Final Report

Genetic Diversity of the Neosho Madtom (*Noturus placidus*)

Submitted to:
Jordan Hofmeier
Kansas Department of Wildlife and Parks
512 SE 25th Ave.
Pratt, KS 67124

Prepared by:
Marlis R. Douglas
Zachery D. Zbinden
Tyler K. Chafin
Jeremy S. Tiemann
David R. Edds
Bradley T. Martin
Michael E. Douglas

Arkansas Conservation & Molecular Ecology Laboratory
University of Arkansas
Biological Sciences
850 W Dickson St
Fayetteville, AR 72701

Respectfully submitted:
15 November 2023

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Overview

The Neosho Madtom, *Noturus placidus,* is a small-range, endemic catfish that persists as isolated populations in the heavily impounded Neosho River system. But the species has experienced declines and was federally listed as threatened in 1990, primarily due to range reduction caused by anthropogenic habitat alterations. Therefore, to promote proactive management and understand Neosho Madtom population boundaries, connectivity, viability, and adaptation, we quantified and analyzed patterns of genetic diversity from Neosho Madtom collected throughout the Neosho River system using a population genomic approach.

Goals

This study aimed to identify boundaries, connectivity, viability, and adaptation of Neosho Madtom populations via genetic characteristics surveyed across the Neosho River system.

Objectives

We addressed six objectives to inform conservation planning:

1. We conducted comprehensive spatial sampling by collecting Neosho Madtom tissues (N=192) from localities (N=14) throughout the Neosho River system (Neosho, Cotttonwood, and Spring rivers).

2. We quantified genetic diversity within and among populations by genotyping thousands of genetic markers (loci; N=2,725) for each sample.

3. We determined distinct genetic population boundaries, quantified population connectivity (gene flow), and delineated potential management units by determining spatial genetic structure.

4. We identified environmental features that affect Neosho Madtom dispersal by correlating spatial genetic divergence with riverscape characteristics between sampling locations.

5. We determined demographic status, population viability, and trends by deriving standard population genetic parameters (e.g., effective population size, N_e).

6. We tested for local adaptation by identifying loci under selection along environmental gradients and provided validation of these candidate markers by elucidating the biological processes with which they are associated.

Approach

We conducted non-lethal tissue sampling in August 2021 across 19 sites within the Neosho River system (Table 1; Figure 1). The target sample size for Neosho Madtom was 16 individuals per site, with individuals documented via photography as digital vouchers and represented by a fin clip tissue sample for DNA extraction and analysis. We collected 249 fin clips representing 192

Neosho Madtom from 14 sites; another 57 fin clips represented other *Noturus* species from 15 sites, including 36 Stonecat (*N. flavus*), 19 Freckled (*N. nocturnus*), and 2 Slender madtoms (*N. exilis*).

DNA was extracted and sequenced, and genetic variation was quantified across 2,725 genetic markers called SNPs (= Single Nucleotide Polymorphisms). This method encapsulates variation across the genome by consistently sequencing the same loci for each individual, making it more efficient and cost-effective than whole-genome sequencing and allowing more individuals and sites to be included in analyses. These genome-wide markers were analyzed to determine genetic population boundaries, quantify genetic variation within and among populations, and investigate whether environmental and contemporary changes, such as dams (Figure 2), have influenced genetic divergence within the Neosho Madtom.

Key Findings

Six Genetic Populations

- Neosho Madtom persists as six genetically distinct populations that could be interpreted as management units (i.e., 'stocks'; Figures 3, 4, 5).
- Populations in the Upper Neosho River (NRU) and Spring River (SPR) are genetically most distinct (Tables 2, 3; Figure 4).

Riverscape Features Impact Connectivity

- All dams, including low head dams, represent barriers to the dispersal of Neosho Madtom, reducing population connectivity (Figures 6, 7).
- No recent dispersal (i.e., first-generation migrants) was observed in any population, but genetic data reflected a signal of historic downstream-biased dispersal.

Standard Population Genetic Parameters

- Genetic patterns are consistent with the species' life history (e.g., low dispersal, short life spans, small populations; Tables 4, 5).
- Short-term extinction risk is low for all populations (N_e>50), but long-term population viability is uncertain for half (3/6) of populations (N_e<500; Table 5).
- All populations show signals consistent with recent bottlenecks (Table 5).
- Each population harbors unique alleles not shared among populations (Table 4), underscoring unique diversity and low dispersal among populations.

Local Adaptation

- Populations of Neosho Madtom appear to be locally adapted along environmental gradients (Figures 9, 10).
- Local adaptation occurred along the upstream to downstream (i.e., 'longitudinal') gradient and correlates with hydrologic/physiographic, landcover, geology, climate, and anthropogenic factors (Figure 10)
- Genetic markers showing signals of local adaptation are related to biological processes involving development, sensory input, stress response, and metabolism (Table 6; Figure 11)

- Patterns also indicate that different selection pressures due to environmental variation across the Neosho River system may be responsible for driving adaptive differences between populations (e.g., populations above and below John Redmond Reservoir; Figure 12).

Conclusion

The presence of six genetically discrete populations, each harboring unique genetic diversity and distinct signals of local adaptation, should be considered when planning management and conservation actions. Localized genetic diversity and population boundaries reflect both natural processes and recent habitat alterations (i.e., dams and flow modifications). Impoundments limit population connectivity and will be detrimental to the long-term viability of Neosho Madtom across its range. Translocations and supplementations may be advisable to ensure the species' long-term persistence, but careful consideration of local genetic diversity, genetic structure, genetic viability, and local adaptation is advised.

I. INTRODUCTION

Neosho Madtom

The Neosho Madtom (NMT), *Noturus placidus* Taylor, 1969 is a diminutive catfish (*Ictaluridae*) species (commonly <75 mm TL) endemic to the Neosho River system in southeastern Kansas, southwestern Missouri, and northeastern Oklahoma. Although typically short-lived (1-2 years), often reproducing within a single season, propagated individuals have exhibited a five- to eight-year lifespan within a laboratory setting (Wildhaber, 2011).

Today, Neosho Madtom occurs in three distinct regions separated by reservoirs (Figure 1): (i) the Neosho and Cottonwood rivers upstream of John Redmond Reservoir in Kansas, (ii) the Neosho River downstream of John Redmond Reservoir in Kansas and Oklahoma, and (iii) the Spring River in Missouri, Kansas, and Oklahoma (USFWS, 1991).

Historically, the range of NMT encompassed the rivers of the Neosho system and extended to the Illinois River in Oklahoma, where the species has since been extirpated. Presently, the species inhabits roughly two-thirds of its original range, a reduction primarily attributable to habitat loss and fragmentation induced by impoundments, such as Grand Lake o' the Cherokees and Lake Tenkiller (USFWS, 2013). The U.S. Fish and Wildlife Service (USFWS) federally listed NMT as a threatened species in 1990, with a recovery plan ratified in the subsequent year (USFWS, 1991). Impoundments are one of the leading causes of anthropogenic habitat alterations in the Neosho River system (Tiemann et al., 2004).

The species prefers relatively shallow riffles and bars with loose gravel situated in medium to moderately large rivers with perennial flow (Wildhaber, 2011). Such habitat preferences largely confine NMT to the Neosho River system's main stem and larger tributaries (Wildhaber et al., 2000). Rarely is it found in smaller tributary streams (stream order <3), and on such occasions, it tends to be located proximal to the main stem (USFWS, 2013). The population density of NMT is notably higher in the Cottonwood and Neosho rivers upstream of John Redmond Reservoir, with a decline observed downstream and in the Spring River (Wildhaber, 2011).

A species' competitive abilities generally trade off with its dispersal abilities (Pellissier, 2015), and NMT seems to 'compete in place' rather than 'move in space.' Because NMT is small-bodied and prefers stable, specialized habitats (shallow gravel with consistent flow), we expect its evolution to favor competition over dispersal (Pellissier, 2015). Indeed, NMT showcases robust competitive traits with a low propensity for dispersion. Competition with other benthic fish species, including Slender Madtom (*Noturus exils*), has not been identified as a limiting factor for NMT (Wildhaber et al., 1999; Tiemann et al., 2004). Moreover, in a one-year study of dispersal, only a single instance of even fine-scale inter-riffle movement across a series of riffles on the Cottonwood River was recorded (Fuselier & Edds, 1994).

Neosho River System

The Neosho River originates in the Flint Hills of east-central Kansas (Figure 1), flowing southeast over a distance of approximately 770 km before merging with the Arkansas River near Fort Gibson, Oklahoma—it is called the Grand River after its confluence with the Spring River

in Oklahoma. The system encompasses an area of 20,374 km^2, with 10,114 km^2 in Kansas (Branson, 1967).

Human activities have notably impacted the Neosho River system (Wildhaber, 2011), evidenced by heavy metal contamination, gravel mining, and numerous impoundments (Figure 2). These impoundments consist of four substantial state/federal dams constructed between 1940 and 1964 and 20 low head dams established between 1860 and 1995 (Fencl et al., 2015). Nine dams have existed for over a century (Table S1; Figure 3).

The Neosho River receives water from three major tributaries: the Cottonwood, Spring, and Elk rivers. At the point of confluence east of Emporia, Kansas, the Cottonwood River is larger than the Neosho River, exhibiting a Horton-Strahler Stream Order of 4 *vs.* 3 and an average discharge of 32.1 m^3/s compared to 13.3 m^3/s (Horton, 1945; Linke et al., 2019). These metrics suggest that the system's main stem, in a strict sense, encompasses the Cottonwood River rather than the upper Neosho River, designating the latter as a tributary to the main stem. This fact can be important when interpreting NMT population boundaries and divergence.

Need

Narrative Outline #112 of the USFWS Neosho Madtom Recovery Plan states, “Molecular systematics research of specimens from throughout the range of the Neosho Madtom should be conducted to help define the boundaries of the populations” (USFWS, 1991). Whitacre (2022) evaluated genetic variation in the Neosho Madtom using whole-genome sequencing. But the spatial scope of that study was limited, comprising 10 individuals collected from three study sites, thus limiting inferences that can be drawn regarding population boundaries, status, and trends, which all are essential information necessary to guide proactive management.

This Study

Our project provides a robust population genetic analysis needed to accurately delineate units for management in threatened stream fishes (Chafin et al., 2020). It is based on a comprehensive spatial sampling design comprising sites and individuals collected throughout the Neosho River system. This design allowed us to conduct a rigorous analysis of genetic variation and population structure in Neosho Madtom and represents an appropriately dense geographic and genomic sampling scheme.

To thoroughly evaluate genetic diversity genome-wide, we applied a genotyping method that assays genetic variation across a reduced but representative sample of the genome (Chafin et al., 2017), referred to as double digest restriction-site associated DNA sequencing (ddRAD; Peterson et al., 2012) of Single Nucleotide Polymorphisms (SNPs). This approach has become standard in conservation genetics and has been successfully applied in many species, including rare and endangered fishes (Bangs et al., 2018, 2020a, 2020b; Chafin et al., 2019, 2020; Mussmann et al., 2020a) as well as entire communities of freshwater fishes (Zbinden et al., 2022, 2023).

Goals

We quantified the genetic variation of this imperiled fish in the highly impounded Neosho River system to determine the boundaries, connectivity, and genetic characteristics of Neosho Madtom populations across its range. Insights from this study can now be applied to other facets of its recovery plan, including:

- Task 11: "Determine population size and mobility of Neosho Madtoms"
- Task 15: "Study the feasibility of artificial propagation"
- Task 43: "Develop site-specific reintroduction plans"

II. RESEARCH OBJECTIVES

We addressed six objectives to inform conservation planning:

1. We conducted comprehensive spatial sampling by collecting Neosho Madtom tissues (N=192) from localities (N=14) throughout the Neosho River system (Neosho, Cotttonwood, and Spring rivers).

2. We quantified genetic diversity within and among populations by genotyping thousands of genetic markers (loci; N=2,725) for each sample.

3. We determined distinct genetic population boundaries, quantified population connectivity (gene flow), and delineated potential management units by determining spatial genetic structure.

4. We identified environmental features that affect Neosho Madtom dispersal by correlating spatial genetic divergence with riverscape characteristics between sampling locations.

5. We determined demographic status, population viability, and trends by deriving standard population genetic parameters (e.g., effective population size, N_e).

6. We tested for local adaptation by identifying loci under selection along environmental gradients and provided validation of these candidate markers by elucidating the biological processes with which they are associated.

III. METHODS

Sampling and Tissue Acquisition

All necessary collection permits were approved before the sampling commenced. Sampling was conducted from August 8 to 13, 2021, across 19 sites within the Neosho River system (Table 1). At each site, fishes were collected from a 4.5 m² area by agitating the substrate 3 m upstream from a stationary 1.5 m wide, 3 mm mesh seine and moving downstream to the seine in a zigzag pattern. Following each seining event, *Noturus* individuals were temporarily housed in a bucket filled with aerated stream water for processing. Sampling ceased once 16 NMTs were captured—or the habitat was exhaustively surveyed—at a site, marking the onset of non-lethal processing.

Each *Noturus* individual was first photographed alongside a unique identifier for cross-referencing. The upper lobe of each individual's caudal fin was then excised using scissors and secured in an individually labeled tube containing 95% ethanol before releasing the fish back into the water. Procedures were approved under University of Illinois IACUC protocol (#20123).

Lab Work – DNA Extraction

The procedures for samples processing, genomic DNA extraction and generation of genetic data followed established, published methods (Chafin et al., 2017; Zbinden et al., 2022, 2023). Comprehensive details of these protocols can be found in Appendix 1.

The initial step entailed extracting DNA from each tissue sample, the template for generating genetic data through sequencing. The QIAamp Fast DNA Tissue Kit extraction protocol we used produced highly pure DNA samples (Seufi et al., 2020). The quantity and quality of DNA were assessed via fluorometry and gel electrophoresis to ensure adequacy for subsequent analysis.

Lab Work – Genetic Data Generation

Genetic diversity within and among madtom species was quantified to detect potential hybridization, evaluate population structure, assess connectivity, and determine local adaptation. This assessment was accomplished by genotyping loci (genetic markers) known as Single Nucleotide Polymorphisms (SNPs) from each individual DNA sample. SNPs represent the most common type of genetic mutation. Notably, most SNPs are assumed 'neutral', meaning they are not under selection (i.e., adaptive); most reflect variation in non-coding parts of the genome regions and do not result in changes in amino acids of proteins (Kimura, 1991). SNPs are widely utilized in wildlife and fisheries management to deduce genetic population structure, designate conservation units (Funk et al., 2012; Mussmann et al., 2020a), and detect hybridization among closely related species (Bangs et al., 2018; Russell et al., 2019; Zbinden et al., 2023). For simplicity, we refer to these SNPs as 'loci'.

For genetic data generation, we devised protocols employing double digest restriction-site associated DNA sequencing (ddRAD; Peterson et al., 2012). This approach encapsulates genetic diversity across the genome by consistently subsampling regions from each individual's DNA. Sequencing the entire genome for hundreds of individuals would be financially and logistically

impractical, and the ddRAD method reduces the genome repeatably across each sample, enabling analysis of the same/homologous loci (genome regions). Detailed procedures on genetic data generation and bioinformatic processing can be found in Appendix 1 and 2.

The initial data processing step involved bioinformatic screeing of the sequencing output to eliminate low-quality sequences using standard criteria. This quality assurance is critical to ensure the reliability of our findings. We then aligned the sequences across samples, matching the same loci (homologous sequences) for each individuals. For this alignment, we utilized the published Neosho Madtom (NMT) genome as a reference to construct the assembly (Whitacre et al., 2022), and we employed the IPYRAD software for this purpose (Eaton & Overcast, 2020).

To ensure consistency of data across individuals and study sites and to maximize accuracy while minimizing missing data, we applied standard bioinformatic filtering approaches (Chafin et al., 2017; Eaton et al., 2017). These approaches are crucial for refining the data set, ensuring that only the most reliable genetic information is included in our analysis. The result was a comprehensive panel of genomic loci genotyped across individuals, with thousands of loci identified for both multispecies (*Noturus*; *N*=6,826 loci) and single-species (NMT; *N*=2,725 loci) analyses. These data enabled us to thoroughly assess admixture (genetic mixing) among madtom species, evaluate fine-scale population structures, explore the environmental factors correlating with genetic divergence, and detect signs of local adaptation. Such detailed genetic insights are crucial for effective conservation strategies and understanding evolutionary processes in these species. Details about bioinformatic processing can be found in Appendix 2.

Analysis – Screening for Hybrid Madtoms

To focus our analyses on genetic patterns of 'pure' NMT, we excluded any individuals exhibiting signs of genetic admixture or hybridization with other species: Slender Madtom (*Noturus exilis*), Freckled Madtom (*N. nocturnus*) or Stonecat (*N. flavus*). We screened for admixed/hybrid individuals in our genetic dataset using two clustering techniques:

- Principal Component Analysis (PCA) is a technique used to capture the most significant patterns of variation within the data (among species variation) by finding principal components (orthogonal axes) along which the data vary the most, focusing on preserving global structure (Lewis et al., 2011).

- Sparse Non-Negative Matrix Factorization (sNMF) is a method that breaks down the data into simpler, meaningful components, helping to unveil patterns or groups. One of its advantages is that it provides a way to measure the ancestry proportions of individuals to different groups, offering a more nuanced approach to identifying individuals with mixed ancestry, which is beneficial in screening for admixed or hybridized individuals in genetic studies (Frichot et al., 2014).

These techniques facilitate identifying individuals with 'intermediate' genotypes that reflect ancestry from multiple species. In subsequent analyses, admixed individuals and individuals

representing other *Noturus* species (not NMT) were excluded. Detailed information about these methods can be found in Appendix 3.

Analysis – Inferring Population Structure

The designation of management units or stocks hinges on understanding a study species' natural genetic structure or population boundaries (Mussmann et al., 2020a). We evaluated whether NMT persists as a single or multiple genetic populations to help inform management by clustering individuals based on genetic diversity. Again, we used multiple methods to validate our findings. Comprehensive details are provided in Appendix 3.

Three methods were employed:

- *K*-means clustering with Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010): Groups individuals into specified number (*K*) of distinct clusters based on the genetic variation that distinguishes these clusters (useful for discerning clear-cut groups within the data).

- t-distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten & Hinton, 2008): Visualization of genetic data using a technique that plots it into two dimensions via non-linear dimensionality reduction; it reveals clusters, but highlights local structure. Unlike DAPC, which linearly transforms the data, t-SNE provides a non-linear mapping, making it capable of capturing more complex relationships and fine-scale structures.

- ADMIXTURE: A biogeographical ancestry analysis (Alexander et al., 2009) similar to how commercial DNA tests reveal a person's ancestry makeup. It is a model-based estimation of the proportions of ancestry each individual has from inferred genetic populations. This analysis is beneficial for a more quantitative understanding of individuals' genetic backgrounds, helping to identify mixed ancestry and migrants, which could be crucial in management decisions.

Once genetically discrete clusters were identified, individuals were grouped into populations to explore how genetic diversity was distributed within and among them. We estimated divergence among populations using three different approaches:

- We quantified the amount of genetic variation distributed at three hierarchical levels: among populations, among sites within populations, and among individuals within sites. This analysis of molecular variation (AMOVA; Excoffier et al., 1992) allows to explore if differences are among major groups (e.g., river regions or populations) or localized (e.g., sites), but we refrained testing significance and reported the variance components alone (Meirmans, 2015).

