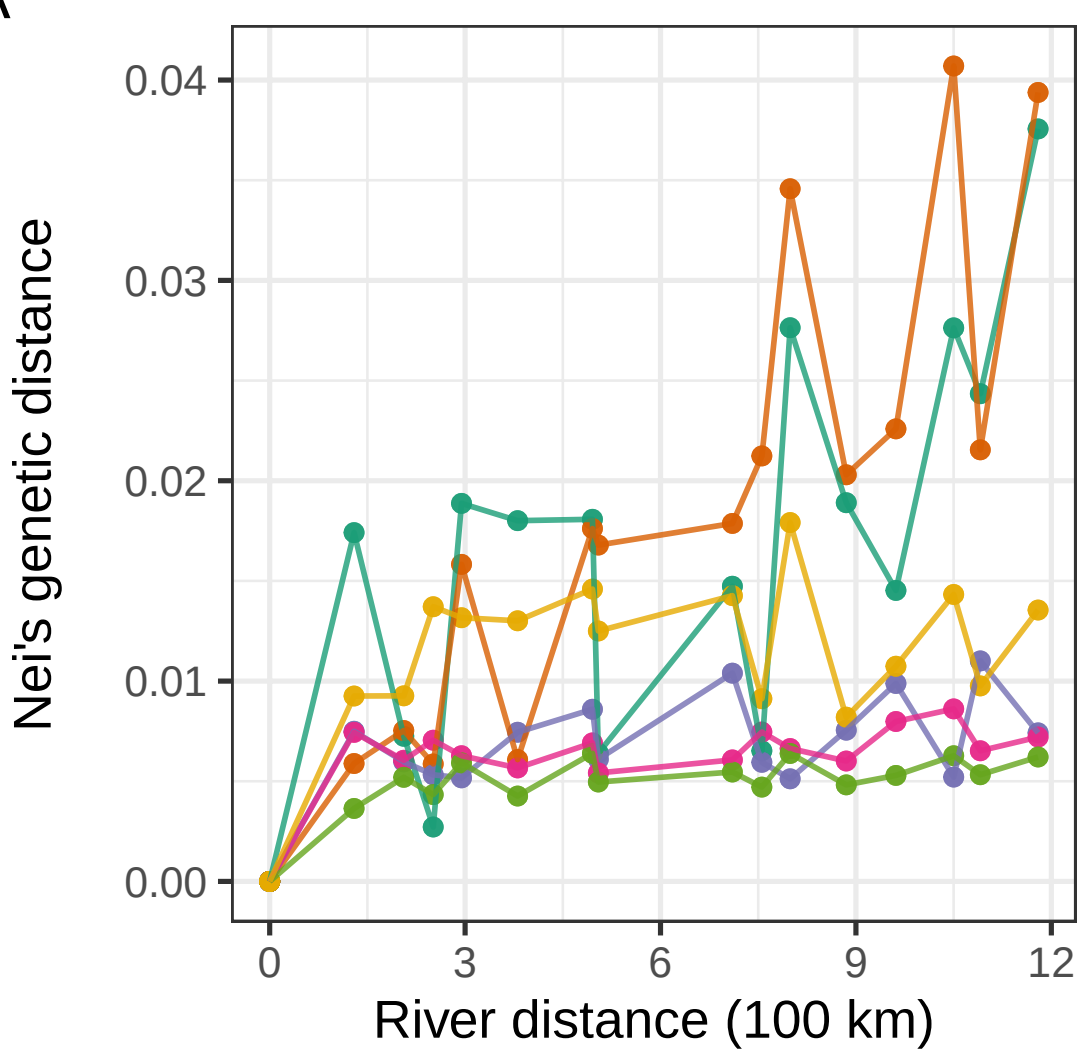
**PCAs with neutral SNPs (WITHOUT putative siblings removed)**



