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

Running Title: Gene flow influences genomic 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 among-population recombination and thus be selected for, but few studies have tested this hypothesis 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 then conducted genome scans for genomic islands of divergence to examine the distribution of adaptive loci and investigated whether these loci were found in inversions. We found that gene flow varied among species, and adaptive loci were clustered more tightly in species with higher gene flow. For example, the two species with the highest overall F_{ST} (0.03 - 0.07) and therefore lowest gene flow showed little evidence of clusters of adaptive loci, with adaptive loci spread uniformly across the genome. In contrast, nearly all adaptive loci in the species with the lowest F_{ST} (0.0004) were found in a single large putative inversion. Two other species with intermediate gene flow (F_{ST} ~ 0.004) also showed clustered genomic architectures, with most islands of divergence clustered on a few chromosomes. These results provide important empirical evidence to support the hypothesis that increasingly clustered architectures of local adaptation are associated with high gene flow. Our study utilized a unique system with species spanning a large gradient of life histories to highlight the importance of gene flow in shaping adaptive divergence.

Keywords: Freshwater fishes, local adaptation, gene flow, genomic islands of divergence, 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, 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 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 population connectivity and genetic diversity by introducing novel genetic variation but it can also impede local adaptation by introducing maladaptive foreign alleles into a 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). Inversions are generally not deleterious and do not impact gene function unless the inversion breakpoint occurs within a gene (Faria *et al.* 2019). However, 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 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 temperature,
101 turbidity, productivity and flow (Figure 1A & C). Our study system provides a unique
102 opportunity to compare the genomic architecture of local adaptation in a natural environment for
103 species with contrasting life histories and assess the influence of gene flow on genomic
104 architecture. Specifically, we test the hypothesis that the genomic islands of differentiation are
105 less frequent but larger for species with relatively high gene flow, whereas genomic islands are
106 more numerous and dispersed throughout the genome for species with low degrees of gene flow.
107 Our multi-species approach investigating six species inhabiting the same environments is unique,
108 as most previous studies have focused on closely related species pairs or ecotypes (Nadeau *et al.*
109 2012; Renaut *et al.* 2012) rather than divergent species inhabiting the same environments.

110 **Materials and Methods**

111 ***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); 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).

Identification of GEA Outliers and Putatively Neutral SNPs

Recent studies have suggested that genotype-environment association (GEA) methods are more robust for identifying adaptive loci than traditional differentiation-based methods (Rellstab *et al.* 2015; Forester *et al.* 2018). Differentiation-based outlier tests 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, providing a more complete view of the genomic landscape of adaptation (Eckert *et al.* 2010; Brauer *et al.* 2016; Forester *et al.* 2018). For these reasons, we focused on identifying putatively adaptive loci (henceforth "adaptive loci") using three GEA methods: 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 putatively adaptive SNPs as the GEA outliers (henceforth "GEA outliers") identified by at least two GEA methods. To determine which environmental PC that 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.

To identify datasets of putatively neutral SNPs (henceforth "neutral SNPs") for each species, we combined results from GEA analyses with results from additional differentiation-based outlier tests. While differentiation-based outlier tests may produce a large number of false positives, they are still useful for conservatively identifying neutral SNPs (Holderegger *et al.* 2006). Therefore, we ran Bayescan, Arlequin, OutFLANK and *pcadapt* on each species (see

Supplementary Methods for details). We defined neutral SNPs as those that were not identified as outliers by any of the aforementioned seven methods.

Neutral Genetic Differentiation

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

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. Channel Catfish is the only species with a high-quality reference genome available in our study. For the other five species, we used high-quality reference genomes (chromosome-level assemblies) from closely related species (Table S3). 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 Fstp per locus 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). Each of these identified 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 (i.e., genomic islands of differentiation), but many did not show especially high levels of differentiation and may be false positives. Therefore, we chose to retain only the genomic islands that contained at least one differentiation outlier SNP identified by Bayescan, Arlequin or OutFLANK. 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 genomic islands in situations where a chromosome only had one island and this island had only one SNP. We also removed islands located on unplaced scaffolds.

Identification and Analysis of Putative Inversions

To identify putative inversions, 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 40-dimension space

multidimensional scaling (MDS) analysis 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 (0 and 2) representing two homozygous groups for alternative orientations and the middle cluster (1) representing the heterozygous group between inversion haplotypes (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 (α = 0.05) to further confirm the middle group had significantly higher heterozygosity. Finally, we calculated linkage disequilibrium, or 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. Since inversions are expected to only suppress recombination in heterokaryotypes, recombination in homokaryotypes should be unaffected.

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 assumed that the more

221 derived inversion arrangement would have lower heterozygosity given its relatively recent origin
222 compared to the ancestral state (Laayouni 2003; Twyford & Friedman 2015; Knief *et al.* 2016).
223 Notably, when examining our data, we found five additional regions with discreteness very close
224 to 0.9 (0.893 - 0.898) that displayed distinct three-cluster PCR patterns, and we included these
225 regions as candidates for putative inversions as well.

226 To investigate patterns of population structure at putative inversions, we calculated genotype
227 frequencies in each river reach for each putative inversion. Additionally, we conducted PCA
228 analyses using all SNPs that were successfully aligned to genomes, SNPs within the identified
229 inversions, and the remaining aligned SNPs after the SNPs in putative inversions were removed
230 to compare patterns of genetic structure inferred from datasets including and not including
231 putative inversions.

232 ***Identification and Analyses of Large Clusters of Adaptive Loci***

233 Adaptive loci can be found across many areas of the genome or can be concentrated (i.e.,
234 clustered) in only a few genomic regions. To determine whether the genomes of our species
235 contained clustered architectures of adaptative loci, we investigated the distribution of GEA
236 outliers and islands of differentiation identified by HMM across the genome. We defined
237 chromosomes exhibiting clustered architecture of adaptation as chromosomes containing at least
238 3 GEA outliers or 20% of the total HMM islands within a given species. We then calculated the
239 following genetic metrics to characterize the genomic properties of clustered architectures that
240 we observed: Fstp, heterozygosity (H_O), absolute differentiation (D_{xy}), and linkage
241 disequilibrium (LD).

Fstp and H_O were calculated using the *basic.stats* function in *hierfstat* as described previously. Pairwise per-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 in the first population and p_2 is the frequency of that allele in the second population (Irwin *et al.* 2016). Allele frequency was estimated using *makefreq* function (missing = "mean") in *adegenet*. Overall D_{xy} was calculated as the mean of all pairwise D_{xy} values. LD (r^2) was calculated using PLINK v1.9 for SNPs with MAF > 0.01. We included D_{xy}, an absolute measure of genetic differentiation, because defining adaptive genomic regions based solely on relative measures of differentiation, such as Fstp, may identify regions resulting from processes other than adaptation (Cruickshank & Hahn 2014).

For each chromosome, we used Wilcoxon tests ($\alpha = 0.05$) to test for significant differences in genetic metrics between SNPs within the islands and SNPs outside the islands (chromosomal background). We also visualized differences in these metrics with boxplots. To ensure the differences we observed in four genetic metrics were not due to island size, we randomly selected five windows outside of the islands as chromosomal background with window size (number of SNPs) set as the average size of all HMM islands found on the corresponding chromosomes. When calculating the average size of the HMM islands, we removed the HMM islands containing only one SNP to avoid downward biasing the window size of random windows in the chromosomal background. We also made sure the randomly selected windows spanned similar distance (± 10% bp) compared to the average of all HMM islands found on the same chromosomes.

Lastly, we conducted Gene Ontology (GO) enrichment tests to test for functional enrichment of genes in the HMM islands located within the six chromosomes displaying clustered architecture.