- We calculated pairwise genetic divergence between populations using an unbiased F_{ST} estimate (G''_{ST}; Meirmans et al., 2011), wherein the values can range from 0 (indicating no difference) to 1 (indicating complete difference). To ascertain if these genetic

differences were notably greater than zero, we accessed significance through bootstrap re-sampling and Bonferroni correction for multiple tests.

- Because F_{ST} is more influenced by common genetic variants, we also calculated a metric that equally weights both common and rare alleles (β-Differentiation = ${}^{0}H_{\beta}$) and estimates the proportion of alleles *not* shared among each population. An allele is one of two possible versions of a locus (a reference and alternative, e.g., 'A' or 'G'). The two indices can hint at recent population bottlenecks or expansion because they vary for populations that have undergone recent demographic shifts (Sherwin et al., 2017).

Analysis – Riverscape Genetics

To explore what environmental factors affect the dispersal of NMT and, consequently, population connectivity, we placed the findings from our analysis of population genetic structure within the 'riverscape' context, i.e., the environment between sites influences their connectivity. We hypothesized that if a particular environmental feature (e.g., low head dam) hinders individual dispersal (i.e., acts as a barrier), it will manifest as heightened levels of genetic divergence. This inference could only be accomplished indirectly by examining how observed patterns of genetic divergence relate to environmental characteristics. A thorough explanation of this approach is provided in Appendix 3.

- First, we mapped pairwise genetic divergence estimates (F_{ST}) onto a graph network of stream segments utilizing AUTOSTREAMTREE (Chafin et al., 2023a); this yielded an estimated genetic divergence for each segment within the Neosho River stream network.

- Next, we assessed the effect of individual dams within the stream network on inflating genetic divergence by employing the standardized genetic fragmentation index (F_{INDEX}) proposed by Prunier et al. (2020). Populations located immediately upstream and downstream relative to each barrier were sampled to yield a standardized index ranging from 0%, indicating no barrier to dispersal, to 100%, representing a complete barrier to dispersal.

- Finally, we modeled the individual effects of environmental features on genetic divergence using RESISTNET (Chafin et al., 2023b). We compiled a comprehensive array of environmental features (N=281) derived from the HydroATLAS v.0.1 database (Linke et al., 2019). To enhance this dataset, we calculated several variables indicative of dam effects: (i) barrier density per river segment, (ii) age of the oldest barrier within a segment, and (iii) indices of river fragmentation and connectivity (Grill et al., 2019). We evaluated each variable's association with genetic divergence (F_{ST}) using a random forest regression approach (Pedregosa et al., 2011). Then, we modeled movement resistance by employing the retained environmental features as inputs for RESISTNET (Chafin et al., 2023b).

Analysis – Estimating Population Genetic Parameters

Genetic diversity indicates a population's viability, resilience, and adaptability to changing environmental conditions. Populations with higher genetic diversity are generally more adept at navigating stressors such as diseases, climate change, and habitat degradation. Consequently, the levels of genetic diversity within each discrete population can be informative about its potential for long-term persistence. Comprehensive details of methods to estimate population genetic parameters are provided in Appendix 3.

The comparison of population-level diversity was conducted with various estimates of genetic diversity derived from the genotype data. These estimates include:

- mean allelic richness per locus
- count of alleles unique to each population (private alleles)
- inbreeding extent within a population (F_{IS})
- heterozygosity (expected and observed across polymorphic sites)
- evolvability based on the Shannon Index
- A q-profile of the effective number of alleles (Hill Numbers) across different estimates (Sherwin et al., 2017)

We also calculated the genome-wide heterozygosity (H_{GW}) from the unfiltered alignment containing polymorphic and non-polymorphic loci for comparison to other studies. Heterozygosity based only on polymorphic loci will be biased upwards, and so genome-wide heterozygosity facilitates better comparison across studies (Schmidt et al., 2021).

We estimated each population's effective population size (N_e) with NeEstimator v.2 (Do et al., 2014), and we tested for evidence of a recent population bottleneck using Tajima's *D* (Tajima, 1989).

Analysis – Loci Under Selection and Local Adaptation

We assessed signals of local adaptation in NMT via genotype-environment association analysis (GEA; Lotterhos & Whitlock, 2015). A comprehensive set of environmental variables (N=281) was compiled from the HydroATLAS v.0.1 database (Linke et al., 2019) and categorized into five groups: (i) hydrological and physiographic, (ii) climate, (iii) landcover, (iv) geology and soils, and (v) anthropogenic characteristics. To mitigate collinearity, variables within each category were consolidated into composite variables using Robust Principal Component Analysis (ROBPCA; Reynkens, 2018).

The relationships between loci and the composite environmental variables were evaluated using Redundancy Analysis (RDA), a multivariate extension of multiple linear regression (Forester et al., 2018). Adaptive loci were inferred as deviating more than ±3 standard deviations from the mean loading of all loci on the canonical axes generated by RDA that predict genotype-environment correlations. Notably, most loci are anticipated to be neutral (i.e., not adaptive) as they predominantly occur in non-coding regions of the genome or result in synonymous mutations that do not alter amino acids (Kimura, 1991).

Our next objective was to discern whether distinct populations or areas within the Neosho River system exhibited variations in local adaptation (i.e., different adaptations to different environments). To accomplish this, we repeated population structure analysis but only used loci identified as potentially adaptive in the RDA. Because of t-SNE's proficiency in rapidly and accurately depicting hierarchical structure (Li et al., 2017), we used this approach to re-examine population structure based on adaptive loci. We expected that geographical regions with different local adaptations would form separate clusters in the t-SNE ordination space.

Finally, we provided validation of the biological importance of these adaptive loci by exploring the biological processes that might drive local adaptation in NMT. For this, we needed to identify if our loci mapped onto genes with known biological functions. We accomplished this by first matching functional annotations (i.e., known genes) from the published Channel Catfish (*Ictalurus punctatus*) genome (Liu et al., 2016) to the NMT genome used to guide our data alignment (Whitacre et al., 2022). The available NMT genome does not have such functional annotations. This matching was done by mapping each feature (e.g., exons, transcripts) using an iterative alignment procedure (Shumate & Salzberg, 2021). Functional effects and gene associations of the candidate adaptive loci were predicted using SnpEff (Cingolani et al., 2012), and information known about each associated gene (gene ontology) was compiled using the QuickGO database (Binns et al., 2009). Comprehensive details related to local adaptation and all other analyses are provided in Appendix 3.

IV. RESULTS

Sampling

Across the Neosho River system, 19 areas were sampled, with madtoms captured at 15 sites, yielding *N*=249 tissue samples for DNA analysis (Table 1). Neosho Madtom tissues were collected at 14 sites, including locations from the Cottonwood (*N*=3), Neosho (*N*=9), and Spring (*N*=2) rivers (Figure 1), resulting in *N*=192 individual fin clips representing NMT (number per site =2–16; mean =13.2).

Additionally, to screen for potential hybridization, fin clips were collected from other madtom species across 15 sampling sites (Table 1), including the Freckled Madtom (*N*=19), Slender Madtom (*N*=2), and Stonecat (*N*=36).

Genetic Data and Screening for Hybrids

We screened for hybrids using a multi-species loci alignment produced with the IPYRAD software. These data comprised 85,364 SNPs across 18,713 reads (100 base-pair regions) for *N*=249 individuals representing the four madtom species (Neosho, Freckled, Slender, Stonecat). We examined genetic patterns across 6,826 loci/SNPs (one SNP per read) for *N*=238 individuals (Figure S1).

Hybridization between NMT and Stonecat was apparent from the mixed genetic ancestry of three individuals. The first individual (86NNLAB09), phenotypically identified as NMT, possessed 28% Stonecat ancestry; the other two, identified as Stonecats, revealed 13% (86TNROY22) and 2% (86TNOSW02) NMT ancestry, respectively (Figure S2). All three individuals, collected from sites in the Neosho River downstream of John Redmond Reservoir (NROY, NOSW, NLAB; Figure 1), were excluded from subsequent analyses. The remaining *N*=235 individuals displayed an ancestry proportion >99.9%, concordant with their phenotypic species identification.

For the population genetic analyses, we created a genotype dataset incorporating only NMT samples using a single-species loci alignment with IPYRAD. The resulting data consisted of 9,345 SNPs across 5,584 reads (100 base-pair regions) for *N*=185 NMT individuals. This dataset contained fewer SNPs and reads because genetic variation is elevated when compared among species but reduced within species when many loci are expected to be fixed (i.e., invariant within NMT).

After standard filtering, our genetic panel for analyses contained 2,725 loci/SNPs (one SNP per read) across *N*=178 NMT individuals (Figure S3). Sequencing depth, a measure of how many times a particular locus was represented by a sequence (Peterson et al., 2012), was high across the dataset (mean depth = 78.5x), giving us high confidence in the accuracy of our genotype data. The mean missing data per individual (11.1%) was low and desirable for a SNP genotype dataset (Zbinden et al., 2022, 2023).

Population Structure

NMT samples clustered into six genetic populations based on results from several analyses (Tables 2, 3; Figure 3). Analyses to explore population structure, including Admixture analysis (Figure 4) and indices of genetic divergence, suggested genetic structure is hierarchical, with the upper Neosho River sites (NDUN & NAME) and the Spring River sites (SMCP & SWAC) being relatively more diverged from the other sites (Figure 4). Admixture validation showed that number of populations between K=2 and K=6 were better models than a single population (K=1; Figure S4). Because not all populations are equally divergent, and differ much more (Meirmans, 2015), one measure of fit (cross-validation error) indicated K=3 was optimal, while another measure (log-likelihood) did not clearly plateau with increased populations (Figure S4).

Based on a mapping of the population structure results (Figure 5), we grouped the NMT sampling sites into six potential Management Units: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and Spring River (SPR). Genetic variation was partitioned hierarchically, with AMOVA estimates indicating 12.7% attributed to among the six populations, 1.5% among sites within populations, and 85% among individuals within sites.

Estimates of pairwise genetic divergence (F_{ST}) between populations of NMT all reflected isolation. Values were significantly different from zero (all Bonferroni adjusted p<0.0014) and ranged from 0.02 (CWR *vs.* NCC) to 0.18 (NRU *vs.* SPR) (Table 2). Populations NRU and SPR showed higher divergence from all others ($\bar{x}$ =0.12), with the remaining four differing less amongst themselves ($\bar{x}$ =0.03). These results are consistent with the hierarchical structure observed in the Admixture analysis (above).

Genetic divergence based on the proportion of unshared alleles among populations (β-Differentiation = ${}^{0}H_{\beta}$; Table 3) showed relatively similar patterns. These values were generally larger, ranging from 0.08 (CWR *vs* NCC) to 0.15 (NRU *vs* SPR) because they take unique, rare alleles into account.

Riverscape Genetics

An analysis of genetic patterns within a riverscape context helped us understand how the environmental characteristics of river reaches between populations influence the dispersal of NMT. We were especially interested in whether the presence of dams affected population connectivity (i.e., gene flow).

Mapping pairwise genetic divergences onto stream segments (AUTOSTREAMTREE) revealed patterns consistent with the population structure analysis results above. Divergence values reflect limited genetic interchange among populations, suggesting reduced connectivity among NMT populations due to dispersal barriers. The largest fitted F_{ST} values (>0.05) were on segments separating the two most distinct populations, SPR and NRU, from the remaining populations (Figure 6a). Intermediate F_{ST} values (>0.02) were found for the segments separating CWR from NCC, as well as that separating NCC from downstream Neosho reaches (NRM, NRL).

Assessing whether dams across the Neosho River system act as barriers clearly showed that dams inflate genetic divergence (F_{INDEX}). All dams in our dataset seemed to affect the dispersal of NMT between different river sections, reducing gene flow (Figure 6c). John Redmond Dam, the tallest and only non-low head dam between sampling sites, was the most substantial barrier to gene flow.

Dams were also associated with increased network resistance across the riverscape (RESISTNET). Two indices that directly reflect the effects of dams, namely degree of fragmentation and flow regulation, were among the eight most strongly associated variables in the network resistance model (relative importance >0.8; Figure 7c). The degree of fragmentation approximates longitudinal resistance at the reach scale. In contrast, flow regulation captures temporal fluctuations of river flow (i.e., both variability in flow and a shift in the timing of flow events) associated with the same dams. The remaining top-ranked variables were related to landscape features such as soil types, landcover and vegetation types, anthropogenic development, and protected area extent (Figure 7c).

Population Genetic Parameters

The most genetically distinct populations, NRU and SPR, showed lower overall genetic diversity than the other four populations (Table 4). However, SPR revealed the highest number of unique alleles (N=77) (i.e., private alleles only found in the Spring River). By contrast, CWR, NCC, and NRL stood out for their relatively higher allelic richness and Shannon Entropy, a measure of population evolvability (i.e., the capacity to generate heritable and adaptive phenotypic variation; Sherwin et al., 2017).

Rare genetic variants significantly contributed to local genetic diversity (Figure 8). Genetic divergence among populations based on the simple proportion of unshared alleles, ${}^{0}H_{\beta}$, was greater than F_{ST}. These results show that measures of observed heterozygosity (H_O) and its derivatives (F_{ST}) underestimate 'true' diversity because they are insensitive to rare alleles.

All populations showed low levels of inbreeding based on our calculations of the inbreeding coefficient F_{IS}, which can range from -1 to 1. Values near zero indicate random breeding, while positive values indicate inbreeding (a deficit of heterozygotes) and negative values outbreeding (an excess of heterozygotes). Only one population, NRL, had a positive F_{IS} estimate, which was still close to zero (F_{IS} =0.015).

Estimates of effective population size (N_e) generally indicate that the short-term (decades) extinction risk for all NMT populations appears low. At the same time, the long-term viability (centuries) is questionable for half of the populations. The '50/500' rule implies that N_e>50 is necessary to ensure short-term survival, whereas N_e>500 is needed for long-term viability (Rieman & Allendorf, 2001; Jamieson & Allendorf, 2012). The CWR, NRL, and NRM estimates revealed N_e>1,000, whereas NCC, NRU, and SPR had N_e<500 (Table 5). CWR and NRL seemed to be the genetically most diverse populations, with relatively higher local genetic diversity and effective population size.

We also found evidence consistent with a population bottleneck. The neutrality statistic, Tajima's D, was greater than 0 across all populations (Table 5). Historical demographic modeling could

provide more insight into the population divergence times and size changes, but our data might not be suitable for the type of analyses that can explore such patterns.

Selection and Local Adaptation

Our analyses indicated patterns consistent with local adaptation along the upstream-downstream longitudinal gradient of the Neosho River system. Genotype-environment association analysis (GEA) revealed adaptive loci significantly associated with environmental variation. These loci were structured differently among populations, emphasizing differences in regional adaptation.

Five composite environmental factors (robust principal components) accounted for 87% of the total environmental variance across N=281 factors. The five composite factors represented (i) hydrological and physiographic, (ii) climate, (iii) landcover, (iv) geology and soils, and (v) anthropogenic environmental characteristics (Table S2). For context, system-wide summary statistics for the original environmental factors with the highest loadings on the five principal components are available in the Supplemental Material (Table S3). Select factors are reported for each site/population (Table S4).

Environmental factors were significantly related to individual genetic variation (p=0.001), accounting for 6% of the total genetic variance within NMT. Most of the explained variance was accounted for by hydrologic-physiographic (39%), climate (27%), and landcover (16%) (Table S5). Again, most loci should be neutral concerning environmental selection, so 6% of genetic variance attributable to environmental variation is considerable (Meirmans, 2015; Brauer et al., 2018). The canonical axes RDA produced (N=5) represent the association between genetic and environmental variance (Figure 9), and RDA triplots visually represent these relationships (Figure 10). Each environmental variable is correlated to some degree (loading) with each RDA canonical axis (Table S6).

The analysis also estimated the correlation (loading) on each RDA canonical axis for every locus (Figure 9). We screened for adaptive outlier loci that would indicate selection (local adaptation) by isolating those with loadings ±3 standard deviations from the mean on each axis (N=5), and we inferred that N=61 loci were locally adapted outliers. Most of these loci were most strongly correlated with either hydrologic-physiographic (N=20) or landcover (N=18) factors (Table S7). Original factors most correlated with these loci (Table S8) represent a variety of factors but point to the adaptive importance of characteristics both within the stream reach (size) and upstream of the reach (erosion and forest cover). Figure S5 provides for an overview of all the individual environmental factors considered.

Geographic differences in local adaptation to the environment were apparent based on the clustering of adaptive loci (Figures 10, 11). The populations upstream of John Redmond Reservoir consisted of CWR, NRU, and NCC and overlapped in their variation for these adaptive loci (Figure 12). The populations downstream of John Redmond Reservoir consisted of the remaining populations NRM, NRL, and SPR and also overlapped in their adaptive variation (Axis 2 of t-SNE plot, Figure 12). However, differences are also apparent (Axis 1 of t-SNE plot, Figure 12). These results suggest differential local adaptation along the longitudinal gradient of the Neosho River system that is consistent with the directional gradient we see for most of the

environmental variables (Figures S6−S14). Population-level allele frequencies for the adaptive loci can be found in the Supplemental Material (Table S9).

Patterns consistent with local adaptation were also apparent from the associations between these adaptive loci and their genetic/biological consequences. For the *N*=61 adaptive loci: *N*=6 loci were associated with amino acid changes within protein-coding regions (=Moderate); *N*=2 loci were associated with synonymous changes within protein-coding regions (=Low); the remaining *N*=53 loci were modifier mutations in non-coding regions (Table S10). These mutations were collectively associated with *N*=30 genes (Table 6; Figure 11) whose known biological functions showed consistent patterns in that they are associated with anatomical development (head, eyes, brain), response to stimuli (stress, chemical, hypoxia), and metabolism (autophagy, ceramide, nitrogen, ubiquitin).

V. DISCUSSION

Overview

We collected NMT at most sites we visited in the Neosho River system (Figure 1), and abundances were generally high enough to gather our target of 16 tissue samples per site (Table 1, Figure S1). We also obtained tissues from other native *Noturus* species, and it appears that NMT may hybridize with Stonecat, specifically downstream of John Redmond Reservoir (Figure S2).

The genetic population structure of NMT appears hierarchical (Figure 3), with two rather distinct populations (NRU and SPR) and four additional populations along the main stem of the Neosho River system (CWR, NCC, NRM, NRL; Table 2; Figures 4, 5). Riverscape environmental characteristics (i.e., dams; Figure 2) seem to mediate dispersal among populations of NMT (Figures 6, 7).

All populations of NMT exhibited unique genetic diversity (Figure 8). Populations appeared safe from extinction risk due to severe inbreeding depression over the coming decades based on their effective population sizes (Table 5). Three populations (NRU, NCC, SPR) showed indicators of vulnerability to long-term extinction and could benefit from management to maintain genetic diversity. However, conservation efforts must be balanced against potential effects on local adaptation across the system. A clear signal of local adaptation along the upstream-downstream (i.e., 'longitudinal') gradient of the system (Figures 9−11) is consistent with past observations related to turbidity and NMT density along this gradient (both higher upstream than downstream John Redmond Reservoir; Wildhaber, 2011).

Sampling Design and Genetic Data

Our sampling design allowed robust inferences because we sampled widely and densely (Table 1; Figure 1). Other work has demonstrated that genetic diversity studies using next-generation sequencing data should target eight or more individuals per sampling location to guarantee accurate estimates of indices such as genetic divergence (Nazareno et al., 2017). Only two of our 14 sampling locations were represented by fewer than 12 individuals.