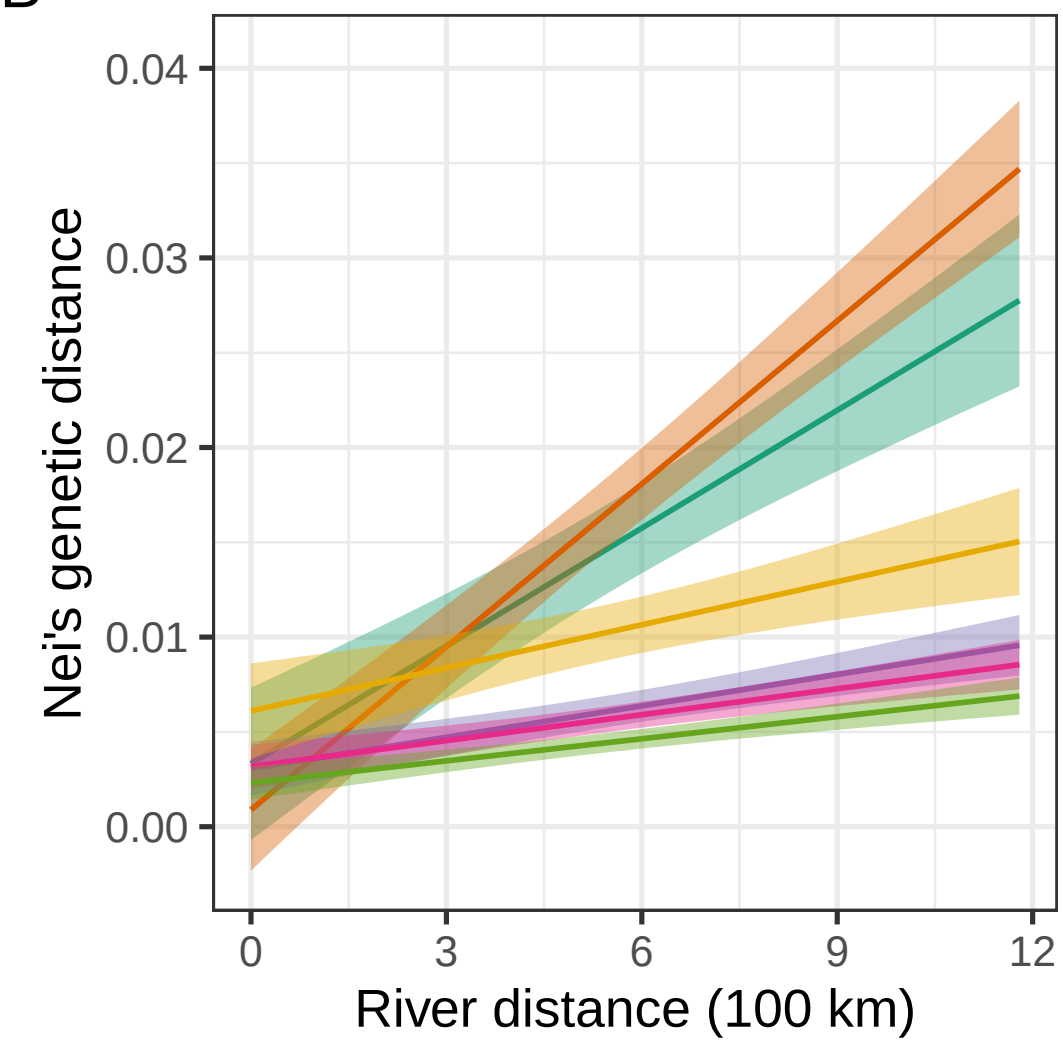

**PCAs with neutral SNPs (WITH putative siblings