We extracted genes within 10 Kb of a SNP for all SNPs located within the islands for all six chromosomes, except for chromosome 9 in Emerald Shiner. Since all of the HMM islands on chromosome 9 in Emerald Shiner were clustered inside of the identified inversion and there was a relatively smaller number of aligned loci, we extracted genes within 20 Kb of a SNP for all SNPs located within the identified inversion on the chromosome 9 in Emerald Shiner instead. See Supplementary Materials for detailed methods about GO enrichment tests.

Results

Summary of Sequencing, GEA Outliers, 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 S3). Out of these polymorphic SNPs, 0.04 % to 0.3% were identified as GEA outliers, and 95.9 % - 99.4% were identified as neutral SNPs in each species (Table S4). 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 S4).

Neutral Genetic Differentiation

Patterns of population structure estimated from the neutral datasets spanned a large gradient of genetic differentiation across species (Figure 2, Table S5). 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 = 0.0302; 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, 0.0004, with pairwise F_{ST} values ranging from -0.0003 to 0.0016.

Results of the 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 grouped separately from other populations along with some individuals from Pool 26 and Open River. In Channel Catfish, individuals from the Open River were slightly separated from all the other reaches. Lastly, Gizzard Shad and Emerald Shiner showed no apparent population structure.

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 S3). Correspondingly, a total of 2 - 25 GEA outliers were aligned with alignment rate per species varying from 25.7% to 100% (Table S4).

306 Genome scan results revealed highly variable genomic landscapes of population differentiation
307 among the six species (Figure 2). In general, GEA outliers and HMM islands in species with
308 lower neutral differentiation were more tightly clustered and found on fewer chromosomes,
309 whereas those in species with higher neutral differentiation were spread out across the genome.
310 Bullhead Minnow (highest neutral population structure) displayed a high level of baseline
311 differentiation without obvious peaks of highly differentiated loci. We only detected 2 islands on
312 2 chromosomes and there were no GEA outliers located within the islands. In Bluegill, the
313 species with the second highest neutral population structure, we identified 83 islands that were
314 dispersed across nearly all chromosomes (22/24) with no chromosomes containing more than 8%
315 of islands. Additionally, 15 out of 21 aligned GEA outliers (71%) were located in 12 islands
316 across 9 chromosomes with only 3 islands having more than one GEA outliers (up to 2).
317 Freshwater Drum had an intermediate level of population differentiation and displayed a more
318 clustered architecture of genomics islands of differentiation compared to Bullhead Minnow and
319 Bluegill. In total, 14 islands were detected across 6 chromosomes and no GEA outliers were
320 found within islands. Of these islands, 3 islands (21%) were clustered on chromosome 7 and 7
321 islands (50%) were clustered on chromosome 17. Channel Catfish had a relatively low level of
322 differentiation and displayed highly clustered architectures of adaptation. We identified 15
323 islands across 10 chromosomes with 3 islands (20%) clustered on chromosome 13. Out of 25
324 aligned GEA outliers, 6 (24%) were located on an island on chromosome 20 and 4 (16%) were
325 located on an island on chromosome 28. Gizzard Shad had a similar level of neutral global Fstp
326 as Channel Catfish, but we did not detect any islands of high differentiation, possibly due to its
327 relatively low genome alignment rate (26.4%). Lastly, Emerald Shiner, the species with lowest
328 overall neutral population differentiation, displayed the strongest signal of clustered architecture

of local adaptation. In Emerald Shiner, 15 islands were detected across 4 chromosomes, of which, 11 (73%) were clustered on chromosome 9. A total of 11 out of 13 aligned GEA outliers (85%) were found 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 S6). 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 inconsistent with the theory that inversions should only suppress recombination in heterokaryotypes, so this region was excluded. 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 S6). The two homokaryotypes presented significant differences in heterozygosity for all three putative inversions (Figure 4B, S5B and S6B) and 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

351 0.28 for the inversion on chromosome 19 (cluster 0; Figure S5C). Moreover, GEA outliers and
352 HMM islands were consistently associated with the putative inversion on chromosome 9 in
353 Emerald Shiner (Figure 2). However, no GEA outliers or HMM islands were found within the
354 inversions on the chromosome 6 and 19.

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

361 ***Genomic Properties of Large Clusters of Adaptive Loci***

362 The following chromosomes exhibited highly clustered architecture with at least 3 GEA outliers
363 or 20% HMM islands within a given species: (1) chromosome 7 and 17 in Freshwater Drum; (2)
364 chromosome 13, 20, and 28 in Channel Catfish; and (3) chromosome 9 in Emerald Shiner
365 (Figure 2). The HMM islands on most of these chromosomes were characterized by high
366 population differentiation and co-located with several GEA outliers strongly associated with
367 environmental variables except for Freshwater Drum, where no GEA outliers were found within
368 clusters of HMM islands. In all six chromosomes with clustered architectures we found, as
369 expected, significantly higher Fstp values within the HMM islands (Figure 6). Comparisons of
370 H_O and D_{xy} between HMM islands and chromosomal background showed three different patterns
371 among six chromosomes: (1) islands on chromosome 7 and 17 in Freshwater Drum and
372 chromosome 9 in Emerald Shiner had similar H_O and D_{xy}; (2) islands on chromosome 13 in

Channel Catfish had significantly higher values of H_O and D_{xy}; (3) islands on chromosome 20 and 28 in Channel Catfish had significantly lower values of H_O and D_{xy} (Figure 6). We also found significantly elevated LD within the islands in all chromosomes except for chromosome 13 in Channel Catfish (Figure 6). Taken together, these results indicate that the HMM islands on the six chromosomes with clustered architectures have higher relative divergence than their chromosomal backgrounds; the islands on chromosome 13 in Channel Catfish also demonstrated higher absolute divergence, though without elevated LD.

A total of 9, 12, and 2 GO terms were significantly enriched ($p < 0.05$) in the HMM islands on chromosome 17 in Freshwater Drum, the island on chromosome 28 in Channel Catfish, and the inversion on chromosome 9 in Emerald Shiner, respectively (Table S7). Enriched GO terms included regulation of cellular component size, cell communication, and regulation of ion transmembrane transport. There were no annotated genes found within the HMM island(s) on chromosome 7 in Freshwater Drum, chromosome 13 and 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

394 studies on similar fish species have generally corroborated the patterns we observed, with nest
395 spawning species such as smallmouth bass (*Micropterus dolomieu*) exhibiting high levels of
396 genetic structure in open systems compared to broadcast spawning species such as walleye
397 (*Sander vitreus)* (Ruzich *et al.* 2019; Euclide *et al.* 2020; 2021)

398 An exception to the pattern described above was Channel Catfish, as they are nest spawners but
399 displayed relatively low levels of differentiation. It is possible that the highly migratory nature of
400 this species mixed with potentially low spawning fidelity (Pellett *et al.* 1998) could explain the
401 low to intermediate levels of population differentiation we observed. Freshwater Drum also
402 deviated from the expected patterns of population structure based on life history, as they are
403 migratory broadcast spawners but displayed an intermediate level of population structure, with
404 individuals from La Grange along with some individuals from southern populations in Pool 26
405 and Open River forming a distinct group. One possible explanation for this pattern is limited
406 movement of Freshwater Drum between the Illinois River, where La Grange is located, and the
407 mainstem Mississippi River. Unfortunately, movement data for this species are generally lacking,
408 making it difficult to corroborate this hypothesis without additional research.

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

410 Most of the GEA outliers that we found were associated with environmental PC1, which had the
411 highest loadings for temperature and turbidity. It is likely that these GEA outliers reflect adaptive
412 divergence driven by the large latitudinal gradient that we sampled. Our study system spans two
413 major Köppen climate zones, with pools 4, 8, and 13 in a humid continental climate
414 characterized by warm summers and very cold winters (below 0 °C) , and Pool 26, Open River,
415 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 strong adaptive divergence driven by temperature that are similar to those found in marine systems, highlighting the fact that populations of continuously distributed riverine species may display the potential for local adaptation across their range.