Our sampling design contrasts with another genetic study of NMT by Whitacre et al. (2022). The latter was based on 10 NMT individuals (5.2% of our sample size) collected from three sites roughly corresponding to our NRU, CWR, and NRM populations. Despite this limited sample size, Whitacre et al. (2022) identified 'weak population structure' via principal component analysis, but concluded that NMT from three sampling locations represented one panmictic population. Prior to Whitacre et al. (2022), the structure of NMT genetic diversity was unknown. We note that the authors did mention, "However, denser geographical sampling is needed to investigate this hypothesis [no structure] further" (Whitacre et al., 2002).

A major contribution of Whitacre et al. (2002) was the generation of whole-genome data, which increased the power of our analyses because it provided more information about our loci (i.e.,

assembly of our loci and linkage to functional genes). We note that more information at the individual level (i.e., whole genome data) is a trade-off, because funds then only allow analysis of fewer samples, limiting spatial coverage and site representation, which can hinder the inference of genetic population structure (Meirmans, 2015; Storfer et al., 2018). Thus, it is crucial to select the appropriate approach for each objective. Comparisons have shown that reduced representation genomic data (e.g., ddRAD, as used in our study) provides the same perspective of population structure as whole genome sequencing when sampling designs are the same (Martchenko and Shafer, 2023).

Whitacre et al. (2022) did not include samples from the Spring River (SPR), one of our study's most genetically distinct populations. However, their results pointed to the genetic distinctness of the Upper Neosho River: Their first principal component of NMT genetic variation separated their Upper Neosho River individuals from the remainder and explained 13.5% of the total variation. The same PCA did not show genetic separation between the Cottonwood and Lower Neosho River populations (but see their Supplementary Figure 2b, which shows three clusters).

Hybridization

Detection of genetic admixture between NMT and Stonecat within our study (Figure S2), albeit rare, may allude to occurrences of hybridization, a phenomenon localized primarily downstream of the John Redmond Reservoir. Such hybridization events are frequently associated with environmental fluctuations and perturbations (Bangs et al., 2017; Douglas and Douglas, 2010; Zbinden et al., 2023). In particular, regulated water releases around spawning periods might inadvertently propel these species to utilize similar habitats, thus facilitating the chances of hybridization.

Spawning of NMT is challenging to study due to high turbidity and river velocity in spring – the species' presumed spawning season. NMT spawning has been linked with increasing photoperiod and temperature concomitant with the spring rise in discharge (Bulger et al. 2002), and spawning may begin as early as March (Moss, 1981) and is assumed to mainly occur from May to July in shallower riffle areas (Bulger & Edds, 2001). Stonecat spawn during the same time period in deep habitats (Brewer et al., 2006).

Thus, flow regulations become especially pertinent when considering the reproductive periods of NMT and Stonecat. It is imperative to reiterate the importance of coordinating flow releases to mimic natural spring flows that align with NMT spawning periods (USFWS, 1991). Such coordination is instrumental in ensuring successful spawning and preserving the reproductive boundaries between NMT and other *Noturus* species, thereby playing a crucial role in the long-term persistence and genetic viability of NMT.

Population Structure

We rejected the hypothesis that NMT represents a single panmictic population based on concordant results across analyses of genetic variation (Tables 2, 3; Figures 3, 4). Our analyses are based on the densest geographic sampling of NMT DNA to date (Table 1).

Clustering analysis to determine the number (*K*) of genetically distinct gene pools revealed a total of six populations. This *K*=6 population model of genetic variation is consistent with the structure observed across the other analyses (*K*-means clustering, DAPC, t-SNE, F_{ST}; Table 3; Figure 3). However, NMT in two areas represented more genetically distinct populations: the Upper Neosho River (NRU) and Spring River (SPR), indicating not all populations are equally divergent, with some more similar to oneanother. This is refered to as hierarchical population structure within a species—when some populations are much more genetically distinct than others— and it can be challenging to tease apart and visualize (i.e., 'the curse of dimensionality;' Schmidt et al., 2023). For example, clustering individuals into three populations with the analysis ADMIXTURE (*K*=3; Figure 4) produced the lowest assignment error via cross-validation, a procedure that helps evaluate how well the model fits the data (Figure S4). This suggests three populations is a good model to reflect genetic structure in NMT. But another indicator of model fit (log-likelihood) did not clearly plateau with more populations and instead increased (Figure S4). Thus, the data need to be interpreted by carefully considering results from different analyses with which we explored our data.

It is important to note that *K* is a useful metric indicative of the number of populations but should also be viewed as a flexible parameter describing one point on a continuous scale of population structure (Jombart & Collins, 2016; Verity & Nichols, 2016). This concept has been seemingly lost in translation since its inception (Pritchard et al., 2010), with most researchers simply reporting the 'best *K*' that minimizes cross-validation error. We advocate for presenting the results from a range of *K* models (as done in Figure 4) and determining the number of genetic populations and management units based on what makes sense biologically and geographically (Meirmans, 2015).

Dispersal of individuals between sites can be obstructed by hard barriers such as dams or soft barriers like sub-optimal habitats, ultimately leading to different genetic populations (Hopken et al., 2013; Zbinden et al., 2022). A lack of dispersal leads to lowered or total absence of exchange of genetic variation (gene flow) between the populations. Consequently, alterations to genetic variation arising from mutation, natural selection, or random loss of genetic variation due to small population size, occur independently within such populations devoid of gene flow. Given enough time, this independence of each isolated population will be reflected as unique genetic variation. These patterns can be detected with population genetic analyses and individuals from discrete populations will be identifiable, and cluster together because they are genetically similar.

A combination of recently isolated and longer-ago separated populations reflecting deeper divergence patterns may drive the hierarchical population structure observed across NMT. Within this context, NRU and SPR became more genetically distinct because they were isolated from the other populations longer ago, and thus could be considered 'older' populations. Alternatively—or perhaps additionally—environmental characteristics might differ among areas

of the Neosho River system and drive adaptive differences. The upper Neosho and Spring rivers are characterized by less average discharge (Table S4; Figure S6) and thus could be considered 'tributaries' to the 'main stem, comprising the Cottonwood and lower Neosho rivers. Given the species' predilection for consistent flows over specific substrates, we hypothesize that NMT individuals naturally tend towards the path with more discharge when dispersing upstream in the pre-impoundment past. This assumption is consistent with their almost exclusively main stem distribution. It could also explain why the populations in the two large tributaries (NRU and SPR) are distinct: these populations could have originated from founder events when some individuals occasionally chose the path with lower discharge instead of the mainstem. Coupled with slightly different habitats in the smaller streams, this could accumulate genetic differences due to stochastic (i.e., genetic drift in small populations) and deterministic processes (i.e., selection in a different environment).

Our data showed signals consistent with historic dispersal being biased in the downstream direction but no evidence of contemporary dispersal. The most upstream populations, NRU and SPR, exhibit minimal admixture with the other populations, which was also true for CWR, save the most downstream site, CESG, which is relatively near the adjacent population, NCC, and not blocked by an impoundment (Figures 4, 5). The population at the upper Neosho and Cottonwood rivers confluence (NCC) showed signs of historic admixture, as most individuals had signs of ancestry from CWR and NRU (upstream). Similarly, the population on the lowest stretch of the Neosho River (NRL) exhibited historic admixture based on ancestry from the upstream populations, NRM and NCC (Figures 4, 5). However, there was no indication of contemporary or first-generation migrants in any population—they would have appeared as an inconsistently-colored vertical bar in Figure 4.

Riverscape Genetics

The riverscape genetics analysis highlighted the effects of environmental factors on populations of NMT. Natural processes and anthropogenic influence modulate dispersal and affect population connectivity and structure. Distinguishing natural from anthropogenic effects is complex, but exploring nuanced signals of genetic patterns allows indirect inferences.

Regarding population structure, the genetic distinction separating the SPR and NRU populations from the remaining NMT populations (Figure 6a) likely reflects historical (i.e., natural) processes that emerged over extended periods before impoundments. Conversely, the more moderate divergences between other populations (Figure 6a), especially around the segment housing John Redmond Dam, correspond with more recent anthropogenic habitat alterations, underscoring the likely role of artificial barriers in blocking dispersal.

Analyses indicated that dams act as barriers to dispersal, reduce population connectivity, and contribute to local isolation of NMT (Figure 6c). John Redmond Dam, the largest among the array that separates NMT populations, emerged as a formidable barrier to NMT dispersal (highest F_{INDEX} value; Figure 6c). This index underscores an inflation of genetic divergence between adjacent populations, attributable to the dam-induced reduction in population connectivity.

The imprint of dams was further highlighted in the RESISTNET analysis, where variables indicative of dam effects, such as degree of fragmentation and flow regulation, were identified as significant factors (Figure 7c). Other landscape-level factors also impede NMT dispersal. The RESISTNET analysis (Figure 7c) shows that soil types and landcover could be proxies for unmeasured in-stream environmental variation. For example, landscape-level differences in vegetation and soil affect runoff and might relate to riverscape-level fluctuations in turbidity or flows (Fox and Magoulick, 2019) that mediate dispersal.

Population Genetic Parameters

Our quantification of genetic parameters within six distinct populations sheds light on nuanced patterns of genetic diversity. One pertinent finding was rare genetic variants unique to each population, significantly contributing to local genetic diversity and not replicated among populations. This rare diversity is particularly relevant for small-range endemics, such as NMT, and underscores low gene flow between populations.

Another relevant finding was the dichotomy in the potential for short-term *vs.* long-term population viability among populations. Estimate of effective population size (N_e; Table 5) indicated that the short-term risk of extinction due to genetic factors is low for all populations (N_e>50), but 50% of populations may be at risk long-term due to N_e<500 (i.e., the '50/500' rule; Rieman & Allendorf, 2001; Jamieson & Allendorf, 2012). Only populations CWR, NRL, and NRM exceed N_e>1,000, in contrast to populations NCC, NRU, and SPR with smaller N_e<500. Although the '50/500' rule serves as a potential benchmark to categorize 'conservation status,' the unique demographic, ecological, and life history traits of the species in question must be considered for a nuanced interpretation of N_e estimates (Waples, 2023).

The signal of historic population processes (i.e., Tajima's D and uneven genetic diversity; Table 5; Figure 8) was consistent with an hypothesis of recent demographic changes such as population bottlenecks or the signature of natural selection across all populations (values greater than zero; Table 5). These are not mutually exclusive scenarios: our results pertaining to local adaptation support the latter, and a decline in effective population sizes beginning about 10,000 years ago for both NMT and Stonecat was inferred by Whitacre et al. (2022)—although we note the relatively small individual sample size (as above). Historical demographic modeling can estimate population divergence times and size changes, augmenting our understanding of the populations' evolutionary trajectories, but our data might not be appropriate for such analyses.

Genetic variability must be considered both in the context of 'diversity' (i.e., within populations) and 'distinctness' (i.e., among populations). For example, populations NRU and SPR, although genetically distinct, manifested lower overall genetic diversity than CWR, NCC, and NRL. The latter exhibited relatively higher 'diversity' as expressed in allelic richness and Shannon Entropy, both indicators of potential evolvability. SPR revealed the highest number of unique alleles, contributing to its genetic distinctiveness.

The minimal levels of inbreeding across populations, save for a marginal positive F_{IS} estimate in NRL, are encouraging. Nevertheless, the observed genetic divergence, especially the greater divergence revealed by ${}^0H_\beta$ compared to F_{ST}, stresses the importance of maintaining population

structure. A prudent approach would be to consider the maintenance of fine-scale population structure to retain unique local genetic variants and local adaptation. However, if translocation emerges as a requisite, our overall results suggest a CWR to NCC and NRU and an NRL to NRM and SPR translocation strategy that could maintain local adaptation and bolster genetic diversity and effective population sizes.

Selection and Local Adaptation

Understanding the landscape of local adaptation is essential when managing populations or brood stocks (Schmidt et al., 2023). Introducing individuals adapted to different environments may result in offspring ill-suited to the local conditions and thus have low survival, lowering fitness and potentially triggering outbreeding depression. Alternatively, supplementation can cause genetic swamping, where the local gene pool is overwhelmed, erasing locally adapted variation (Flanagan et al., 2018; Hoffmann et al., 2021). Both outcomes are detrimental to population persistence and counter the intended management goal of boosting genetic viability.

Locally adapted differences also can be effectively leveraged (Weeks et al., 2011). Strategic mixing of different brood stocks can serve as a genetic rescue, bolstering genetic diversity, combating inbreeding depression, inducing hybrid vigor, or transferring locally adapted traits to populations where they have not naturally emerged (Whiteley et al., 2015; Hoffmann et al., 2021). However, to distinguish a mere short-term demographic response (i.e., boost in population size due to the introduction of additional individuals) from an actual genetic rescue (i.e., an increase of genetic diversity to enhance fitness), requires quantification of local genetic variation *a priori*, as done herein, as well as assessment of fitness parameters via ecological data (Mussmann et al., 2017). Therefore, the data on local genetic diversity presented in this study will be valuable in informing any management interventions considering supplementation and translocation.

Local adaptation along the Neosho system's upstream-downstream ('longitudinal') gradient was inferred based on GEA analysis and the similarity of populations' adaptive variation (Figure 12). See Appendix 4 for a detailed discussion of GEA analysis caveats. Both turbidity and NMT density have been linked to this gradient previously, and both are generally higher upstream than downstream of John Redmond Reservoir (Wildhaber, 2011) and could be hypothesized to play an underlying role in adaptation. Furthermore, upstream localities fall within smaller river reaches with less precipitation (Figures S6,& S8). NMT upstream of John Redmond Reservoir may experience lower predation due to higher turbidity and optimal interstitial space in the substrate, which provides cover but at the cost of more variable and potentially drier instream conditions that necessitate more inter-riffle movement or movement to other, deeper microhabitats to track environmental fluctuation (Wildhaber, 2011). The reverse may hold for the populations downstream of John Redmond Reservoir, which retains sediments and reduces turbidity, with more stable instream conditions at the cost of higher predation pressure due to lack of cover (Wildhaber et al., 1999). The abundance of black bass species (*Micropterus* spp.), predators of madtoms in the system (Branson, 1967), likely changes along the longitudinal gradient as they prefer larger, more stable habitats (Johnson et al., 2009; Bruckerhoff et al., 2021).

Biological functions associated with the adaptive markers suggest NMT may be adapting to different stressors and resources throughout the Neosho River system. An overarching concern with GEAs is whether outlier loci are indeed biologically relevant, and thus, candidate adaptive markers need further validation to increase confidence in the results (Meirmans, 2015; Schmidt et al., 2023). We provide a first step in validating the candidate adaptive markers by establishing consistent connections between genotypes and phenotypes, i.e., biological functions. The candidate adaptive markers we detected were associated with amino acid changes in protein-coding regions and modifications within non-coding regions that may affect upstream and/or downstream genes and gene complexes. These markers were collectively associated with *N*=30 genes (Table 6) whose known biological functions showed consistent patterns in that they are associated with anatomical development (head, eyes, brain), response to stimuli (stress, chemical, hypoxia), and metabolism (autophagy, ceramide, nitrogen, ubiquitin). Most of the putatively adaptive genes found here have been implicated in the evolution and adaptation of other fishes. See Appendix 4 for references and more discussion about these genes.

Conclusion

Our comprehensive evaluation of genetic diversity and population structure in NMT across its range illuminates a complex interplay between genetic, environmental, and anthropogenic factors. Several forward-focused strategies to safeguard the species' integrity and facilitate its long-term viability can be drawn from Schmidt et al. (2023).

1. **Maintenance of Species Integrity and Genetic Diversity**:
   Preserving the genetic diversity within NMT populations is essential for their adaptive capacity and resilience against environmental changes. A comprehensive understanding and recognition of the existing genetic population boundaries and the variable genetic differences among populations is a prerequisite for such management actions, and this study provides such data to make informed, science-based decisions.

2. **Limiting Hybridization with Other Species**:
   The potential hybridization between NMT and Stonecat, particularly downstream of John Redmond Reservoir, albeit seemingly rare, is of concern because of its ramifications on fitness and population viability. Pro-active management actions should include strategic interventions to coordinate dam releases during spawning times to mitigate such occurrences, ensuring the distinct genetic identity of NMT.

3. **Preserving Local Adaptation**:
   Upholding the historical demographic processes among remnant populations is essential, as it reflects the natural evolutionary trajectory and adaptability of NMT in its native habitat. Importantly, we hypothesize that local adaptation evolved over time before habitat fragmentation and reduced population connectivity due to anthropogenic habitat alterations. Therefore, removing barriers (i.e., dams) would not harm NMT because even unimpeded gene flow is not expected to overwhelm adaptation, given low dispersal propensity of the species. Enhanced natural population connectivity would contribute to long-term population viability by allowing some transfer of genetic diversity among populations, but at a rate low enough to maintain adaptation.

4. **Translocations for Genetic or Evolutionary Rescue**:
   While all populations appear to be genetically viable and with minimal risk for short-term extinction, supplementations aimed to boost population size (demographic response) or enhance genetic diversity (evolutionary capacity) could potentially reduce long-term extinction risk. But benefits of such efforts should be carefully balanced against the potential risk of outbreeding, loss of local adaptation or unique genetic diversity (Weeks et al., 2011).

5. **Genetic Monitoring and Temporal Tracking**:
   Establishing a long-term genetic monitoring program to track genetic changes should be considered as an additional pro-active management strategy. Obtaining such data would provide quantitative metrics that track the evolving genetic landscape of NMT populations and inform adaptive management strategies. A potential implementation tool is the development of a reduced marker panel composed of adaptive and neutral loci (e.g., GTseq; Campbell et al., 2015). Such an approach would facilitate long-term monitoring at a low cost per sample. Our lab has already implemented this strategy for several studies in Arkansas involving game species (i.e., Smallmouth Bass, White-tailed Deer, and Black Bear) and species of conservation concern (i.e., Collard Lizard).

6. **Preservation of Population Structure**:
   A critical consideration for preserving population structure is identifying the underlying processes that shaped this pattern (i.e., historical, natural processes or anthropogenic interventions). Equally important is understanding the extent of local adaptation and its role in population persistence. Mixing populations through supplementation or translocation could potentially jeopardize natural population structure and/or local adaptation, ultimately leading to fitness loss and population decline. Therefore, until a deeper understanding is attained, avoiding human-mediated movement of NMT between populations is prudent, especially among sites upstream and downstream of John Redmond Reservoir. However, human-mediated migration among sites *within* the six management units could promote population viability by increasing population connectivity among sites (i.e., gene flow) and stabilize or enhance effective population size by reducing genetic drift (i.e., prevent random loss of genetic diversity due to small population size).

These outlined strategies underscore a holistic approach toward the genetic conservation and management of NMT populations. The assimilation of genetic knowledge into adaptive management practices will be indispensable in navigating the conservation path ahead to sustain NMT in the Neosho River system and aid in implementing the recovery of the species.

VI. ACKNOWLEDGMENTS

This research was made possible through the generous funding provided by the Kansas Department of Wildlife & Parks (KDWP), with special thanks to project manager Jordan Hofmeier for his invaluable guidance and support. Supplemental funding was provided by generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (MRD), the 21st Century Chair in Global Change Biology (MED).

We are grateful for the permits and cooperation extended by the U.S. Fish and Wildlife Service (USFWS), the KDWP, the Missouri Department of Conservation, the Oklahoma Department of Wildlife Conservation (ODWC), and Emporia State University (ESU). Their contributions were instrumental in the successful execution of this project.

Our work was conducted under the University of Illinois's IACUC protocol (#20123), and we extend our gratitude for their oversight and assistance in ensuring the ethical conduct of our research.

We also wish to acknowledge the support and cooperation of ESU, KDWP, and various other landowners who granted us access to their lands for our fieldwork. Their willingness to support our research endeavors was crucial to our study.