removed)**



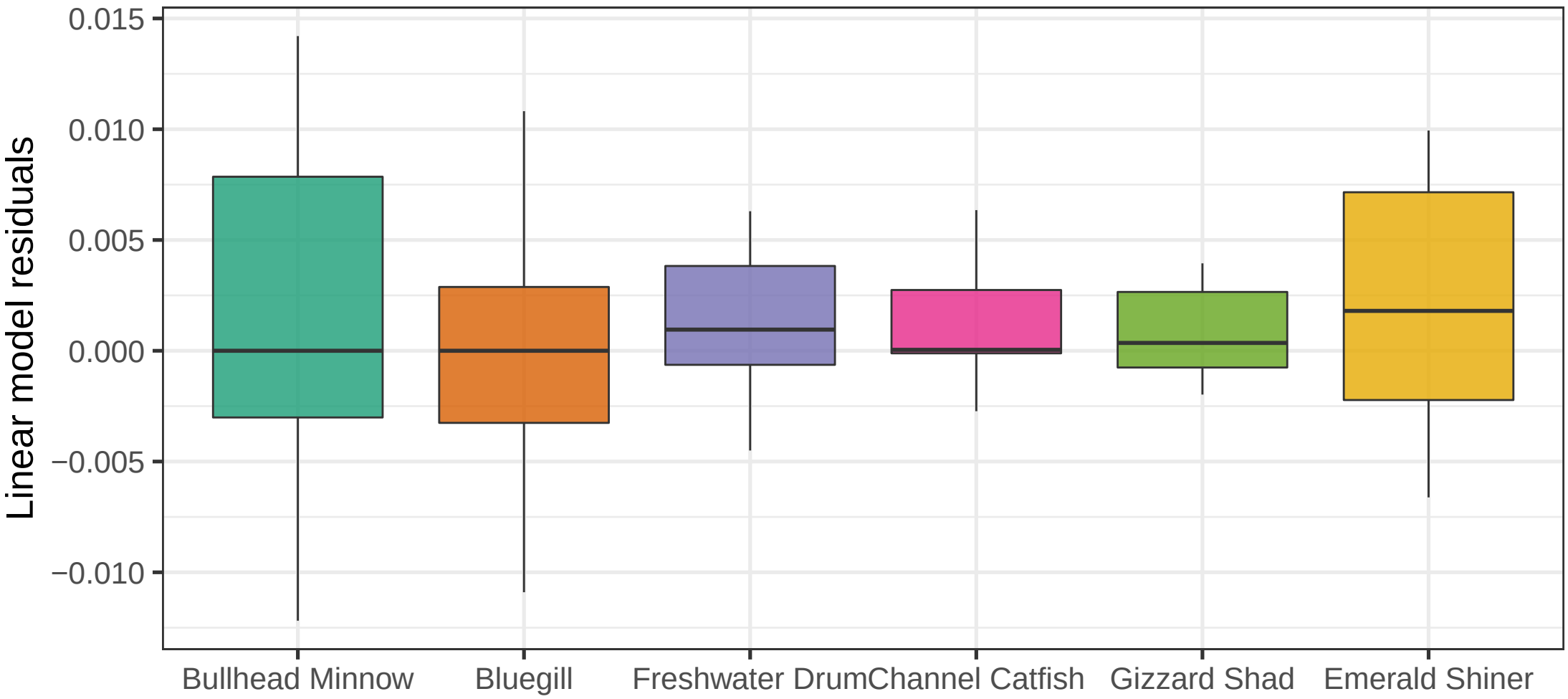
A

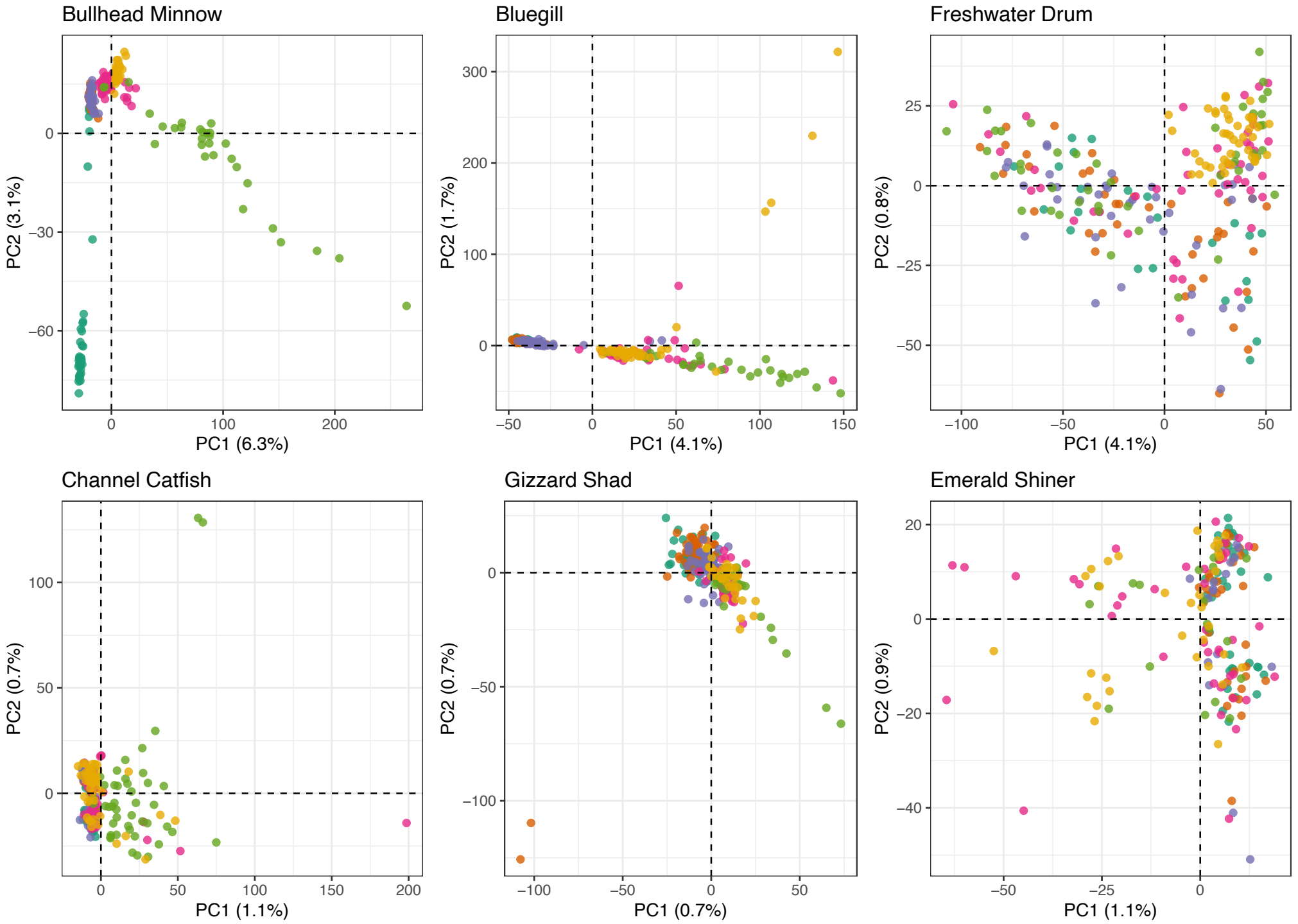

Nei's genetic distance
0.04
0.03
0.02
0.01
0.00
0
3
6
9
12
River distance (100 km)