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 an underlying geomorphological condition correlated with agricultural inputs or to

variation along a lotic-lentic gradient, to which Freshwater Drum are known to respond (Rypel *et al*. 2006; Jacquemin *et al*. 2015).

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 Drum) displayed more clustered architecture of adaptation than low gene flow species (Bullhead Minnow and Bluegill). Our results are somewhat similar to a recent study which examined adaptive divergence of four flatfish species across a strong salinity gradient in the Baltic Sea (Le Moan *et al.* 2019). Specifically, Le Moan *et al.* (2019) found more evidence of clustered architectures of adaption in species displaying low genetic differentiation compared to those displaying higher differentiation. However, Le Moan *et al.* (2019) sampled a much smaller gradient of genetic differentiation (overall F_{ST} from 0.005 – 0.02) than our study and examining the effects of gene flow on landscapes of adaptive differentiation was not a central goal of their study.

Though our finding that clustered genomic architectures of adaptation (i.e., genomic islands of divergence) increase with gene flow is in line with theoretical expectations and the results from

Le Moan *et al.* (2019), this finding is inconsistent with other studies positing that islands of divergence are the result of variation in intrinsic recombination rate rather than the combination of gene flow and selection (Roesti *et al.* 2012; Renaut *et al.* 2019). In fact, there is considerable debate over the mechanisms that lead to islands of divergence, with past research suggesting that these islands can be caused by variation in recombination rates (Roesti *et al.* 2012; Renaut *et al.* 2019), linked selection (Cruickshank & Hahn 2014; Burri *et al.* 2015), divergence hitchhiking (Via 2012), genomic rearrangements including chromosomal inversions (Rogers *et al.* 2013; Yeaman 2013), and elevated linkage preserving locally adapted alleles (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 other islands are less clear.

To investigate the genomic mechanisms that created the islands on the remaining five chromosomes exhibiting clustered architecture, we calculated the following four metrics: LD, F_{ST}, H_O, and D_{xy}, and compared these metrics between islands and chromosomal background. While islands on all five chromosomes displayed elevated F_{ST} as expected, we did observe differences in the remaining three metrics among the chromosomes. Islands on all but one chromosome displayed elevated LD; H_O was elevated or similar to neighboring neutral regions in islands on three out of five chromosomes, and D_{xy} was elevated or similar to neutral regions in islands 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 the two chromosomes with reduced diversity (islands on Channel Catfish chromosomes 20 and 28) may have been created by linked selection (Cruickshank & Hahn 2014; Burri *et al.* 2015), while the islands on Channel Catfish chromosome 13 and

Freshwater Drum chromosomes 7 and 17 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 islands created by divergent selection should not (Cruickshank & Hahn 2014; Irwin *et al.* 2016). Applying this method to our data provided evidence that islands on three of the chromosomes in our study were created by divergent selection and are likely involved in adaptive divergence with gene flow whereas the islands on the other two chromosomes were likely a result of ancient linked selection that acted to reduce diversity in particular genomic regions but is not influencing contemporary adaptive divergence.

Interestingly, we observed variation in the mechanisms putatively responsible for creating islands both within and among species, as islands on one of the three chromosomes in Channel Catfish were likely created by divergent selection while islands on the other two chromosomes were likely the result of linked selection. This result indicates that, while many studies have tried to generalize the proposed mechanisms responsible for creating islands of divergence, these mechanisms vary across and even within species (Ottenburghs *et al.* 2020; Liu *et al.* 2020; Wilder *et al.* 2020). Thus, it is vital to examine both absolute and relative measures of

differentiation as well as diversity in newly discovered islands of divergence to help clarify the mechanisms responsible for creating these islands.

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

Our results and those of previous empirical and theoretical studies suggest that divergent selection can result in clusters of adaptive loci through mechanisms such as divergence 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

528 involved in adaptive divergence have been documented in many fishes including Atlantic cod
529 (*Gadus morhua*) (Kirubakaran *et al.* 2016), lingcod (*Ophiodon elongatus*) (Longo *et al.* 2020),
530 rainbow trout (*Oncorhynchus mykiss*) (Arostegui *et al.* 2019; Pearse *et al.* 2019), Pacific herring
531 (*Clupea pallasii*) (Petrou *et al.* 2021), Atlantic silverside (*Menidia menidia*) (Wilder *et al.* 2020),
532 and European plaice (*Pleuronectes platessa*) (Le Moan *et al.* 2019). However, all of these studies
533 were conducted on marine fish or salmonids, making our study the first to provide evidence of a
534 putative adaptive inversion in a non-salmon freshwater fish. It is likely that the lack of previous
535 evidence for adaptive inversions in freshwater fish is due to the generally higher genetic structure
536 observed in these species, making inversions less necessary for adaptation. However, our study
537 illustrates that inversions are likely a larger component of adaptive divergence in freshwater fish
538 than previously assumed, highlighting the importance of future studies aimed at characterizing
539 them in additional species.

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

551 would be useful for assessing the functional role of these inversions in the process of adaptive
552 divergence.

553 **Conclusions**

554 Our data from six riverine fish species in UMRS displaying a large gradient of life history
555 strategies suggest that higher gene flow leads to increasingly concentrated genomic architectures
556 of adaptation. Additionally, our results provide evidence that the mechanisms that create islands
557 of divergence can be highly variable across and within species, with both ancient linked selection
558 and more contemporary divergent selection playing important roles in creating genomic islands
559 of differentiation. Additionally, our study provides further evidence that chromosomal inversions
560 are important for facilitating adaptive divergence in continuously distributed species with
561 extremely high gene flow and also sheds light on the documented importance of inversions in
562 freshwater fish. Taken together, our findings represent a significant contribution towards
563 understanding the evolutionary processes that influence the genomic landscape of adaptation in
564 non-model organisms. However, our study used RADseq, which does not assess the full suite of
565 adaptive divergence across the genome. Future studies should focus on whole genome
566 resequencing to better understand variation within genomic islands of divergence and to assess
567 the functional role of these islands in promoting adaptive divergence.

568 **Acknowledgements**

569 This study was funded by the U.S. Army Corps of Engineers Upper Mississippi River
570 Restoration Program (grant number 96514790661277) and supported by the Turing High
571 Performance Computing cluster at Old Dominion University. We thank U.S. Geological Survey
572 Upper Midwest Environmental Sciences Center, Iowa Department of Natural Resources,

Minnesota Department of Natural Resources, Wisconsin Department of Natural Resources, 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

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.

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.