Our field efforts were greatly assisted by the dedication and expertise of T. Ratliff (KDWP), B. Brown (ODWC), C. Gainer (ODWC), B. Johnston (ODWC), and K. Robbins (ODWC). Their contributions were vital in the collection and analysis of our data.

Lastly, we thank the University of Arkansas and Arkansas High-Performance Computing Center for providing the computational resources that significantly enhanced our research capabilities.

VII. REFERENCES CITED

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Balmori-Cedeño, J., Liu, J. T., Misk, E., Lillie, B. N., & Lumsden, J. S. (2019). Autophagy-related genes in rainbow trout *Oncorhynchus mykiss* (Walbaum) gill epithelial cells and their role in nutrient restriction. *Journal of Fish Diseases*, *42*(4), 549–558.

Bangs, M. R., Douglas, M. R., Thompson, P., & Douglas, M. E. (2017). Anthropogenic impacts facilitate native fish hybridization in the Bonneville Basin of Western North America. *Transactions of the American Fisheries Society*, *146*(1),16–21.

Bangs, M. R., Douglas, M. R., Mussmann, S. M., & Douglas, M. E. (2018). Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthyes: Catostomidae). *BMC Evolutionary Biology*, *18*(86), 1–16.

Bangs, M. R., Douglas, M. R., Brunner, P. C., and Douglas, M. E. (2020a). Reticulate evolution as a management challenge: Admixture in endemic fishes of Western North America. *Evolutionary Applications*, *13*(6), 1400–1419.

Bangs, M. R., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2020b). Gene flow and species delimitation in fishes of Western North America: Flannelmouth (*Catostomus latipinnis*) and Bluehead sucker (*C. pantosteus discobolus*). *Ecology and Evolution*, *10*(13), 6477–6493.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'donovan, C., & Apweiler, R. (2009). QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, *25*(22), 3045–3046.

Branson, B. A. (1967). Fishes of the Neosho River system in Oklahoma. *American Midland Naturalis*t, *78*(1), 126–154.

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

Brewer, S. K., Papoulias, D. M., & Rabeni, C. F. (2006). Spawning habitat associations and selection by fishes in a flow-regulated prairie river. *Transactions of the American Fisheries Society*, *135*(3), 763–778.

Bruckerhoff, L. A., Gido, K. B., Estey, M., & Moore, P. J. (2021). Disentangling effects of predators and landscape factors as drivers of stream fish community structure. *Freshwater Biology*, *66*(4), 656–668.

Buckley, B. A., & Somero, G. N. (2009). cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. *Polar Biology*, *32*(1), 403–415.

Bulger, A. G., & Edds, D. R. (2001). Population structure and habitat use in Neosho madtom (*Noturus placidus*). *The Southwestern Naturalist*, *46*(1), 8–15.

Bulger, A. G., Wildhaber, M., & Edds, D. (2002). Effects of photoperiod on behavior and courtship of the Neosho madtom (Noturus placidus). *Journal of Freshwater Ecology*, *17*(1), 141–150.

Calboli, F. C., Delahaut, V., Deflem, I., Hablützel, P. I., Hellemans, B., Kordas, A., … & Volckaert, F. A. (2021). Association between Chromosome 4 and mercury accumulation in muscle of the three-spined stickleback (*Gasterosteus aculeatus*). *Evolutionary Applications*, *14*(10), 2553–2567.

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855–867.

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309.

Chafin, T. K., B. T. Martin, M. R. Douglas, S. M. Mussmann, and M. E. Douglas (2017). FRAGMATIC: *in silico* locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources, 10*(3), 325–328.

Chafin, T. K., M. R. Douglas, B. T. Martin, and M. E. Douglas (2019). Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity*, *123*(6), 759–773.

Chafin, T. K., Mussmann, S. M., Douglas, M. R., & Douglas, M. E. (2023a). AUTOSTREAMTREE: Genomic variant data fitted to geospatial networks. *bioRxiv*, 2023–05.

Chafin, T. K., Mussmann, S. M., Douglas, M. R., & Douglas, M. E. (2023b). Quantifying isolation-by-resistance and connectivity in dendritic ecological networks. *bioRxiv*, 2023–07.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

Corl, A., Bi, K., Luke, C., Challa, A. S., Stern, A. J., Sinervo, B., & Nielsen, R. (2018). The genetic basis of adaptation following plastic changes in coloration in a novel environment. *Current Biology*, *28*(18), 2970–2977.

DeRaad, D. A. (2022). SNPfiltR: an R package for interactive and reproducible SNP filtering. *Molecular Ecology Resources*, 22(6), 2443–2453.

Dietrich, M. A., Hliwa, P., Adamek, M., Steinhagen, D., Karol, H., & Ciereszko, A. (2018). Acclimation to cold and warm temperatures is associated with differential expression of male carp blood proteins involved in acute phase and stress responses, and lipid metabolism. *Fish & Shellfish Immunology*, *76*(1), 305–315.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214.

Douglas, M. R., and Douglas, M. E. (2010). Molecular approaches to stream fish ecology, pp. 157–195 – *In: Community Ecology of Stream Fishes*. K. Gido and D. Jackson (eds.). *American Fisheries Society Symposium 73.*

Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594.

Eaton, D. A., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, *66*(3), 399–412.

Etchegaray, E., Dechaud, C., Barbier, J., Naville, M., & Volff, J. N. (2022). Diversity of Harbinger-like Transposons in Teleost Fish Genomes. *Animals*, *12*(11), 1429.

Excoffier, L., Smouse, P. E., & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491.

Fencl, J. S., Mather, M. E., Costigan, K. H., & Daniels, M. D. (2015). How big of an effect do small dams have? Using geomorphological footprints to quantify spatial impact of low-head dams and identify patterns of across-dam variation. *PLoS ONE*, *10*(11), e0141210.

Feng, J., Song, G., Wu, Y., Chen, X., Pang, J., Xu, Y., ... & Zhang, M. (2021). Plasmalogens improve swimming performance by modulating the expression of genes involved in amino acid and lipid metabolism, oxidative stress, and ferroptosis in an Alzheimer's disease zebrafish model. *Food & Function*, *12*(23), 12087–12097.

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, *11*(7), 1035–1052.

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233.

Fox, J. T., & Magoulick, D. D. (2019). Predicting hydrologic disturbance of streams using species occurrence data. *Science of the Total Environment*, *686*(10), 254–263.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496.

Fuselier, L., & Edds, D. (1994). Seasonal variation in habitat use by the Neosho madtom (Teleostei: Ictaluridae: *Noturus placidus*). *The Southwestern Naturalist*, *39*(3), 217–223.

Goudet, J., & Jombart, T. (2022). hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.5-11. https://CRAN.R-project.org/package=hierfstat

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., ... & Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.

Han, Z. Q., Guo, X. Y., Liu, Q., Liu, S. S., Zhang, Z. X., Xiao, S. J., & Gao, T. X. (2021). Whole-genome resequencing of Japanese whiting (*Sillago japonica*) provide insights into local adaptations. *Zoological Research*, *42*(5), 548–561.

Hoffmann, A. A., Miller, A. D., & Weeks, A. R. (2021). Genetic mixing for population management: From genetic rescue to provenancing. *Evolutionary Applications*, *14*(3), 634–652.

Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, *56*(3), 275–370.

Jamieson, I. G., & Allendorf, F. W. (2012). How does the 50/500 rule apply to MVPs? *Trends in Ecology & Evolution*, *27*(10), 578–584.

Ji, S., Sun, J., Bian, C., Huang, X., Chang, Z., Yang, M., ... & Ji, H. (2020). cAMP-dependent protein kinase A in grass carp *Ctenopharyngodon idella*: Molecular characterization, gene structure, tissue distribution and mRNA expression in endoplasmic reticulum stress-induced adipocyte lipolysis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *250*(1), 110479.

Jiang, T., Sun, J. L., Gu, Y., Yao, F. C., Liang, Y. S., Liu, Y. F., ... & Luo, J. (2023). Hypoxia alters glucose and lipid metabolisms in golden pompano (*Trachinotus blochii*). *Aquaculture*, *562*(1), 738747.

Johnson, R. L., Christian, A. D., Henry, S. D., & Barkley, S. W. (2009). Distribution, population characteristics, and physical habitat associations of black bass (*Micropterus*) in the lower Eleven Point River, Arkansas. *Southeastern Naturalist*, *8*(4), 653–670.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.

Jombart, T., & Collins, C. (2015). A tutorial for discriminant analysis of principal components (dapc) using adegenet 2.0.0. https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15.

Kalinowski, S. T., Meeuwig, M. H., Narum, S. R., & Taper, M. L. (2008). Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. *Canadian Journal of Fisheries and Aquatic* Sciences, *65*(12), 2752–2760.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*(1), e281.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788.

Kimura, M. (1991). The neutral theory of molecular evolution: a review of recent evidence. *The Japanese Journal of Genetics*, *66*(4), 367–386.

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, *14*(1), 1–14.

Krijthe. J. H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation. https://github.com/jkrijthe/Rtsne

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., & Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PLoS ONE*, *6*(4), e18007.

Li, S., Liu, X., Lin, T., Feng, G., Wang, X., & Zhang, D. (2023). Muscle fiber plasticity, stress physiology, and muscle transcriptome determine the inter-individual difference of swimming performance in the large yellow croaker (*Larimichthys crocea*). *Aquaculture*, *567*(1), 739247.

Li, W., Cerise, J. E., Yang, Y., & Han, H. (2017). Application of t-SNE to human genetic data. *Journal of Bioinformatics and Computational Biology*, *15*(4), 1750017.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., ... & Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 283.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., ... & Waldbieser, G. C. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature Communications*, *7*(1), 11757.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046.

Lou, F., Zhang, Y., Song, N., Ji, D., & Gao, T. (2020). Comprehensive transcriptome analysis reveals insights into phylogeny and positively selected genes of *Sillago* species. *Animals*, *10*(4), 633.

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*(13), 3223–3231.

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, *11*(1), 5–18.

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods in Ecology and Evolution*, *13*(10), 2150–2158.

Miller, J. M., Cullingham, C. I., & Peery, R. M. (2020). The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity*, *125*(5), 269–280.

Moss, R. (1981). Life history information for the Neosho madtom (*Noturus placidus*). Report to Kansas Fish & Game Commission, Pratt, KS

Mussmann, S. M., Douglas, M. R., Anthonysamy, W. J. B., Davis, M. A., Simpson, S. A., Louis, W., & Douglas, M. E. (2017). Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. *Royal Society Open Science*, *4*(2), 160736.

Mussmann, S. M., Douglas, M. R., Oakey, D. D., & Douglas, M. E. (2020a). Defining relictual biodiversity: Conservation units in speckled dace (Leuciscidae: Rhinichthys osculus) of the Greater Death Valley ecosystem. *Ecology and Evolution*, *10*(19), 10798-10817.

Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2020b). AdmixPipe: population analyses in Admixture for non-model organisms. *BMC Bioinformatics*, *21*(1), 1–9.

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

Ogawa, Y., & Corbo, J. C. (2021). Partitioning of gene expression among zebrafish photoreceptor subtypes. *Scientific Reports*, *11*(1), 17340.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, *26*(3), 419–420.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*(1), 2825–2830.

Peel, D., Waples, R. S., Macbeth, G. M., Do, C., & Ovenden, J. R. (2013). Accounting for missing data in the estimation of contemporary genetic effective population size (Ne). *Molecular Ecology Resources*, *13*(2), 243–253.

Pellissier, L. (2015). Stability and the competition-dispersal trade-off as drivers of speciation and biodiversity gradients. *Frontiers in Ecology and Evolution*, *3*(1), 52.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution, 31*(7), 1929–1936.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Prunier, JG, Poesy, C., Dubut, V., Veyssière, C., Loot, G., Poulet, N., & Blanchet, S. (2020). Quantifying the individual impact of artificial barriers in freshwaters: A standardized and absolute genetic index of fragmentation. *Evolutionary Applications*, *13*(10), 2566–2581.

Quigley, K. M., Bay, L. K., & van Oppen, M. J. (2020). Genome-wide SNP analysis reveals an increase in adaptive genetic variation through selective breeding of coral. *Molecular Ecology*, *29*(12), 2176–2188.

Reynkens (2018). rospca: Robust Sparse PCA using the ROSPCA Algorithm. R package version 1.0.4. https://CRAN.R-project.org/package=rospca

Rieman, B. E., & Allendorf, F. W. (2001). Effective population size and genetic conservation criteria for bull trout. *North American Journal of Fisheries Management*, *21*(4), 756–764.

Rimoldi, S., Benedito-Palos, L., Terova, G., & Pérez-Sánchez, J. (2016). Wide-targeted gene expression infers tissue-specific molecular signatures of lipid metabolism in fed and fasted fish. *Reviews in Fish Biology and Fisheries*, *26*(1), 93–108.

Russell, T., Cullingham, C., Kommadath, A., Stothard, P., Herbst, A., & Coltman, D. (2019). Development of a novel mule deer genomic assembly and species-diagnostic SNP panel for assessing introgression in mule deer, white-tailed deer, and their interspecific hybrids. *G3: Genes, Genomes, Genetics*, *9*(3), 911–919.

Schmidt, T. L., Jasper, M. E., Weeks, A. R., & Hoffmann, A. A. (2021). Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and Evolution, 12*(10), 1888–1898.

Schmidt, T. L., Thia, J. A., & Hoffmann, A. A. (2023). How Can Genomics Help or Hinder Wildlife Conservation? *Annual Review of Animal Biosciences*, *12*(1), 53.

Seufi, A. M., & Galal, F. H. (2020). Fast DNA Purification Methods: Comparative Study: DNA Purification. *WAS Science Nature*, *3*(1), 2766–7715.

Shafer, A. B. (2023). Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity*, *131*(1), 273–281.

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information theory broadens the spectrum of molecular ecology and evolution. *Trends in Ecology & Evolution*, *32*(12), 948–963.

Shumate, A., & Salzberg, S. L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics*, 37(12), 1639–1643.

Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, 9(1), 68.

Streelman, J. T., Peichel, C. L., & Parichy, D. M. (2007). Developmental genetics of adaptation in fishes: the case for novelty. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 655–681.

Suo, N., Zhou, Z. X., Xu, J., Cao, D. C., Wu, B. Y., Zhang, H. Y., ... & Zhao, Z. X. (2022). Transcriptome analysis reveals molecular underpinnings of common carp (*Cyprinus carpio*) under hypoxia stress. *Frontiers in Genetics*, *13*(1), 907944.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Tiemann, J. S., Gillette, D. P., Wildhaber, M. L., & Edds, D. R. (2004). Correlations among densities of stream fishes in the upper Neosho River, with focus on the federally threatened Neosho madtom *Noturus placidus*. *Transactions of the Kansas Academy of Science 107*(1), 17–24.

U.S. Fish and Wildlife Service (1991). Neosho madtom recovery plan. Denver, Colorado.

U.S. Fish and Wildlife Service (2013). Neosho Madtom (*Noturus placidus*) 5-year review: Summary and evaluation. Manhattan, Kansas.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*(11), 2579–2605.

Veilleux, H. D., Ryu, T., Donelson, J. M., Ravasi, T., & Munday, P. L. (2018). Molecular response to extreme summer temperatures differs between two genetically differentiated populations of a coral reef fish. *Frontiers in Marine Science*, *5*(1), 349.

Venkatachalam, A. B., Parmar, M. B., & Wright, J. M. (2017). Evolution of the duplicated intracellular lipid-binding protein genes of teleost fishes. *Molecular Genetics and Genomics*, *292*(1), 699–727.

Verity, R., & Nichols, R. A. (2016). Estimating the number of subpopulations (K) in structured populations. *Genetics*, *203*(4), 1827–1839.

Vieira, J. C. S., Braga, C. P., de Queiroz, J. V., Cavecci-Mendonça, B., de Oliveira, G., de Freitas, N. G., ... & de Magalhães Padilha, P. (2023). The effects of mercury exposure on Amazonian fishes: An investigation of potential biomarkers. *Chemosphere*, *316*(1), 137779.

Waples, R. S. (2023). Practical application of the linkage disequilibrium method for estimating contemporary effective population size: A review. *Molecular Ecology Resources*, In press.

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., ... & Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications*, *4*(6), 709–725.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Whitacre, L. K., Wildhaber, M. L., Johnson, G. S., Durbin, H. J., Rowan, T. N., Tribe, P., ... & Decker, J. E. (2022). Exploring genetic variation and population structure in a threatened species, *Noturus placidus*, with whole-genome sequence data. *G3: Genes, Genomes, Genetics*, *12*(4), jkac046.

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, *30*(1), 42–49.

Wildhaber, M. L. (2011). The Neosho Madtom and the multifaceted nature of population limiting factors. *American Fisheries Society Symposium*, *77*, 281–294.

Wildhaber, M. L., Allert, A. L. & Schmitt, C. J. (1999). Potential effects of interspecific competition on Neosho madtom (*Noturus placidus*) populations. *Journal of Freshwater Ecology*, *14*(1), 19–30.

Wildhaber, M. L., Tabor, V. M., Whitaker, J. A. E., Allert, A. L., Mulhern, D. W., Lamberson, P. J., & Powell, K. L. (2000). Ictalurid populations in relation to the presence of a main-stem reservoir in a Midwestern warmwater stream with emphasis on the threatened Neosho madtom. *Transactions of the American Fisheries Society*, 129(6), 1264–1280.

Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T., & Christie, M. R. (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, *27*(20), 4041–4051.

Xiao, W. (2015). The hypoxia signaling pathway and hypoxic adaptation in fishes. *Science China Life Sciences*, *58*(1), 148–155.

Yuan, M., Zhang, X., Louro, B., Li, X., Canário, A. V., & Lu, W. (2021). Transcriptomics reveals that the caudal neurosecretory system in the olive flounder (*Paralichthys olivaceus*) is more responsive in bold individuals and to chronic temperature change. *Aquaculture*, *544*(1), 737032.

Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2022). Riverscape community genomics: A comparative analytical approach to identify common drivers of spatial structure. *Molecular Ecology*, In press.

Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2023). A community genomics approach to natural hybridization. *Proceedings of the Royal Society B*, *290*(1999), 20230768.

Zhang, Q., Tyler-Smith, C., & Long, Q. (2015). An extended Tajima's D neutrality test incorporating SNP calling and imputation uncertainties. *Statistics and its Interface*, *8*(4), 447.

Zheng, K., & Li, T. (2021). Prediction of ATPase cation transporting 13A2 molecule in Petromyzon marinus and pan-cancer analysis into human tumors from an evolutionary perspective. *Immunogenetics*, *73*(4), 277–289.

VIII. GLOSSARY

Admixture: The result of interbreeding between two or more previously isolated populations within a species.

AIC (Akaike Information Criterion): A measure used in statistics to compare different statistical models. The AIC estimates each model's quality relative to the other models.

Allele: An allele is one of two or more versions of a gene or locus. An individual inherits two alleles for each gene, one from each parent. These alleles may be the same (homozygous) or different (heterozygous), and they can result in different traits or characteristics. Alleles differ by a single or multiple mutations (differences in the DNA sequence).

Autophagy: A cellular process for degrading and recycling proteins and other components within the cell.

Biallelic: Refers to a genetic locus with two different alleles or forms.

Ceramide: A family of lipid molecules composed of sphingosine and a fatty acid. Ceramides are found in high concentrations within the cell membrane and are involved in various cellular signaling pathways.

DAPC (Discriminant Analysis of Principal Components): A multivariate method that combines principal component analysis and discriminant analysis for identifying and describing clusters of genetically related individuals.

ddRAD (Double Digest Restriction-site Associated DNA): A method used in genomics to reduce genome complexity, making it easier to sequence specific regions of interest.