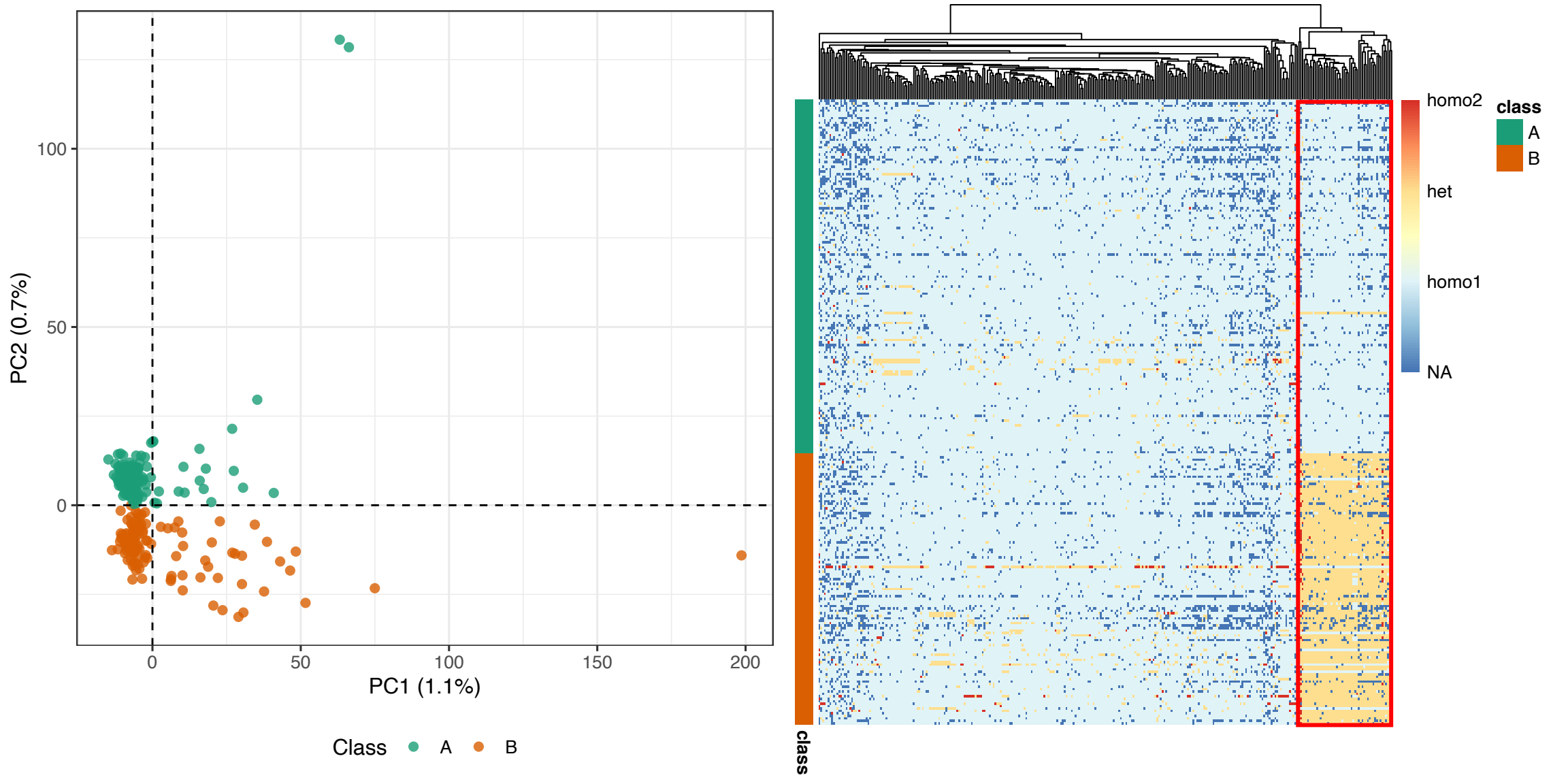
B


Nei's genetic distance
0.04
0.03
0.02
0.01
0.00
0
3
6
9
12
River distance (100 km)