607 Euclide PT, MacDougall T, Robinson JM *et al.* (2021) Mixed-stock analysis using Rapture
608 genotyping to evaluate stock-specific exploitation of a walleye population despite weak
609 genetic structure. *Evolutionary Applications*.
610 Euclide PT, Ruzich J, Hansen SP, Zorn TG, Larson WA (2020) Genetic structure of Smallmouth
611 Bass in the Lake Michigan and Upper Mississippi River drainages relates to habitat, distance,
612 and drainage boundaries. *Transactions of the American Fisheries Society*, 383–397.
613 Faria R, Johannesson K, Butlin RK, Westram AM (2019) Evolving inversions. *Trends in*
614 *Ecology & Evolution*, **34**, 239–248.
615 Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting
616 multilocus adaptation with multivariate genotype-environment associations. *Molecular*
617 *Ecology*, **27**, 2215–2233.
618 Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics.
619 *Molecular Ecology Notes*, **5**, 184–186.
620 Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the genomic basis of local adaptation:
621 pitfalls, practical solutions, and future directions. *The American Naturalist*, **188**, 379–397.
622 Hofer T, Foll M, Excoffier L (2012) Evolutionary forces shaping genomic islands of population
623 differentiation in humans. *BMC Genomics*, **13**, 1–13.
624 Hoffmann AA, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From
625 Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of*
626 *Ecology, Evolution, and Systematics*, **39**, 21–42.
627 Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications
628 for landscape genetics. *Landscape Ecology*, **21**, 797–807.
629 Hondorp DW, Breitburg DL, Davias LA (2010) Eutrophication and fisheries: separating the
630 effects of nitrogen loads and hypoxia on the pelagic□to□demersal ratio and other measures
631 of landings composition. *Marine and Coastal Fisheries*, **2**, 339–361.
632 Huang K, Rieseberg LH (2020) Frequency, origins, and evolutionary role of chromosomal
633 inversions in plants. *Frontiers in Plant Science*, **11**, 1–13.
634 Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH (2020) Multiple chromosomal
635 inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular*
636 *Ecology*, **29**, 2535–2549.
637 Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL (2016) Recurrent selection explains
638 parallel evolution of genomic regions of high relative but low absolute differentiation in a
639 ring species. *Molecular Ecology*, **25**, 4488–4507.
640 Jacobson PC, Hansen GJA, Bethke BJ, Cross TK (2017) Disentangling the effects of a century of
641 eutrophication and climate warming on freshwater lake fish assemblages. *PLoS ONE*, **12**,
642 e0182667.
643 Jacquemin SJ, Doll JC, Pyron M, Allen M, Owen DAS (2015) Effects of flow regime on growth
644 rate in freshwater drum, *Aplodinotus grunniens*. *Environmental Biology of Fishes*, 993–1003.
645 Johannesson K, Le Moan A, Perini S, André C (2020) A darwinian laboratory of multiple
646 contact zones. *Trends in Ecology & Evolution*, 1–16.
647 Jombart T (2008) *adegenet*: an R package for the multivariate analysis of genetic markers.
648 *Bioinformatics*, **24**, 1403–1405.
649 Kirubakaran TG, Grove H, Kent MP *et al.* (2016) Two adjacent inversions maintain genomic
650 differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular*
651 *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, 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.

Liu X, Glemin S, Karrenberg S (2020) Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (*Silene*). *Molecular Ecology*, **11**, 545–15.

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, 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.

Ottenburghs J, Honka J, Müskens GJDM, Ellegren H (2020) Recent introgression between Taiga Bean Goose and Tundra Bean Goose results in a largely homogeneous landscape of genetic differentiation. *Heredity*, **125**, 73–84.

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

Pearse DE, Barson NJ, Nome T *et al.* (2019) Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, 1–30.

Pellett TD, Van Dyck GJ, Adams JV (1998) Seasonal migration and homing of Channel Catfish in the lower Wisconsin River, Wisconsin. *North American Journal of Fisheries Management*, 85–95.

Petrou EL, Fuentes-Pardo AP, Rogers LA *et al.* (2021) Functional genetic diversity in an exploited marine species and its relevance to fisheries management. *Proceedings of the Royal Society B: Biological Sciences*, 20202398.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.

Renaut S, Grassa CJ, Yeaman S *et al.* (2019) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 1–8.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Roda F, Walter GM, Nipper R, Ortiz-Barrientos D (2017) Genomic clustering of adaptive loci during parallel evolution of an Australian wildflower. *Molecular Ecology*, **26**, 3687–3699.

Roesti M (2018) Varied Genomic Responses to Maladaptive Gene Flow and Their Evidence. *Genes*, **9**, 298–16.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Rogers SM, Mee JA, Bowles E (2013) The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology*, **59**, 53–71.

Ruzich J, Turnquist K, Nye N, Rowe D, Larson WA (2019) Isolation by a hydroelectric dam induces minimal impacts on genetic diversity and population structure in six fish species. *Conservation Genetics*, **20**, 1421–1436.

Rypel AL, Bayne DR, Mitchell JB (2006) Growth of Freshwater Drum from lotic and lentic habitats in Alabama. *Transactions of the American Fisheries Society*, 987–997.

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.

Stanley RRE, DiBacco C, Ben Lowen *et al.* (2018) A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, **4**, eaaq0929.

Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the 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.

743 Thompson NF, ANDERSON EC, Clemento AJ *et al.* (2020) A complex phenotype in salmon
744 controlled by a simple change in migratory timing. *Science*, 609–613.
745 Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*,
746 **25**, 2144–2164.
747 Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower *Mimulus guttatus* is
748 maintained by a chromosomal inversion. *Evolution*, **69**, 1476–1486.
749 Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological
750 speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological*
751 *Sciences*, **367**, 451–460.
752 Wellenreuther M, Bernatchez L (2018) Eco-evolutionary genomics of chromosomal inversions.
753 *Trends in Ecology & Evolution*, **33**, 427–440.
754 Wilder AP, Palumbi SR, Conover DO, Therkildsen NO (2020) Footprints of local adaptation
755 span hundreds of linked genes in the Atlantic silverside genome. *Evolution Letters*, **4**, 430–
756 443.
757 Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci.
758 *Proceedings of the National Academy of Sciences*, **110**, E1743–E1751.
759 Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-
760 selection balance. *Evolution*, **65**, 1897–1911.

761 **Data Accessibility**

762 Upon acceptance, demultiplexed RAD sequencing data used in this study will be released in the
763 NCBI with BioProject ID, PRJNA674918. Bioinformatic scripts along intermediate datasets
764 (post-filtering *vcf*, neutral *vcf*, adaptive *vcf*) supporting this article will be available as a Github
765 repository.

766 **Author Contributions**

767 YS, KLB, AB and WAL conceived of the study, designed the study, and coordinated the study.
768 YS and WD carried out the molecular lab work. YS and WAL conducted data analyses and
769 drafted the manuscript; GJM helped interpreted the results regarding putative inversions. MC
770 supervised the project. All authors commented the manuscript and gave final approval for
771 publication.

772 **Conflict of Interest**

773 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.

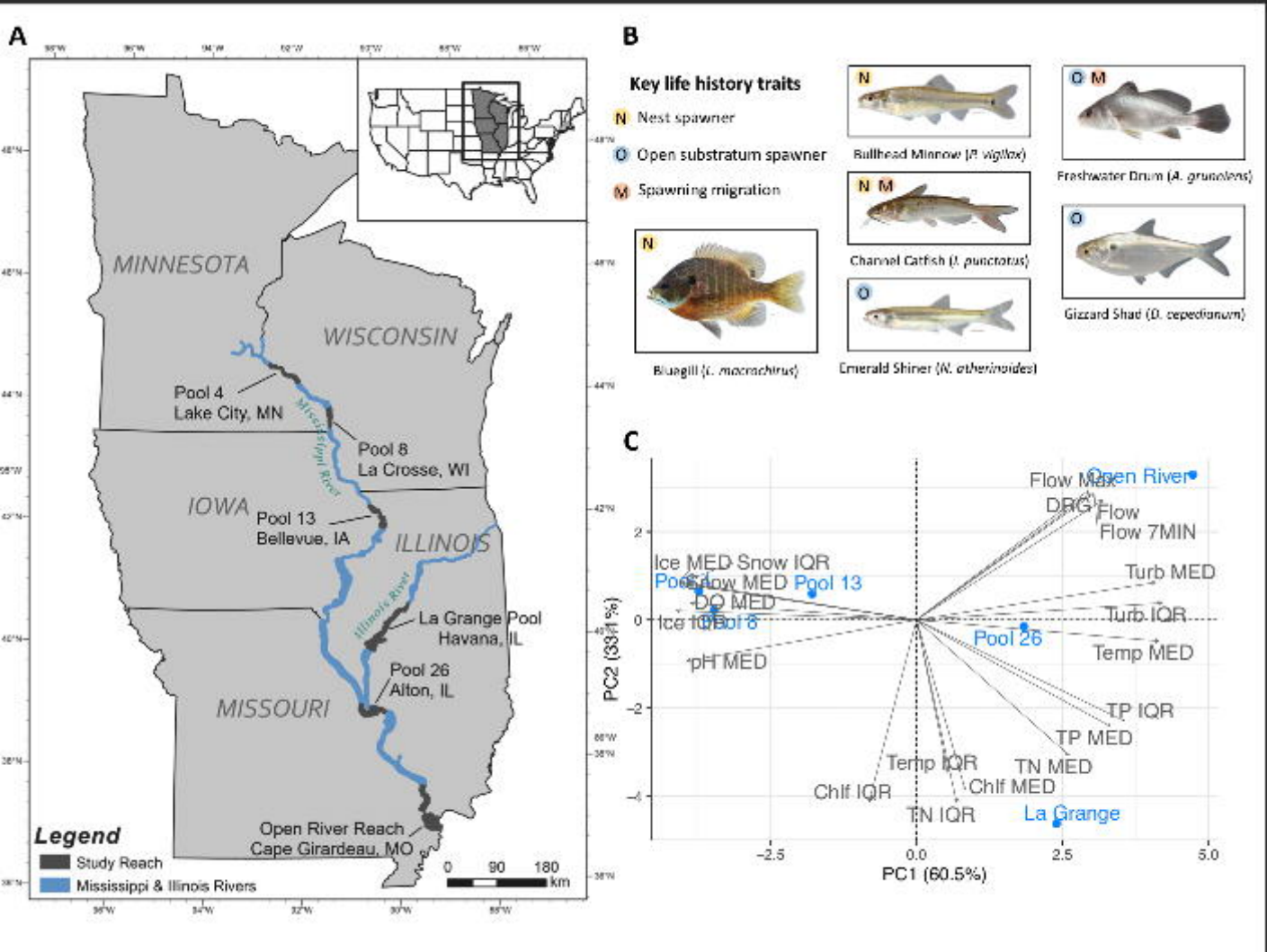
Figure 2. Manhattan plots depicting the genomic landscape of differentiation (Fstp, corrected F_{ST}) across the genomes for the six study species. Species are ordered based on neutral population differentiation, with neutral global Fstp values labeled next to the species name. At the top of each plot, genomic islands of differentiation identified using HMM after filtering are in blue, and islands located within the chromosomes showing clustered architecture are in purple. GEA outliers found within islands are in red, whereas those found outside of islands of genomic divergence are in orange. Reference genomes and alignment summary can be found in Table S3.