DNA (Deoxyribonucleic Acid): The molecule that carries genetic instructions for the growth, development, functioning, and reproduction of all known living organisms and many viruses.

***F*ST (Fixation Index):** A measure of population isolation due to genetic structure; the metric informs about levels of population connectivity.

GEA (Genotype Environment Association): The study of how genetic variation in a population correlates with environmental factors.

Gene Ontology: A framework for the model of biology that represents genes and gene product attributes across all species.

Hill numbers: A way of measuring biodiversity that considers the number of variants present (different alleles) and the relative abundance of each variant (frequency).

***K*-means clustering:** A vector quantization method, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Loci (plural of locus): Specific, fixed positions on in the genome where a particular gene is located on a chromosome. Often interchangeably used with the term 'genetic marker.'

***N*e (Effective Population Size):** The number of individuals in a population who contribute offspring to the next generation, which differs from the actual number of adult individuals (i.e., population size *N*).

Oligogenic (inheritance): A mode of inheritance for a trait influenced by a few genes, often with a major gene playing a predominant role.

Outbreeding Depression: Reduced biological fitness in a given population due to mating of unrelated individuals that produce offspring with reduced viability.

Paralog: A gene that has evolved by duplication within a genome and may evolve new functions.

RDA (Redundancy Analysis): A method of canonical analysis that seeks to explain the variance of one set of variables that is predictable from another.

ROBPCA (Robust Principal Component Analysis): A method of PCA resistant to the effect of outlying observations and skewed distributions.

Sequencing Depth: The number of times a locus is read during the sequencing process. Higher depth increases the accuracy of the sequencing data and thus confidence in the results.

Shannon Entropy: A measure of the uncertainty or randomness in a set of data, often used in ecological studies to measure biodiversity and population genetics to measure evolvability.

sNMF (Sparse Non-negative Matrix Factorization): A method used in bioinformatics for reducing the dimensionality of large genetic datasets, inferring structure, and assigning individual ancestry.

SNP (Single Nucleotide Polymorphism): A variation in a single nucleotide (mutation) that occurs at a specific position in the genome (site).

t-SNE (t-Distributed Stochastic Neighbor Embedding): A machine learning algorithm for dimensionality reduction that is particularly well suited for visualizing high-dimensional datasets, such as genomic sequence data, multi-locus data or SNP genotypes.

TL (Total Length): A morphometric measurement used in fisheries biology that refers to the total length of an organism. In fishes, TL is measured from the tip of the snout to the tip of the caudal fin.

Ubiquitin: A small regulatory protein found in almost all tissues of eukaryotic organisms, where it directs protein recycling.

IX. TABLES

Table 1: Overview of sampling sites across the Neosho River system. A total of 19 areas were sampled, with madtoms captured at 15 sites, yielding 249 tissue samples for DNA analysis. Sites are grouped by population inferred from analyses of genetic structure. Identifying information for each site where madtoms were collected included its river, state, county, specific location details, latitude, and longitude. Additionally, the number of individuals collected for each of four madtom species is provided.

Population	Site Name	River	State	County	Specific Location	Latitude	Longitude	Neosho Madtom	Freckled Madtom	Slender Madtom	Stonecat
CWR	CCFE	Cottonwood	KS	Chase	2.5 mi E Cottonwood Falls	38.37416	-96.49278	14	0	0	0
CWR	CCOT	Cottonwood	KS	Chase	Cottonwood Falls	38.37469	-96.53825	2	0	0	6
CWR	CESG	Cottonwood	KS	Chase	Emporia, Soden's Grove	38.38527	-96.18112	16	0	0	1
NRU	NAME	Neosho	KS	Lyon	Near Americus	38.50611	-96.31167	16	9	0	8
NRU	NDUN	Neosho	KS	Lyon	Near Dunlap	38.55083	-96.34723	16	5	0	1
NCC	NECW	Neosho	KS	Lyon	Emporia, Campus Woods	38.42666	-96.17222	16	0	0	1
NCC	NRAP	Neosho	KS	Lyon	Neosho Rapids	38.36916	-96.00918	16	0	0	0
NRM	NIOL	Neosho	KS	Allen	Between Neosho Falls & Iola	37.97889	-95.50028	16	0	0	1
NRM	NROY	Neosho	KS	Coffey	Near LeRoy	38.06388	-95.65361	16	1	0	5
NRL	NLAB	Neosho	KS	Labette	US 160 bridge	37.16638	-95.06223	16	2	0	1
NRL	NMIA	Neosho	OK	Ottawa	NW Miami, Stepp's Ford bridge	36.92777	-94.96056	16	0	0	3
NRL	NSTP	Neosho	KS	Neosho	Near St. Paul	37.44889	-95.14361	4	0	0	4
SPR	SMCP	Spring	OK	Ottawa	10 mi E Miami, former Centennial Park	36.96167	-94.72163	12	1	0	0
SPR	SWAC	Spring	MO	Jasper	2 mi E Waco, Maple Road	37.25087	-94.57384	16	0	2	0
-	NOSW	Neosho	KS	Labette	Oswego	37.17666	-95.10225	0	1	0	5

CWR=Cottonwood River
NCC=Neosho-Cottonwood Confluence
NRL=Lower Neosho River
NRM=Middle Neosho River
NRU=Upper Neosho River
SPR=Spring River

Table 2: Population connectivity evaluated as pairwise unbiased F_{ST} (G''_{ST}) between six populations of Neosho Madtom. Estimates are based on SNP data of 178 individuals genotyped across 2,725 loci. Values range from 0 (total connectivity) to 1 (total isolation). All estimates significantly differed from zero based on bootstrap re-sampling to generate confidence intervals. Note: NRU and SPR show much higher isolation ($\bar{x}$ =0.12) than the others ($\bar{x}$ =0.032).

	CWR	NCC	NRL	NRM	NRU
NCC	0.020				
NRL	0.032	0.020			
NRM	0.053	0.037	0.031		
NRU	0.119	0.090	0.098	0.112	
SPR	0.124	0.114	0.097	0.102	0.184

CWR=Cottonwood River
NCC=Neosho-Cottonwood Confluence
NRL=Lower Neosho River
NRM=Middle Neosho River
NRU=Upper Neosho River
SPR=Spring River

Table 3: Population uniqueness evaluated as pairwise proportion of alleles (variants) not shared (β-Differentiation = ${}^{0}H_{\beta}$) among six Neosho Madtom populations. Estimates are based on SNP data of 178 individuals genotyped across 2,725 loci. Note: values are greater than F_{ST} estimates, because rare alleles are weighted in the calculation. In constrast, F_{ST} values based on heterozygosity, which is affected more by common than rare alleles.

	CWR	NCC	NRL	NRM	NRU
NCC	0.079				
NRL	0.096	0.093			
NRM	0.106	0.096	0.087		
NRU	0.141	0.138	0.146	0.144	
SPR	0.149	0.150	0.147	0.142	0.152

CWR=Cottonwood River
NCC=Neosho-Cottonwood Confluence
NRL=Lower Neosho River
NRM=Middle Neosho River
NRU=Upper Neosho River
SPR=Spring River

Table 4: Local genetic diversity estimates for six Neosho Madtom populations based on SNP data genotyped across N=178 individuals collected from 14 sampling sites. All measures were calculated based on N=2,725 genetic loci, except genome-wide heterozygosity (H_{GW}) based on the total unfiltered genomic alignment.

Pop	PA	AR	H_E	H_O	H_{GW}	F_{IS}	ENT
CWR	21	1.72	0.206	0.206	0.00065	-0.006	0.321
NCC	19	1.76	0.212	0.213	0.00067	-0.003	0.333
NRL	49	1.77	0.213	0.209	0.00067	0.015	0.336
NRM	20	1.54	0.291	0.209	0.00068	-0.010	0.290
NRU	25	1.61	0.187	0.189	0.00061	-0.008	0.287
SPR	77	1.50	0.255	0.193	0.00066	-0.005	0.273

PA: Number of private alleles unique to each population
AR: Mean allelic richness per locus (at most 2 for biallelic SNPs)
H_E: Expected heterozygosity across polymorphic sites
H_O: Observed heterozygosity across polymorphic sites
H_{GW}: Genome-wide observed heterozygosity based on the unfiltered alignment (polymorphic and non-polymorphic sites)
F_{IS}: Mean inbreeding coefficient ($F_{IS} = 0$ is neutral HWE; $F_{IS} < 0$ indicates an excess of heterozygotes due to outbreeding; $F_{IS} > 0$ indicates a deficit of heterozygotes due to inbreeding)
ENT: Shannon Information (entropy), a measure of evolvability

CWR=Cottonwood River
NCC=Neosho-Cottonwood Confluence
NRL=Lower Neosho River
NRM=Middle Neosho River
NRU=Upper Neosho River
SPR=Spring River

Table 5: Genetic viability and historic population demographic estimates for six Neosho Madtom populations. Estimates are based on SNP data of 178 individuals genotyped across 2,725 loci. The effective population size (N_e) was estimated using the linkage disequilibrium method NEESTIMATOR. Mean N/Locus is the mean number of individuals across the loci used to estimate the effective population sizes (N_e). Tajima's *D* is a test of population neutrality, and values greater than 0 are consistent with a population bottleneck.

Population	Mean N/locus	N_e	Tajima's D
CWR	29.5	2326	0.74
NCC	29.5	279	0.67
NRL	29.1	1622	0.45
NRM	29.6	1138	0.69
NRU	29.3	349	1.17
SPR	24.2	404	0.83
	Sum =	6118	

CWR=Cottonwood River
NCC=Neosho-Cottonwood Confluence
NRL=Neosho River Lower
NRM=Neosho River Middle
NRU=Neosho River Upper
SPR=Spring River

Table 6: Thirty genes were associated with outlier adaptive loci across Neosho Madtom populations. The gene ontologies related to biological processes are shown here. These ontologies are grouped for inference; each gene may be associated with multiple processes. Note: two genes not shown here had no matching association to a known biological process (cpped1, harbi1).

Gene	Associated Biological Process (Gene Ontology)
nptn	anatomical development (cell adhesion/organization)
aim1	anatomical development (lens development in eye)
usp28	anatomical development (nervous system morphogenesis)
atp13a2	anatomical development (cerebellum)
atg13	anatomical development (chondrocytes/cartilage)
efnb2	anatomical development (nervous system)
eftud2	anatomical development (neuron/eye/head development)
atp13a2	behavior (locomotion)
bmx	response to stimulus
efnb2	response to stimulus
htr1e	response to stimulus
nkd2	response to stimulus
rac3	response to stimulus
fgb	response to stimulus/stress
usp28	response to stimulus/stress
rnf103	response to stimulus/stress/chemical
fmn2	response to stimulus/stress/chemical/hypoxia
atp13a2	response to stimulus/stress/chemical
bmx	adaptive immune response
fgb	blood coagulation/platelet activation/fibrin
gramd4	cellular processes: apoptosis/cell death
atg13	metabolism (autophagy/cell recycling)
atp13a2	metabolism (autophagy/cell recycling)
ptpre	metabolism (dephosphorylation)
bmx	metabolism (organic, nitrogen)
eftud2	metabolism (organic, nitrogen)
fgb	metabolism (organic, nitrogen)
fmn2	metabolism (organic, nitrogen)
nkd2	metabolism (organic, nitrogen)
pwwp2a	metabolism (organic, nitrogen)
samd11	metabolism (organic, nitrogen)
cerk	metabolism (organic, nitrogen, ceramide)
cers1	metabolism (organic, nitrogen, ceramide)
prkar1a	metabolism (organic, nitrogen, protein, phosphorylation)
rnf103	metabolism (organic, nitrogen, protein, ubiquitination)
ubap1l	metabolism (organic, nitrogen, ubiquitin)
usp28	metabolism (organic, nitrogen, ubiquitin)
lpcat1	metabolism (organics, lipids)
fmn2	DNA damage response / meiotic spindle localization
usp28	DNA damage signaling
pwwp2a	DNA transcription
samd11	DNA transcription
immt	mitochondrial development (cristae formation)

X. FIGURES

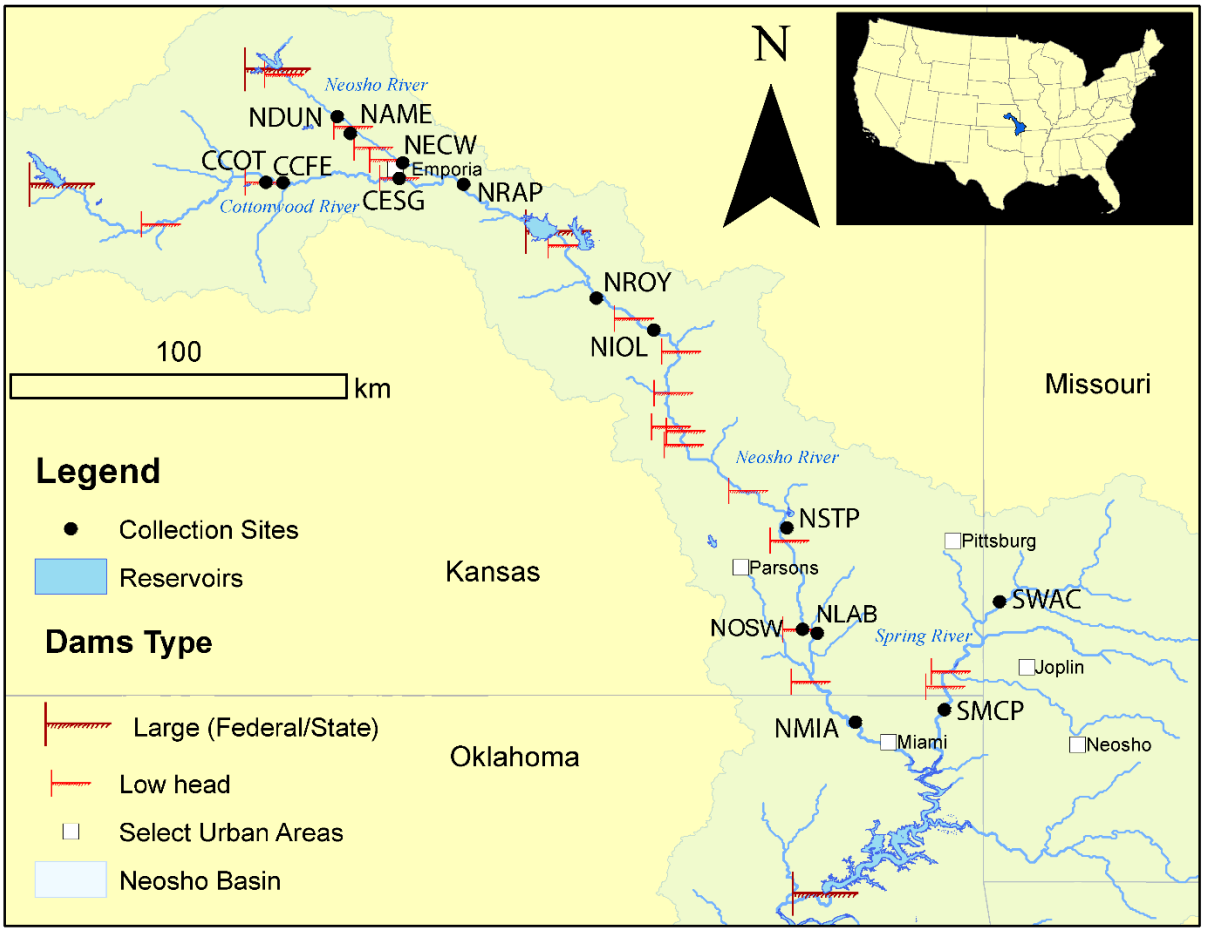

Figure 1: Map of the Neosho River system and major tributaries (Cottonwood and Spring rivers) showing the sampling locations (N=15) where madtoms were collected (black circles labeled with site ID). Neosho Madtom was not detected at site NOSW, but sample of two other *Notorus* species were collected. The light blue highlighted areas indicates Neosho system watershed. The system is highly regulated, and dams denoted with red symbols. John Redmond Reservoir, downstream of site NRAP, demarcats the separation the northern and southern sites on the Neosho River, whereas Grand Lake o' the Cherokees separates the Spring River sites from the remainder. Detailed information on sites is provided in Table 1.

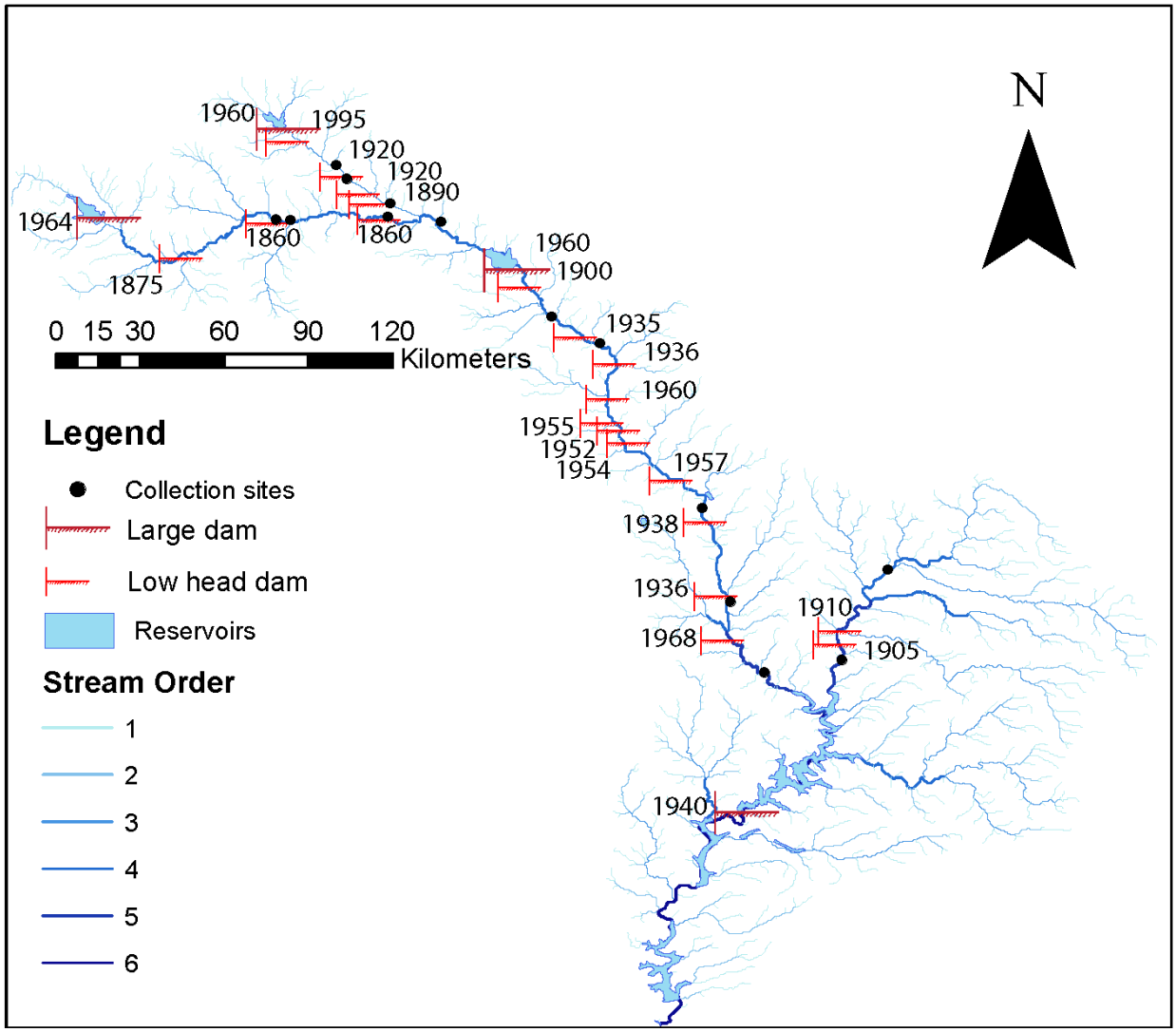

Figure 2: Map of the Neosho River system showing dams (*N*=24) with their approximate earliest construction dates. Sites where Neosho Madtom (NMT) was collected (*N*=14) are indicated with black circles. Note: Only *N*=18 of the dams are positioned between NMT sampling locations; the other *N*=6 are upstream (see dams on upper Cottonwood and Neosho rivers) or downstream (Grand Lake) of all sites. Figure 1 and Table 1 provide details on sampling locations. See Supplement Table S1 for more details on dams.