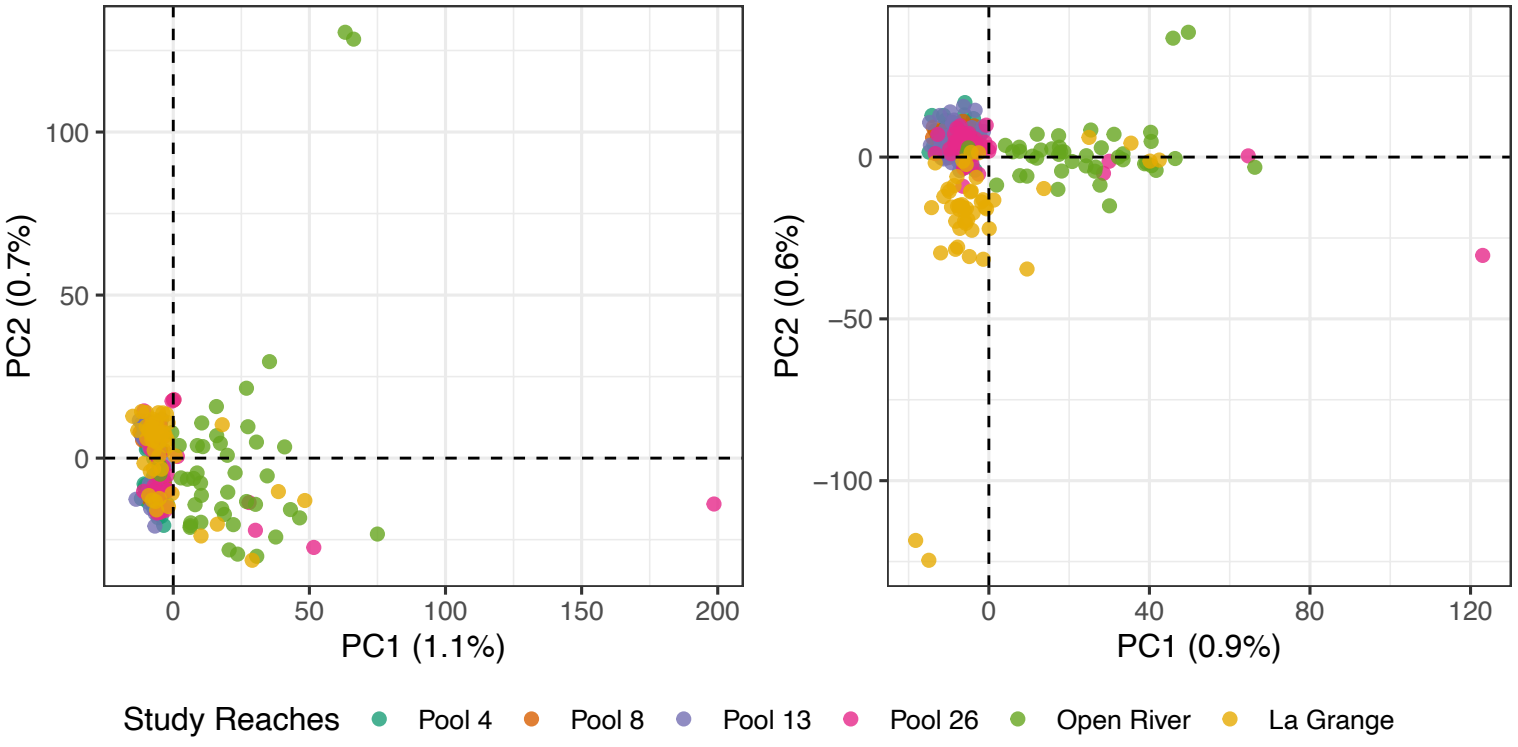
C


Linear model residuals
0.015
0.010
0.005
0.000
−0.005
−0.010
Bullhead Minnow
Bluegill
Freshwater Drum
Channel Catfish
Gizzard Shad
Emerald Shiner


Bullhead Minnow
PC2 (3.1%)
PC1 (6.3%)
Bluegill
PC2 (1.7%)
PC1 (4.1%)
Freshwater Drum
PC2 (0.8%)
PC1 (4.1%)
Channel Catfish
PC2 (0.7%)
PC1 (1.1%)
Gizzard Shad
PC2 (0.7%)
PC1 (0.7%)
Emerald Shiner
PC2 (0.9%)
PC1 (1.1%)
Study Reaches
Pool 4
Pool 8
Pool 13
Pool 26
Open River
La Grange


100
50
0
PC2 (0.7%)
0
50
100
150
200
PC1 (1.1%)
Class
A
B
class
homo2
het
homo1
NA
class
A
B

## Channel Catfish WITHOUT vs. WITH removing SNPs of high LD (316 SNPs)




## Emerald Shiner WITHOUT vs WITH removing SNPs of high LD (101 SNPs)