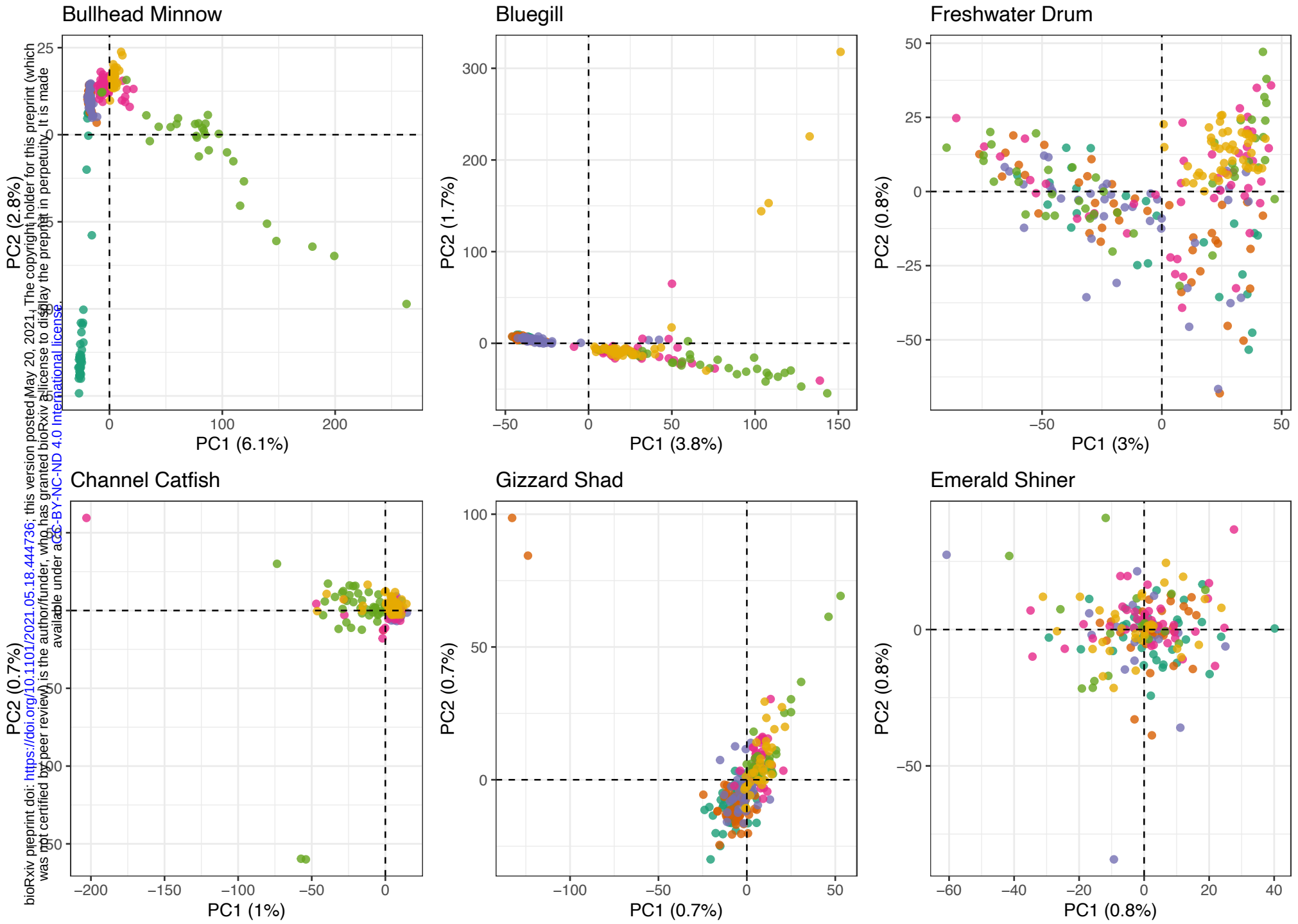
Figure 3. Principal component analyses using neutral SNPs only for the six study species. The percentage of variance explained by each principal component (PC) is labeled on the *x*- and *y*-axes.