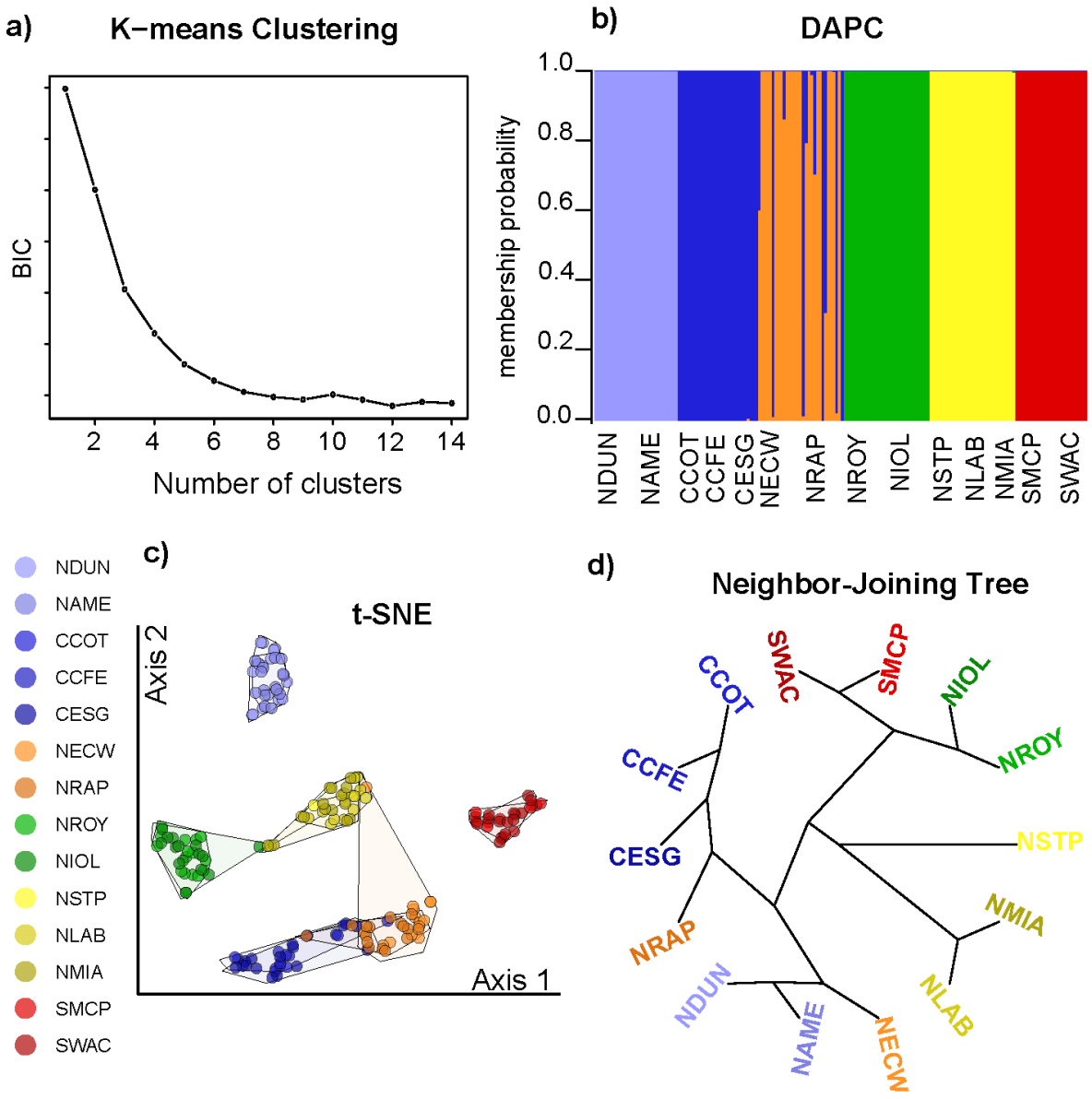

Figure 3: Results of cluster analyses that grouped Neosho Madtom into six genetic populations. Several clustering approaches were employed to analyze SNP date derived from *N*=178 individuals genotyped across *N*=2,725 loci. (a) *K*-means clustering shows Bayesian Information Criterion (BIC) declining with increasing clusters (=better fit). (b) Membership probabilities to clusters based on Discriminant Analysis of Principal Components (DAPC) show clearly defined clusters and some possible migrants. (c) Clustering individuals based on their loci using dimension reduction technique t-distributed stochastic neighbor embedding (t-SNE) with sites colored by inferred populations. (d) Neighbor-joining tree of sites based on genetic divergence (F_{ST}). Site acronyms defined in Table 1 and geographic distribution of samples showing in Figure 1.

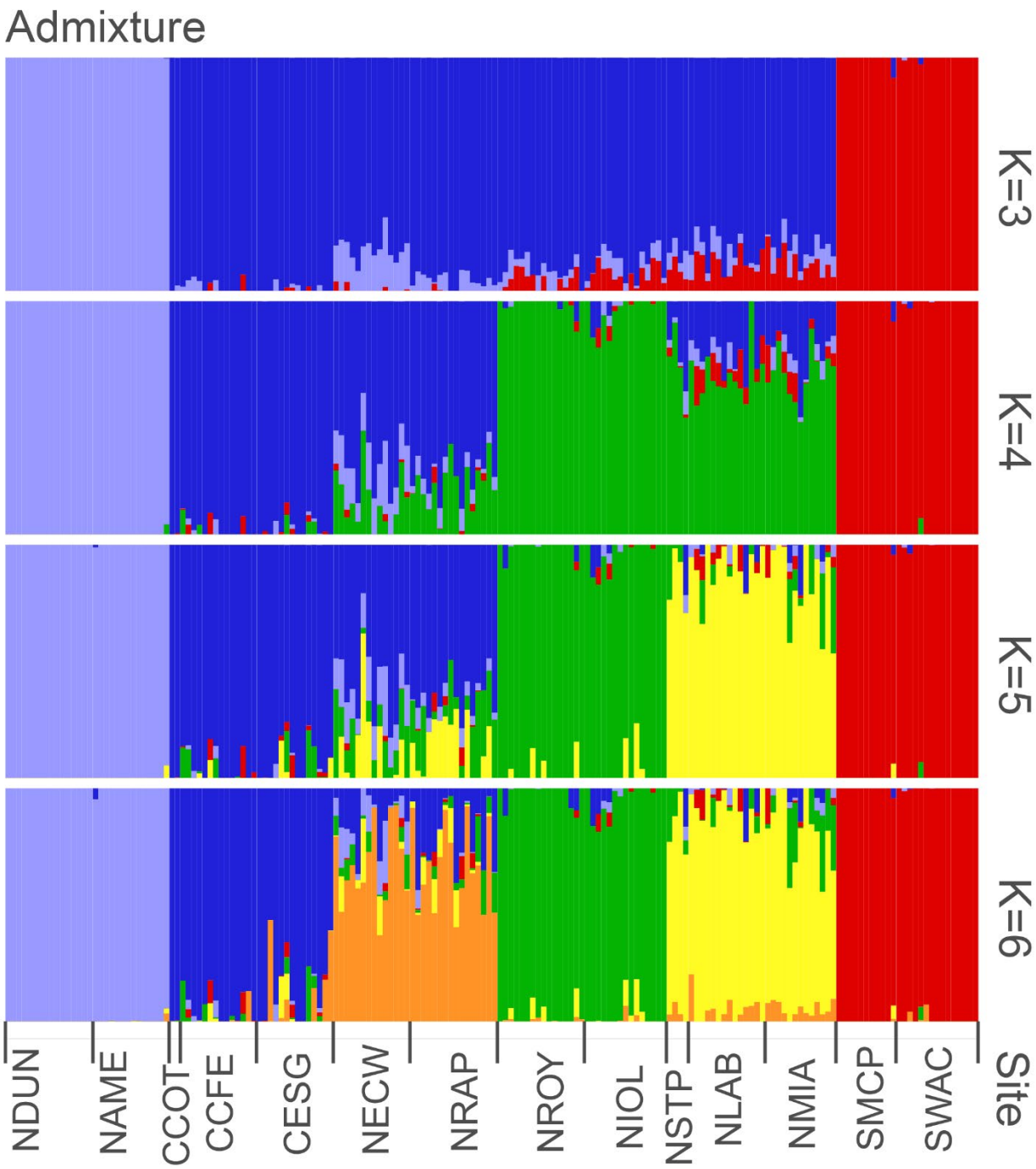

Figure 4: Evaluation of genetic population structure using ADMIXTURE analysis of Neosho Madtom individuals (*N*=178) collected at 14 sites and genotyped at *N*=2,725 loci. Individuals are grouped by sampling sites (bottom) and represented as vertical bars in the plot. The color of an individual's vertical bar is determined by the proportion of its ancestry in each cluster (*K*). Sites are ordered geographically, from upstream to downstream. Four models of hierarchical population structure are shown, from three (*K*=3) to six (*K*=6) discrete populations. Mixed ancestry denotes gene flow between groups (admixture). Site acronyms defined in Table 1.

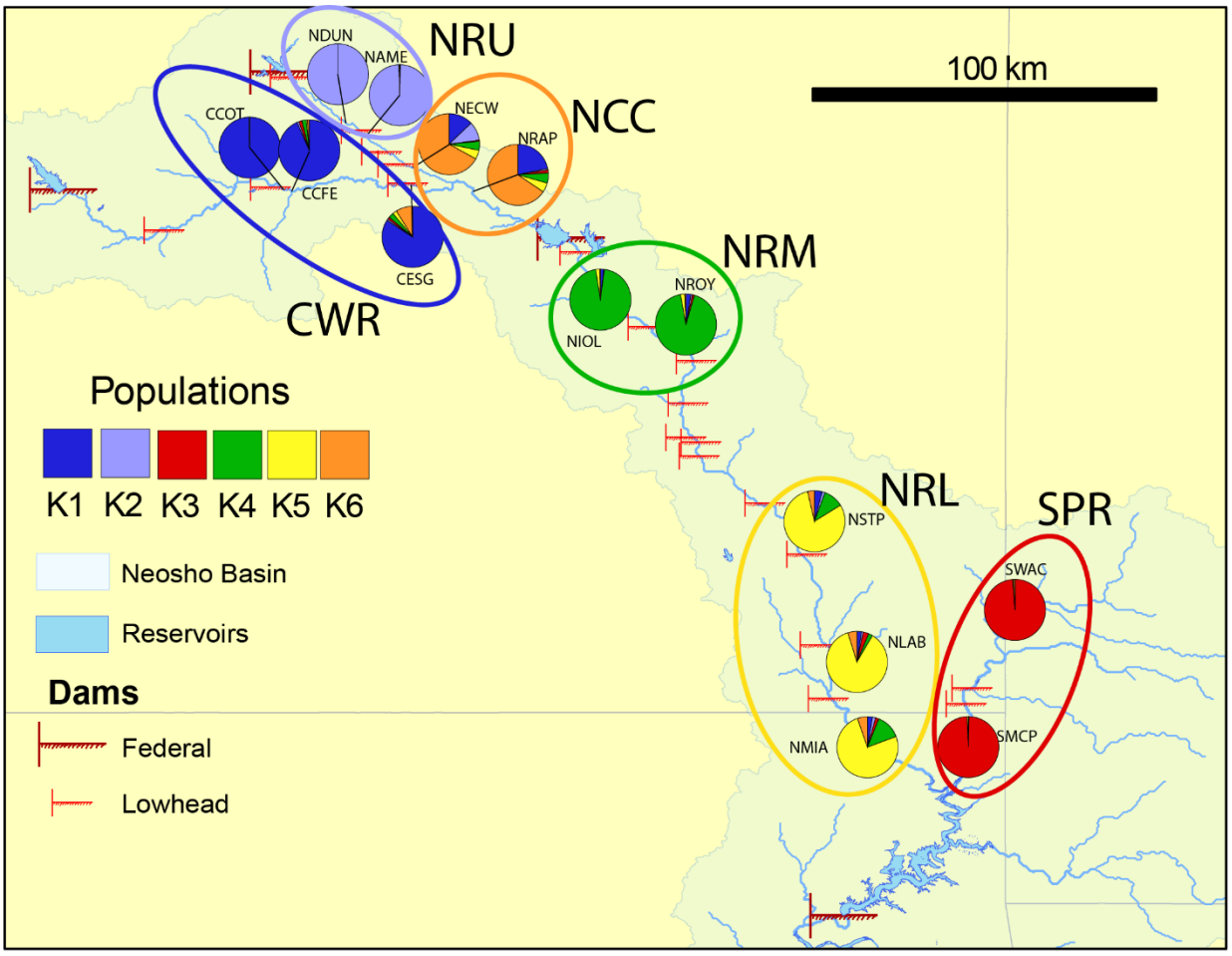

Figure 5: Potential Management Units based on the population structure of Neosho Madtom. Distinct populations were inferred from clustering approaches SNP data of *N*=178 individuals collected across 14 sites and genotyped *N*=2,725 polymorphic loci. Genetic composition at each sampling site is pictured as pie chart representing population ancestry proportions based on ADMIXTURE analysis (*K*=6). Based on these ancestry results, the sites can be grouped into six potential genetic populations (gene pools): Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on sites and collections are listed in Table 1.

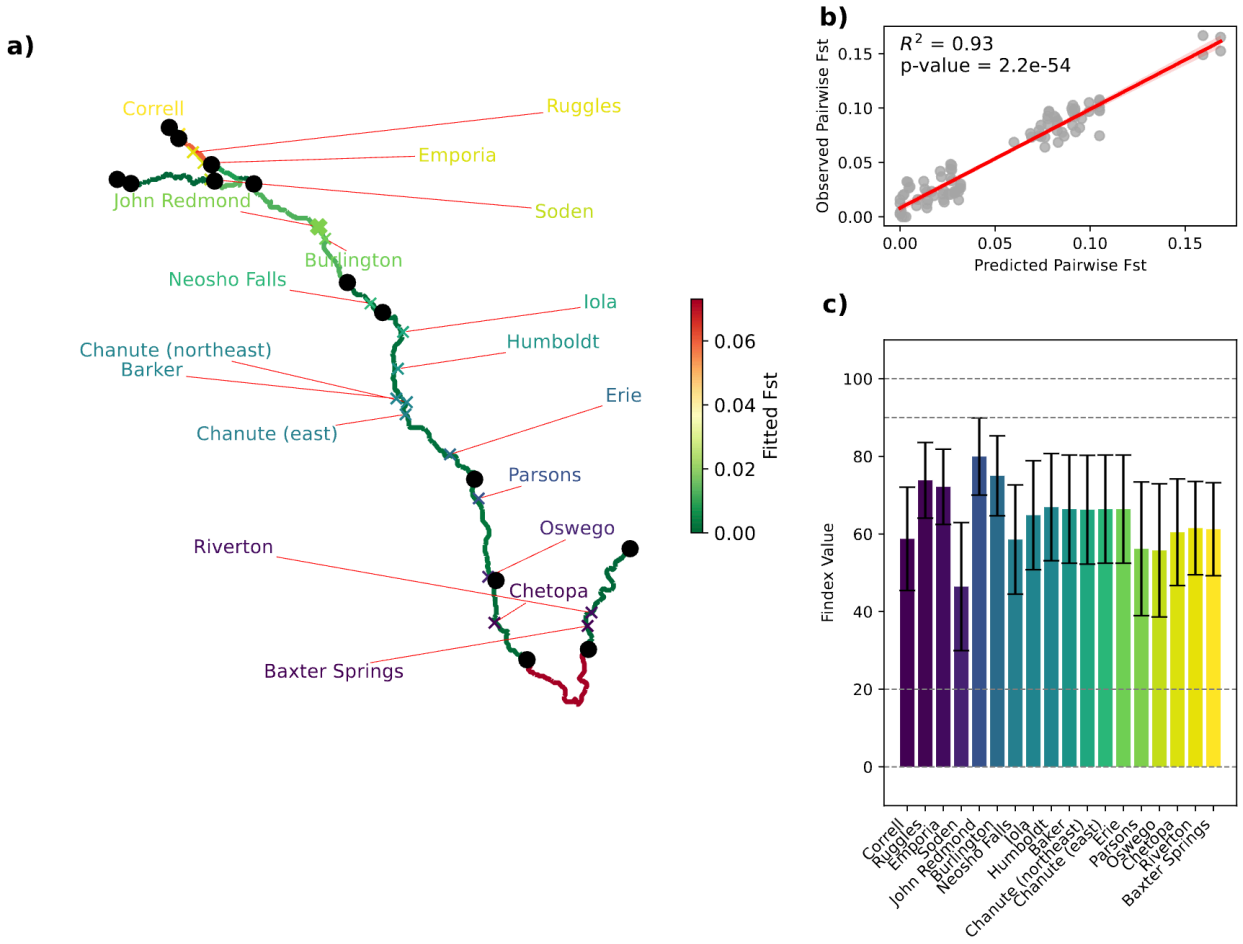

Figure 6: Riverscape analysis of population connectivity relative to dam locations (i.e., dispersal barriers) for Neosho Madtom. Results are inferred from analysis of SNP data for N=178 individuals collected across 14 sites and genotyped across N=2,725 loci. (a) Colors reflect level of potential isolation for each river segment, based on fitted genetic divergence (F_{ST}) per river reach segment using the STREAMTREE model, with estimates ranging from low (green) to high genetic divergence (red). Sample sites are annotated as black circles, with dams (colored by latitude, as in c) as X's and labeled with names. John Redmond, the only sizeable federal dam between sampling sites, is indicated in bold. Note: Only N=18 dams are positioned between NMT sampling locations and were analyzed here. (b) Fit of STREAMTREE-optimized pairwise genetic divergence (F_{ST}) compared with the observed values. (c) Dam effects on population connectivity (gene flow): Normalized index of genetic fragmentation (F_{INDEX}) where 0=no barrier effect and 100=complete reduction in gene flow (95% confidence intervals are indicated).

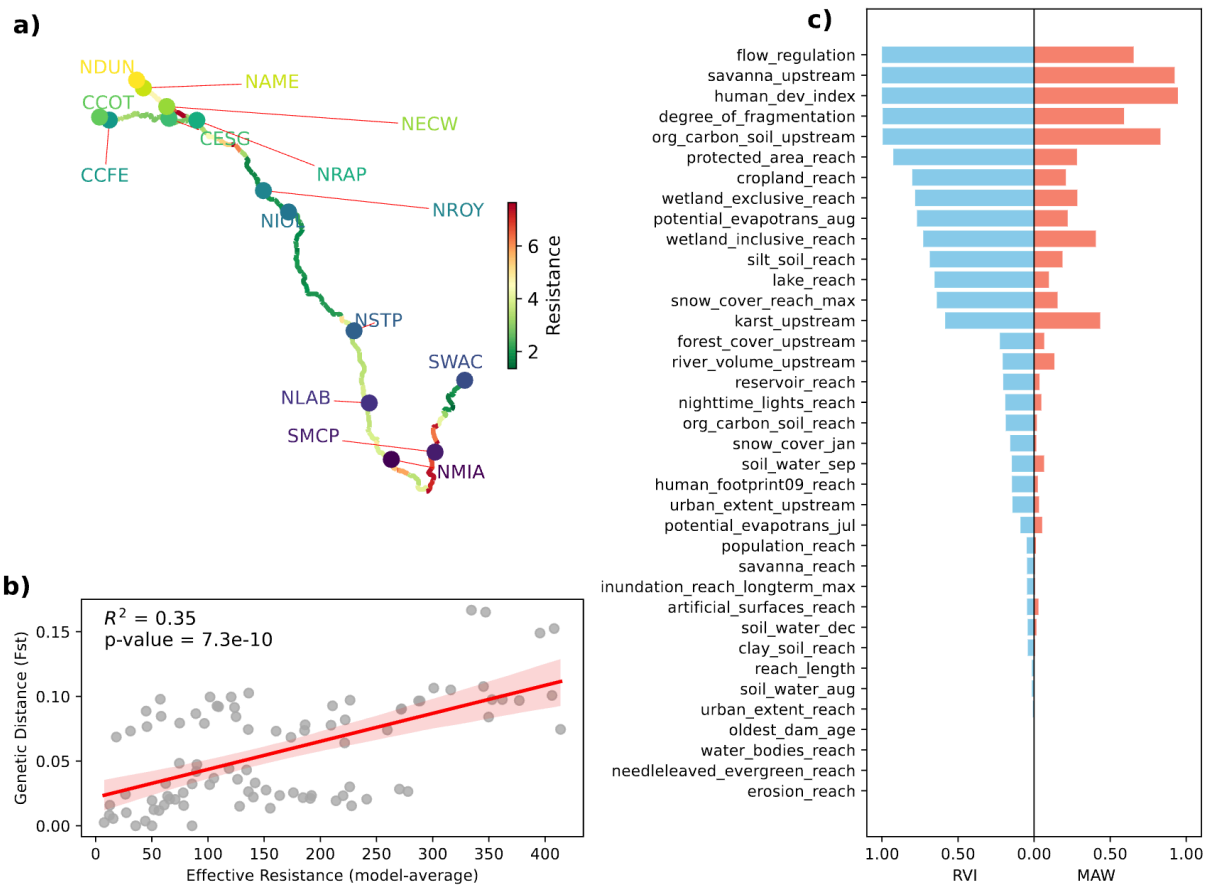

Figure 7: Effective resistance network modeling for Neosho Madtom across the range. Estimates were inferred from SNP data of *N*=178 individuals collected across 14 sites and genotyped across *N*=2,725 loci. (a) Model-averaged effective resistance attributed to each stream reach, ranging from low (=0) to high (=10) resistance to individual movement. (b) Linear regression of pairwise effective resistance, expressed as the sum of model-averaged reach-wise values along the least-cost distance path between samples, compared to their observed genetic divergence (F_{ST}). (c) Relative variable importance (RVI; blue) and model-averaged weights (MAW; red) for environmental features used in ResistNet. Details on sampling sites provided in Table 1.

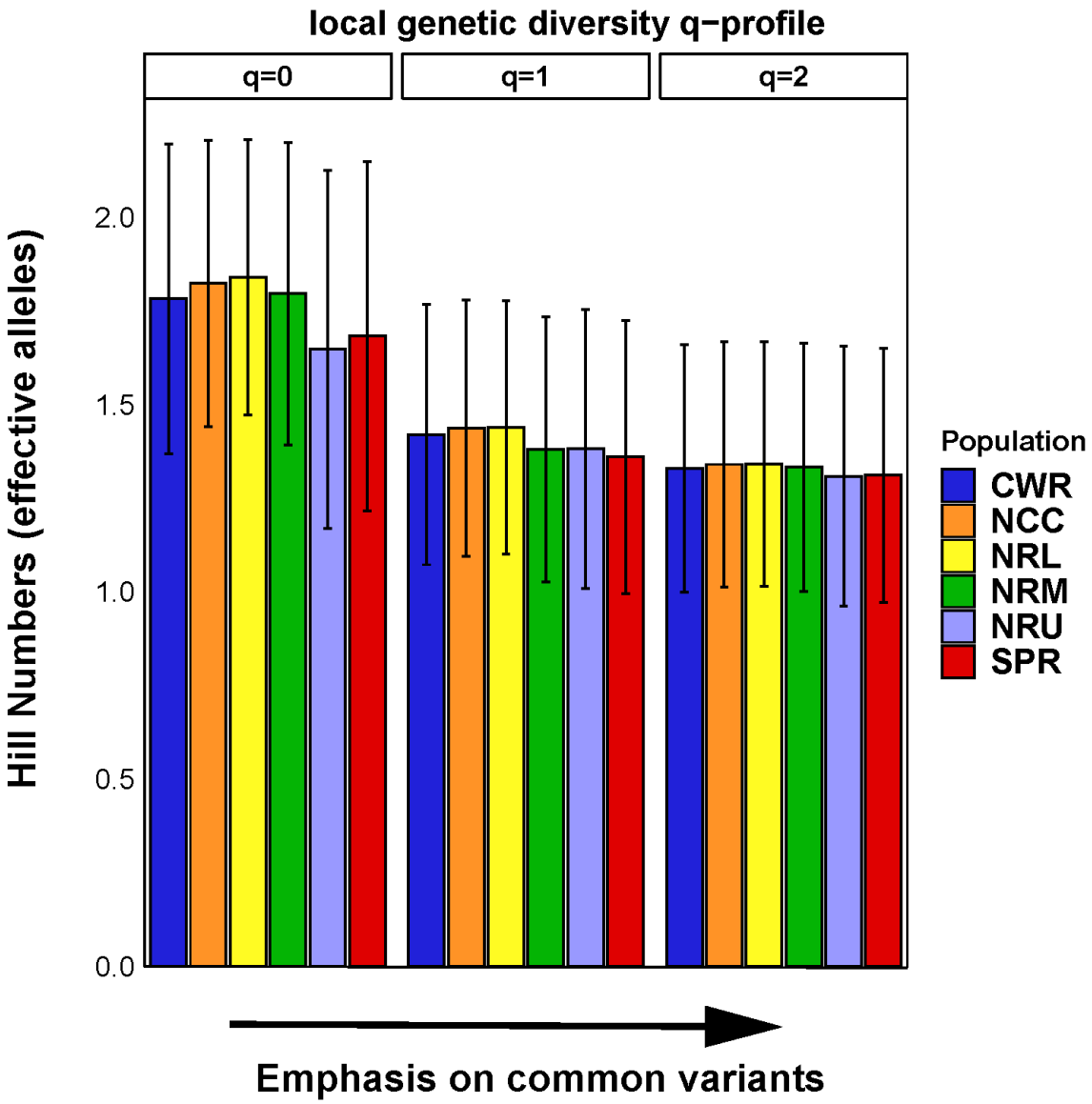

Figure 8: Three metrics of local, within-population genetic diversity (α diversity) for each of the six genetic populations of Neosho Madtom (i.e., Management Units). Estimates are based on anlysis of SNP data from *N*=178 individuals collected from 14 sampling sites and genotyped across *N*=2,725 loci. Metrics include allelic richness (q=0), Shannon information (q=1), and heterozygosity (q=2). Each metric differs in how it is reflects rare *vs.* common genetic variants. For better comparison, the three metrics are standardized to the effective number of alleles (Hill numbers), a standard scale representing the number of equally frequent alleles needed to yield the observed diversity value. Solid bars represent the mean value, with standard deviations indicated as lines. Colors reflect the six genetic populations, considered potential Management Units: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on populations are provided in Table 1, and geographic locations depicted in Figure 5.

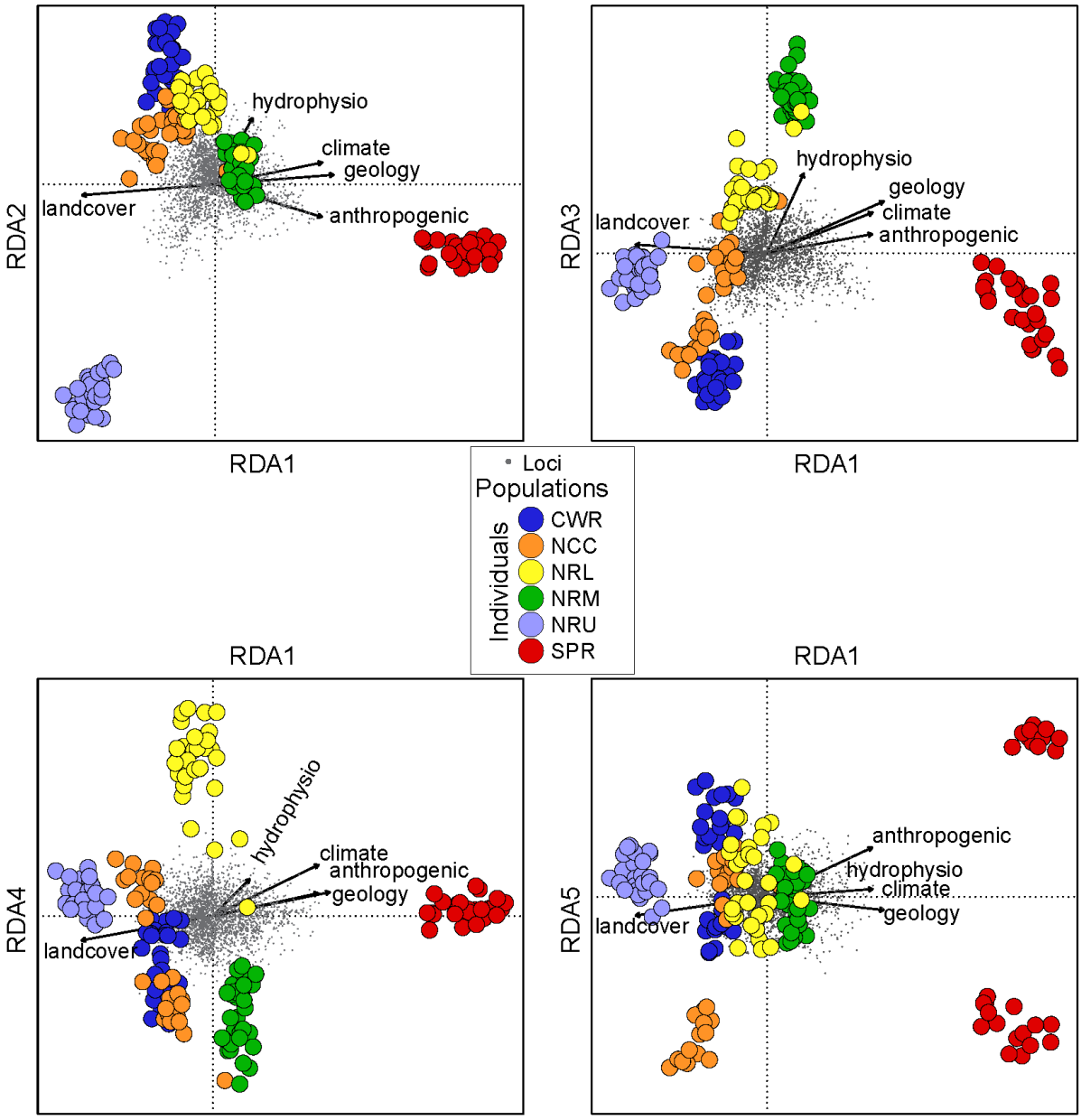

Figure 9: Assessment of local adaptation of Neosho Madtom along environmental gradients. Evaluations are based on genotype-environment association (GEA) via redundancy analysis (RDA). The RDA modeled the relationship between individual genetic variation and five environmental factors (five principal components representing landcover, hydro-physiographic, geology, climate, and anthropogenic factors). This model produced five RDA canonical axes on which loci and environmental factors have correlations (loadings) and can be used to ordinate individuals in the "RDA space." Colors reflect six genetic populations: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on populations provided in Table 1 and geographic locations depicted in Figure 5.

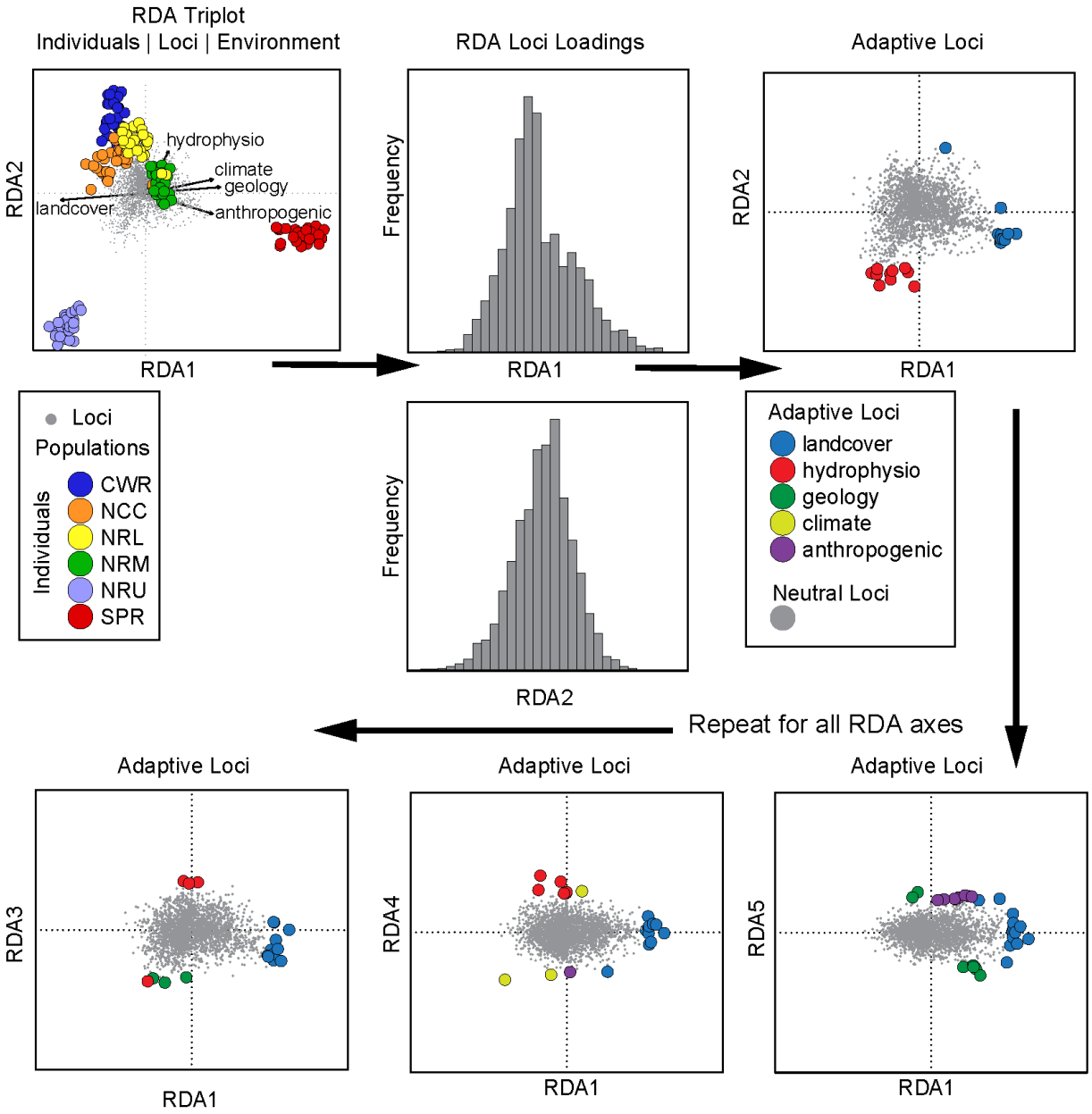

Figure 10: Local adaptation of Neosho Madtom assessed using genotype-environment association via redundancy analysis modeling the relationship between individual genetic variation and environmental factors. This model produced five canonical axes on which loci and environmental factors have correlations. The distribution of loci loadings on each RDA axis was assessed (top middle), and locally adapted loci were the outliers (±3 standard deviations from the mean). The rest are assumed neutral. The remaining plots (top right and bottom) show these locally adapted outlier loci. For interpretation, loci are colored based on the environmental factor they are most correlated with. Six genetic populations: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on populations are provided in Table 1, and geographic locations depicted in Figure 5.

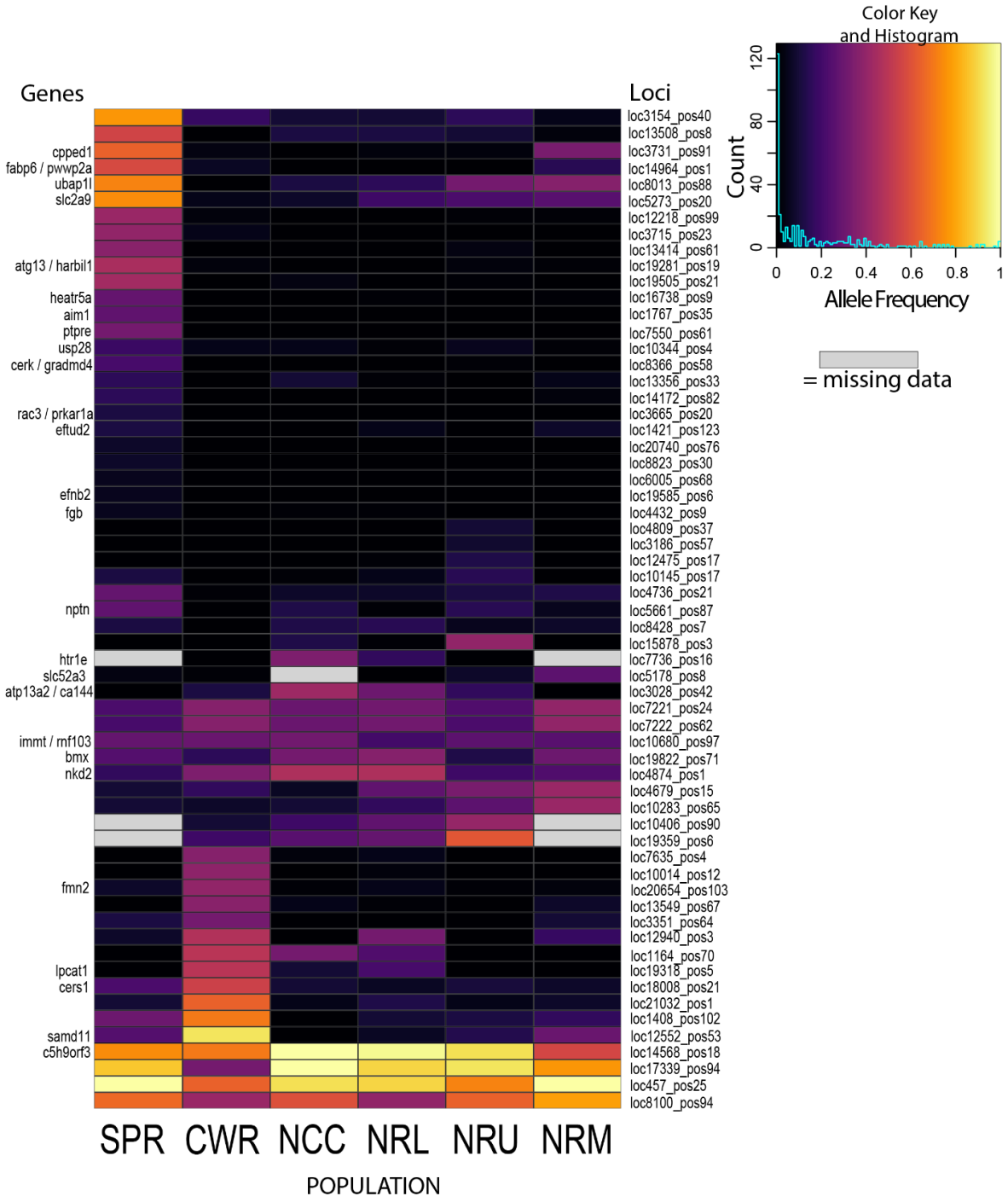

Figure 11: Frequencies of the adaptive loci (*N*=61) for each of the six populations of Neosho Madtom. Color key (upper right): darker colors correspond to lower frequencies, whereas bright colors reflect high frequencies; blue line shows distribution of allele frequencies. Missing data are denoted with gray and indicate a particular locus was not recovered for a population. Locus IDs (right) and related gene names (left) annotate plot rows. Note: a heat map of the alternate allele would produce an inverse image. Six genetic populations: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on populations are provided in Table 1 and geographic distribution is depicted in Figure 5.