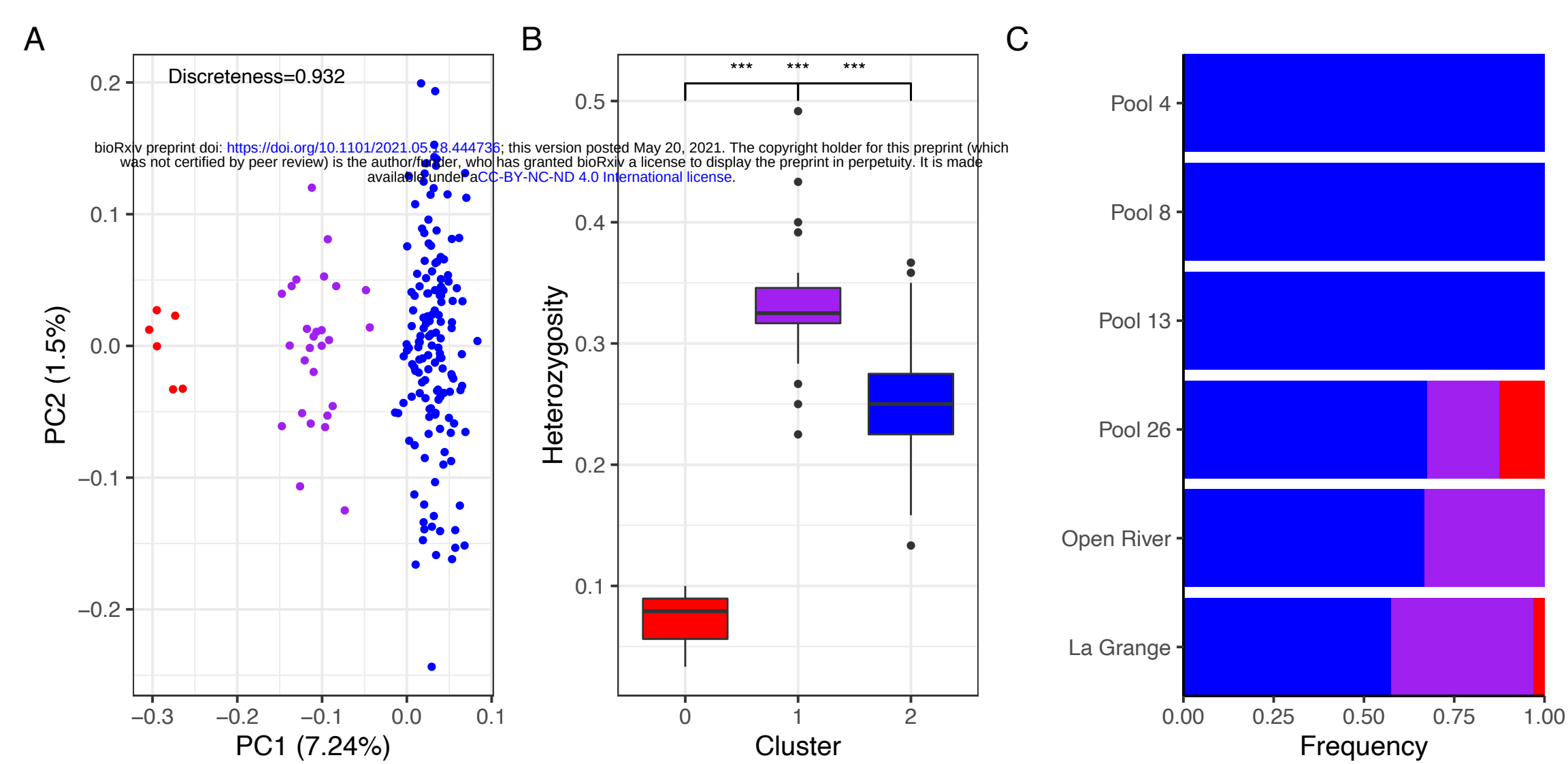
Figure 4. Characterization of putative inversion on chromosome 9 in Emerald Shiner. (A) PCA based on SNPs within the putative inversion region. Three clusters identified using k-means clustering correspond to two homozygote groups (blue and red) and a heterozygote group (purple). The discreteness of 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).

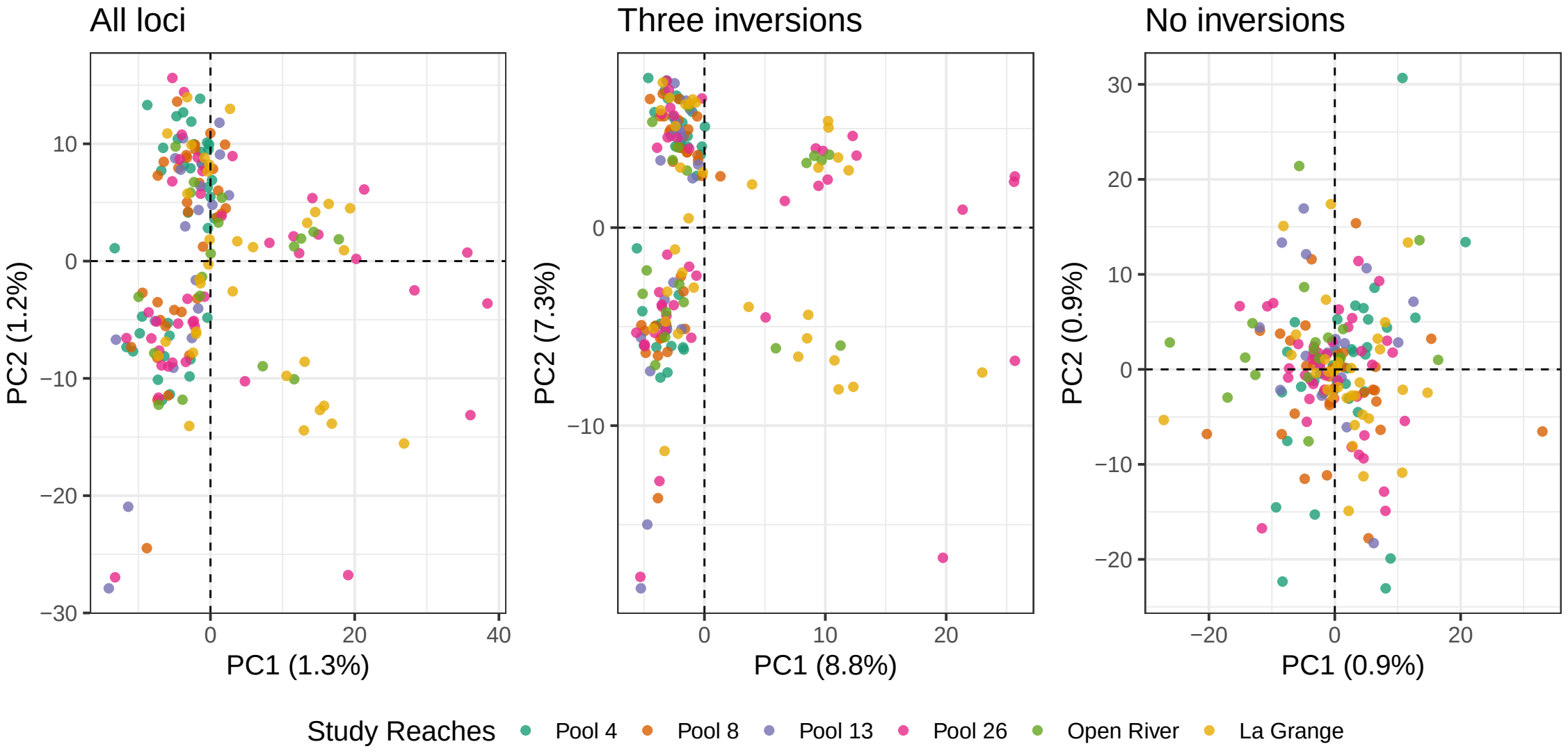
Figure 5. Principal component analyses for Emerald Shiner using different sets of loci: (A) All aligned SNPs (3,348 SNPs); (B) Putative inversions on chromosome 6, 9 and 19 (228 SNPs); (C) After the removal of three putative inversions (3,120 SNPs). The percentage of variance explained by each principal 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 SNPs within the HMM islands (Island, red) and five combined random windows (Background, gray) on the corresponding chromosomes for chromosomes with clustered architecture, including on chromosome 7 and 17 in Freshwater Drum, chromosome 20 and 28 in Channel Catfish, and chromosome 9 in Emerald Shiner. The average distance between pairs of SNPs within islands and random windows were labelled below the bloxplots of LD values (column 4). Significance was assessed between islands and random windows using Wilcoxon tests with alpha level of 0.05. Note: *** = 0.001, ** = 0.01, * = 0.05, NS = not significant. Only HMM islands with more than one SNP were included.

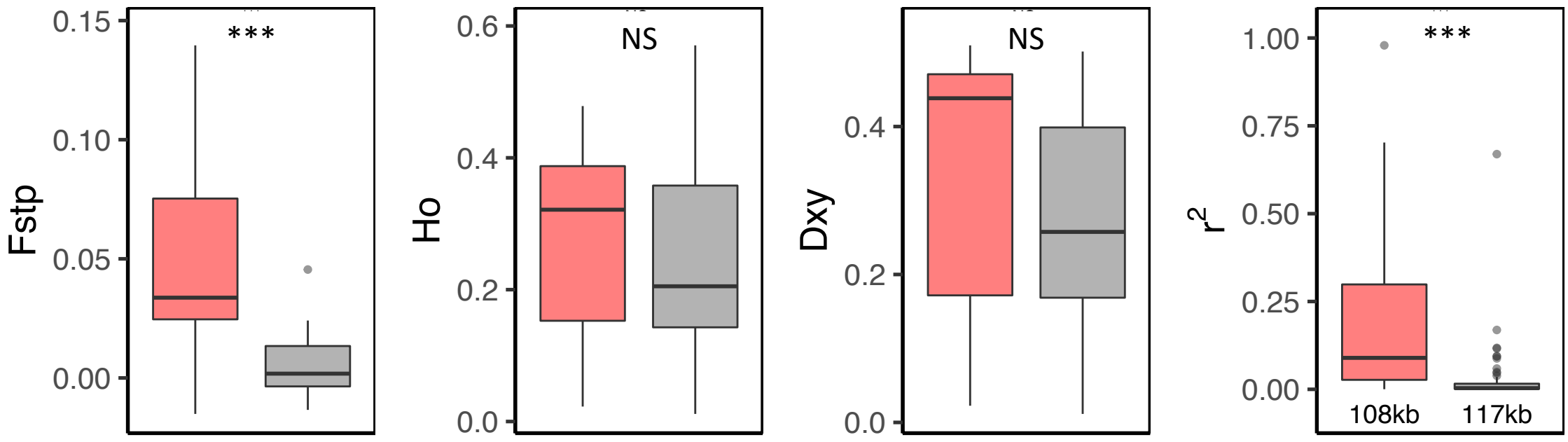

A
MINNESOTA
WISCONSIN
IOWA
ILLINOIS
MISSOURI
Pool 4
Lake City, MN
Pool 8
La Crosse, WI
Pool 13
Bellevue, IA
La Grange Pool
Havana, IL
Pool 26
Alton, IL
Open River Reach
Cape Girardeau, MO
Mississippi River
Illinois River
Legend
Study Reach
Mississippi & Illinois Rivers
0 90 180
km
B
Key life history traits
N Nest spawner
O Open substratum spawner
M Spawning migration
Bullhead Minnow (P. vigilax)
Freshwater Drum (A. grunniens)
Channel Catfish (I. punctatus)
Gizzard Shad (D. cepedianum)
Bluegill (L. macrochirus)
Emerald Shiner (N. atherinoides)
C
Open River
Flow Max
DRG
Flow
Flow 7MIN
Turb MED
Turb IQR
Temp MED
TP IQR
TP MED
TN MED
Chlf MED
La Grange
TN IQR
Chlf IQR
Temp IQR
pH MED
Ice MED
Snow IQR
Snow MED
Pool 13
DO MED
Pool 8
Pool 26
PC1 (60.5%)
PC2 (33.1%)



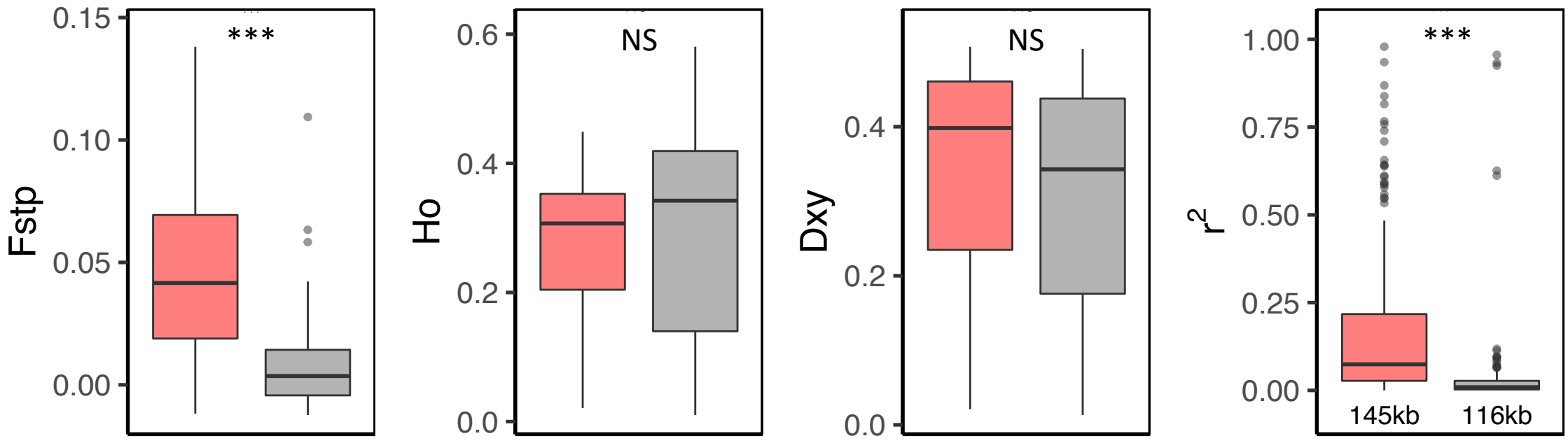



All loci
Three inversions
No inversions
PC2 (1.2%)
PC1 (1.3%)
PC2 (7.3%)
PC1 (8.8%)
PC2 (0.9%)
PC1 (0.9%)
Study Reaches
Pool 4
Pool 8
Pool 13
Pool 26
Open River
La Grange

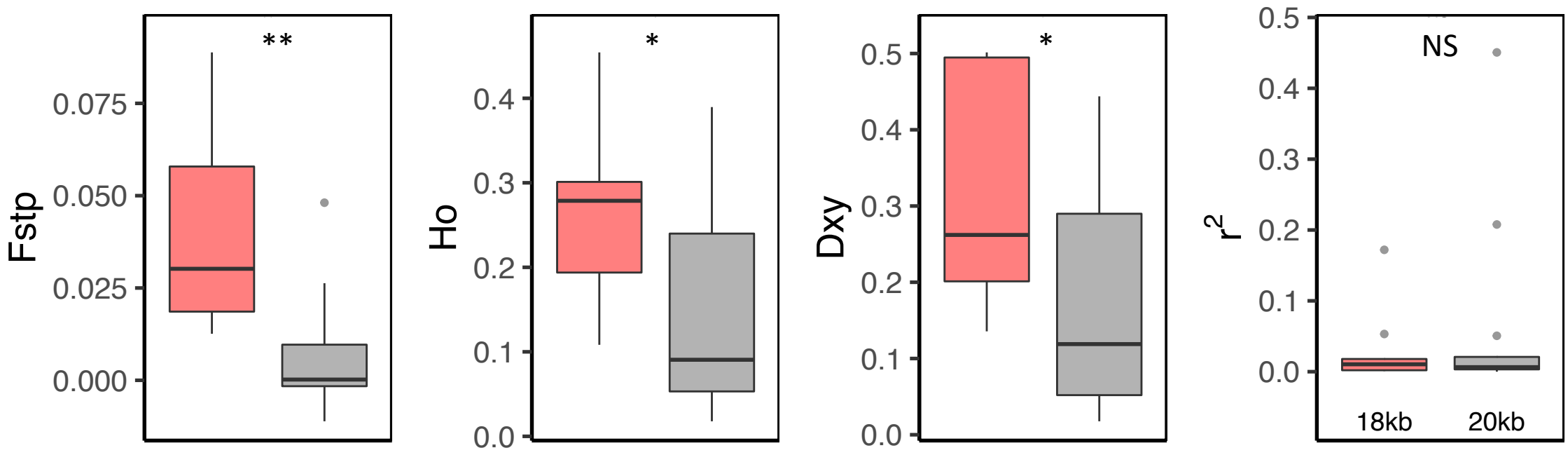
Freshwater Drum Chromosome 7



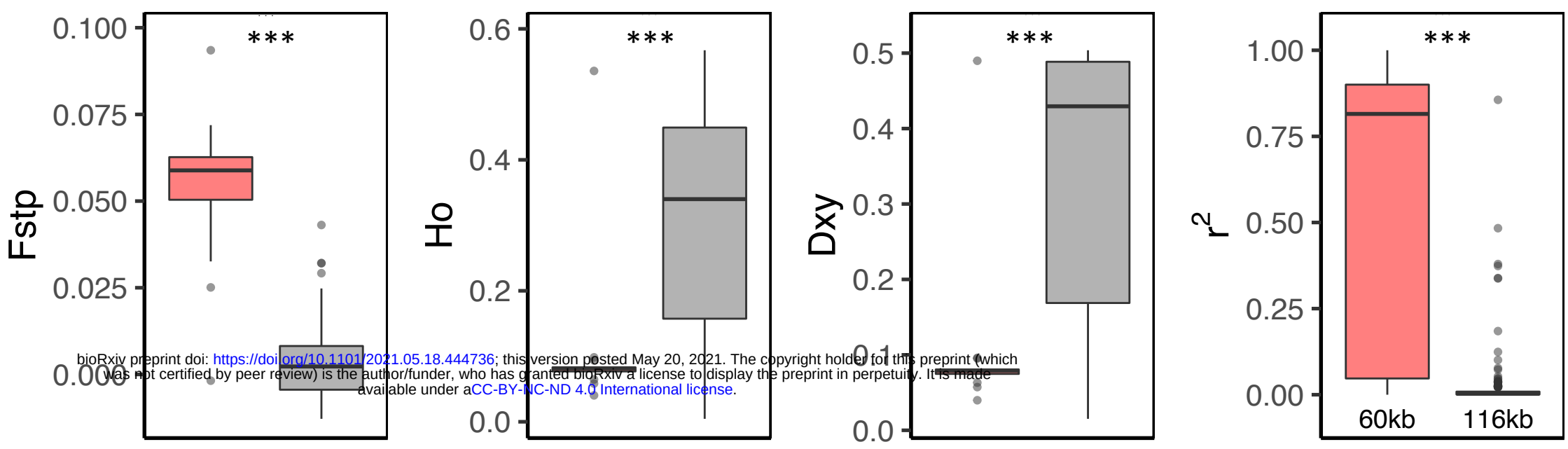
Freshwater Drum Chromosome 17



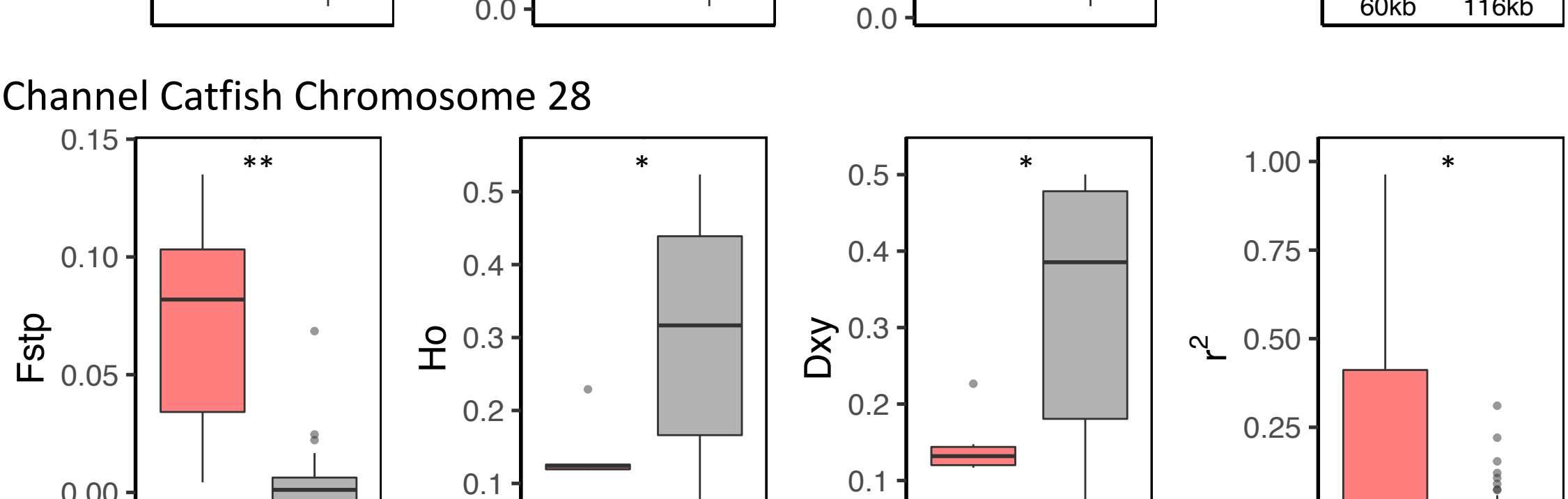
Channel Catfish Chromosome 13



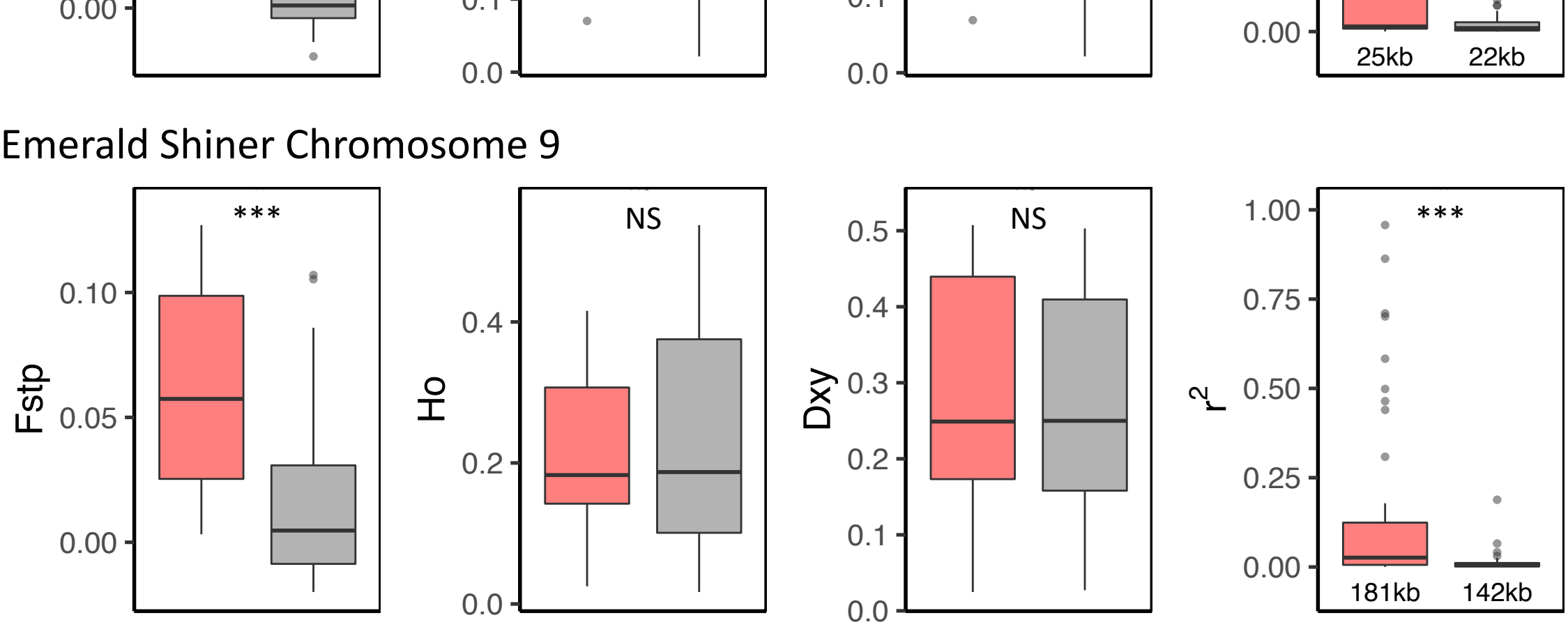
Channel Catfish Chromosome 20



Channel Catfish Chromosome 28



Emerald Shiner Chromosome 9



Island Background