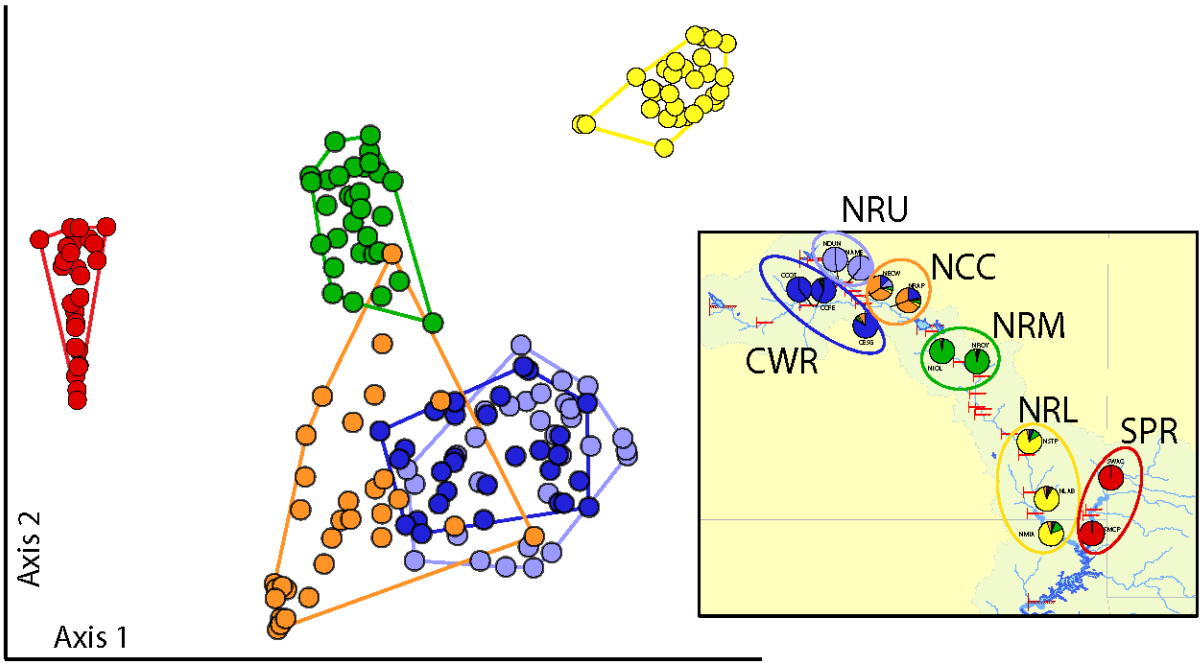

Figure 12: Population structure of Neosho Madtom based on locally adapted genetic loci. If populations have been adapting to the specifics of their local environment, we expect to see different groupings in the plot of genetic variation produced by t-distributed stochastic neighbor embedding (t-SNE). The northern populations (CWR, NRU, NCC) show much overlap in the plot, reflecting similarity in adaptation, whereas there is little (NRM) or no (NRL, SPR) overlap among the southern sites, pointing to possible differences in local adaptation along the longitudinal gradient of the river system. The southern sites appear distinct from each other, indicating unique local adaptation. Colors reflect six genetic populations: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR). Details on populations are listed in Table 1.

XI. APPENDICES

Appendix 1: Laboratory Methods

Genetic DNA Extraction

DNA was extracted from all tissues following the QIAamp Fast DNA Tissue Kit protocol (QIAGEN© Corporation, Maryland, USA). Tissue was homogenized using steel beed mechanical disruption for 1 minute at 24 kHz using a TissueLyser II (QIAGEN© Corporation, Maryland, USA). Samples were then spun down and placed on a thermomixer at 56° C for 10 minutes at 1000 rpm. Finally, samples rested on a thermo-block at 56° C for 120 minutes before proceeding with the remainder of the QIAamp protocol.

The concentration of genetic DNA from each sample was quantified with a Qubit 2.0 Fluorometer (Invitrogen, Inc., Maryland, USA) following the manufacturer's protocol. To ascertain the presence of high-quality genetic DNA (i.e., molecular weight >10kb), a 5µl aliquot of the DNA extract was separated on a 2% agarose gel and visualized using GelGreen on a blue light transilluminator (Gel Doc™ EZ Imager; Bio-Rad). Large DNA fragments migrate more slowly than small fragments in a gel, and high-quality DNA forms a distinct band, whereas degraded or fragmented DNA forms a 'smear' of small fragments.

Genetic Library Preparation and Sequencing

For single nucleotide polymorphism (SNP) loci data generation from each madtom tissue sample, we devised protocols employing a genetic technique termed double digest restriction-site associated DNA sequencing (ddRAD; Peterson et al., 2012).

Standardized DNA concentrations (500 ng) were digested at 37°C with high-fidelity restriction enzymes *MspI* (5'-CCGG-3') and *PstI* (5'-CTGCAG-3') (New England Biosciences), bead-purified (Ampure XP; Beckman-Coulter Inc.), standardized to 50 ng, and then ligated with custom adapters containing in-line identifying barcodes (T4 Ligase; New England Biosciences).

Individual samples were pooled in sets of 41 to 42 and size-selected from 326 to 426 bp, including adapter length (Pippin Prep; Sage Sciences). Illumina adapters and i7 index were added via 12-cycle PCR with Phusion high-fidelity DNA polymerase (New England Biosciences).

A set of six libraries (3x41 + 3x42 =249 individuals) were pooled into a single lane and sequenced single-end on the Illumina NovaSeq 6000 platform (SP 1x118 bp; Genomic & Cell Characterization Core Facility; University of Oregon/ Eugene). Quality control checks, including fragment analysis and quantitative real-time PCR, were performed at the core facility before sequencing.

Appendix 2: Bioinformatics

Data Processing and Assembly

Quality control checks were performed on the raw sequence data using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw sequence files were then de-multiplexed, and reads were assigned to individuals based on their unique DNA barcodes using IPYRAD V.0.9.87 (Eaton and Overcast, 2020), allowing no tolerance for mismatched barcodes. We constructed two alignments for different purposes. We constructed a multispecies alignment (four madtom species) to screen for hybrids and a single-species Neosho Madtom alignment for the rest of the analyses.

For both alignments, individual sequence files were processed with IPYRAD (Eaton and Overcast, 2020) using a reference-guided assembly that first mapped reads to the Neosho Madtom genome (Whitacre et al., 2022). Adapters and primers were removed, and reads with >5 low-quality bases (Phred<20) were discarded. Clusters were removed via conditional criteria to ensure high-quality data: <20x and >1000x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of nucleotides polymorphic; >8 indels present; or presence in <5 individuals. Putative paralogs were removed if clusters displayed >2 alleles per site in consensus sequence or excessive heterozygosity (>5% consensus bases or >50% heterozygosity/site).

Data Filtering

Some additional filtering was performed on the VCF files produced from the multispecies and single-species alignments. For the multispecies alignment, loci were removed for an individual if depth < 20x and > 200x. Loci were further filtered if they were not biallelic (i.e., invariant or more than two alleles), had >75% missing data, and had a minor allele count <2. Finally, one SNP per locus was selected randomly using a thinning distance of 500 bp. Individuals were removed if they contained >70% missing data. The single species (NMT) alignment was filtered in the same way, except we only allowed up to 50% missing data per loci and up to 50% missing data per individual. A combination of R packages was used for visualization, summary, and filtering, including VCFR V.1.13.0 (Knaus and Grünwald, 2017), SNPFILTR V.1.0.0 (DeRaad, 2022), and DARTR V.2.7.2 (Mijangos et al., 2022).

Appendix 3: Analytical Details

Analysis – Screening for Hybrid Madtoms

Utilizing the loci data aligned across all collected madtom species (Neosho, Freckled, Slender, and Stonecat), we screened for admixed individuals employing two clustering techniques implemented in R: Principal Component Analysis (PCA; ADE4 v.1.7-20; Dray and Dufour, 2007) and Sparse Non-Negative Matrix Factorization (sNMF; LEA v.3.6.0, Frichot & François, 2015). These techniques facilitated the identification of 'intermediate' individuals possessing ancestry from multiple species.

Before PCA, SNP data were centered and scaled, and the mean value replaced missing data at a locus. The first two principal components were visualized to identify intermediate individuals between species clusters.

We ran sNMF with 25 repetitions per *K* clusters (1 to 8) and a regularization parameter, $\alpha = 1000$ (Frichot & François, 2015). The 'best' *K* from each sNMF run minimizes the cross-validation entropy criterion. The best K—which was four, the number of species—was then used to impute missing data (impute function using method = "mode" in LEA). The sNMF algorithm was then repeated (as above) using imputed genotypes. The resulting Q-matrix of ancestry coefficients per cluster was used to screen for admixed individuals. Such mixed individuals and individuals representing non-NMT madtoms were excluded before proceeding with subsequent analyses.

Analysis – Inferring Population Structure

Clustering and ancestry analysis

We first visualized clusters by calculating F_{ST} among NMT sampling sites (HIERFSTAT v.0.5-11; Goudet and Jombart, 2022) to inform the construction of a neighbor-joining tree (APE v.5.6-2; Paradis and Schliep, 2019).

Next, we clustered individuals using *K*-means clustering paired with Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010). Discriminant functions combine principal components (PCs) to maximally separate groups determined using K-means clustering. Neither DAPC nor PCA tolerates missing data, so missing data were imputed randomly to avoid introducing bias (DARTR v.2.7.2; Mijangos et al., 2022). Then, the imputed data were clustered using *K*-means clustering with max number of clusters set equal to the number of sampling sites (N=14) and the number of principal components set to 6 based on visualization of the PCA scree-plot (ADEGENET v.2.1.10; Jombart and Ahmed, 2011). Based on the Bayesian Information Criteria plateau, the optimum number of clusters was used to inform the DAPC with an initial run using all PCs and several discriminant axes equal to the number of groups (Jombart et al., 2010). This run defined the optimum number of PCs to use in conjunction with the optimized *a*-score (Miller et al., 2020) before running a final DAPC using those PCs.

We then clustered data using a different dimension reduction analysis called t-distributed stochastic neighbor embedding (t-SNE; van der Maaten & Hinton, 2008). We used the RTSNE v.0.16 package (Krijthe, 2015) using the imputed data matrix (as above). We used the following

parameters for the Rtsne() function: max iterations = 1000, perplexity = maximum ((N individuals -1)/3), initial dimensions = 5, pca = TRUE, and theta = 0.0.

These results were contrasted with model-based biogeographical ancestry analysis termed ADMIXTURE (Alexander et al., 2009). This analysis assumes populations can be divided into *K*-clusters identified by permuting membership to minimize linkage disequilibrium and departure from Hardy–Weinberg expectations. We evaluated a range of K =1–14 using ADMIXPIPE V.3 (Mussmann et al., 2020b) with the remaining default parameters. Cross-validation error and log-likelihood were assessed to compare each *K*.

Genetic divergence

Upon establishing genetic populations, individual genetic data were grouped accordingly. We estimated the genetic variation explained among populations, among sites *within* populations, and among individuals within sites via Analysis of Molecular Variance (AMOVA; Excoffier et al., 1992) by implementing the PEGAS V.1.1 (Paradis, 2010) version of AMOVA and the POPPR V.2.9.4 (Kamvar et al., 2014) wrapper function.

Pairwise genetic divergence was calculated using the index G''_{ST} (Meirmans et al., 2011), wherein the values can range from 0 (indicating no difference) to 1 (indicating complete difference). To ascertain if these genetic differences were notably greater than zero, we accessed significance through bootstrap re-sampling with 10,000 replicates. The *p*-values were multiplied by the number of tests (N=14) to correct for multiple tests (Bonferroni). We employed the R packages DIVERSITY V.1.9.90 (Keenan et al., 2013) and DARTR.

Furthermore, the proportion of alleles *not* shared among each population was calculated (β-Differentiation = ${}^{0}H_{\beta}$; DARTR), offering a metric that equally weights both common and rare alleles; this contrasts with G''_{ST}, which is more influenced by common alleles, meaning these indices can vary for species that have undergone recent demographic shifts (Sherwin et al., 2017).

Analysis – Riverscape Genetic Analysis

To place the results of our population genetic structure analyses into an environmental context, we examined how the 'riverscape' characteristics may impede or promote individual movement through their relationship with observed patterns of genetic divergence. Here, we leveraged the expectation that if a feature (e.g., a barrier or depth) impedes individual movement, this will be reflected in greater levels of genetic divergence.

We first applied a least-squares algorithm (Kalinowski et al., 2008) implemented in AUTOSTREAMTREE (Chafin et al., 2023a) to map genetic distances to stream segments, given pairwise linearized F_{ST} (=F_{ST}/1- F_{ST}), using Weir and Cockerham's estimator (Weir and Cockerham, 1984). We reduced redundancy inherent in our linked SNP data: Phased alleles for each locus were concatenated to form 'pseudo-microhaplotypes' and frequencies used to compute F_{ST}. As in Kalinowski et al. (2008), we employed a constrained optimization procedure to remove negative fitted distances, wherein negative distances are iteratively constrained to zero

and re-running the least-squares procedure. The result is an F_{ST} estimate explained by each segment in the stream network.

We then tested the effects of individual barriers in the stream network as inflating genetic differentiation using the standardized index of genetic fragmentation (F_{INDEX}) developed by Prunier et al. (2020). Here, the immediate upstream and downstream populations relative to each barrier are sampled to compute a theoretical range of genetic variation that could theoretically accumulate – as a function of mutation and genetic drift and the expected heterozygosity of the observed sample – given the age of the artificial barrier. Here, a lower bound reflects a fully permeable barrier, with differentiation only due to stochastic change within a single panmictic population, and an upper bound reflecting an hypothesis of zero gene flow. The index is then standardized within this range to account for theoretical variations in the upper bound as a function of barrier age and effective population size (Prunier et al., 2020), which necessarily will vary across comparisons. The standardized index varies from 0%, where the observed differentiation matches what is expected if the barrier has no effects, to 100%, that of a total barrier to gene flow.

We incorporated a comprehensive set of environmental features (N=281) sourced from the HydroATLAS database (version 0.1; Linke et al., 2019) to model their contributions to genetic differentiation. To augment this dataset, we included several specific variables serving as proxies of dam effects: the density of barriers per river segment, the age of the oldest barrier within a segment (expressed in generations, assuming one generation per year), and indices of river fragmentation and connectivity as globally assessed by Grill et al. (2019). Features were first reduced as a pre-processing step wherein variables demonstrating high pairwise correlations (correlation coefficient > 0.95) were removed to mitigate multicollinearity. Subsequently, we employed a random forest regression approach, utilizing the SCIKIT-LEARN library (Pedregosa et al., 2011), to evaluate the 'importance' of each variable in predicting pairwise genetic differentiation, represented by F_{ST} values. Through this method, we identified and excluded any features that contributed minimally to the model's explanatory power, specifically those with relative importance <0.01.

Retained features were used as inputs to RESISTNET (Chafin et al., 2023b), a graph-based approach employing a genetic algorithm to optimize an 'effective resistance network' representing the input features' multivariate movement resistance, allowing for non-linear transformations. This procedure was conducted in 10 independent replicates, each using a population size of 1000 models and optimizing for a maximum of 600 generations, terminating after 30 successive failures to improve model fit (as evaluated via AIC). Search space was constrained using the *'—allShapes –posWeight –minWeight 0.1 –maxShape 5'* options, and the final resistance model was computed as a model average of the best model (by AIC) selected by each replicate.

Analysis – Estimating Population Genetic Parameters

The comparison of population-level diversity was conducted with various estimates of genetic diversity derived from the loci data. These estimates include (i) mean allelic richness per locus; (ii) the count of alleles unique to each population (private alleles); (iii) the inbreeding extent within a population (F_{IS}); (iv) heterozygosity (expected and observed); (v) evolvability based on

Shannon Index. These estimates were calculated using R packages DIVERSITY, POPPR, and DARTR. We also calculated the genome-wide heterozygosity from the unfiltered alignment containing polymorphic and non-polymorphic loci for comparison to other studies with IPYRAD. Heterozygosity based only on polymorphic loci will be biased upwards (Schmidt et al., 2021).

We estimated each population's effective population size (N_e) with NEESTIMATOR V.2 (Do et al., 2014) with the DARTR wrapper function. We used more strictly filtered data for this calculation by retaining only loci found across >90% of individuals and retaining only one locus per 10,000 bases because the analysis can be strongly affected by missing data and physical linkage (Peel et al., 2013). We assessed N_e using a minor allele frequency threshold of 0.1 while retaining singletons and using the random mating model (Do et al., 2014).

Finally, we tested for evidence of a recent population bottleneck using an estimate of departure from neutral expectations for genetic variability called Tajima's *D* (Tajima, 1989), which is the difference between the mean pairwise genetic differences and the number of segregating sites. For a set of putatively neutrally evolving loci—the expectation for our genetic markers—we expect Tajima's $D = 0$ when the population is not undergoing selection or recent demographic changes. Tajima's $D < 0$ is consistent with a recent selective sweep or population expansion after a recent bottleneck. Tajima's $D > 0$ is consistent with either balancing selection or a sudden population bottleneck. For this calculation, we relaxed data filtering to allow singletons so the number of segregating sites was not underestimated (minor allele count =1) and retained only one locus per 10,000 bases (thin = 10000) to minimize linkage disequilibrium. The precision of singletons is essential for calculating Tajima's *D* (Zhang et al., 2015) and can be difficult for low-coverage (<20x) next-gen sequencing data (Korneliussen et al., 2013). But our mean coverage was very high (78x). Tajima's *D* was calculated using the R package POPGENOME V.2.7.7 (Pfeifer et al., 2014).

Analysis – Loci Under Selection and Local Adaptation

We assessed local adaptation in NMT via genotype-environment association analysis (GEA). The association between each locus and a comprehensive set of environmental characteristics (N=281) from the HydroATLAS v.0.1 database (Linke et al., 2019) was determined. We removed environmental characteristics that were categorical nominal or had extremely low variances of less than one. Variables were then standardized by subtracting the mean and dividing by the standard deviation. These variables were then separated into five categories: (i) hydrological and physiographic, (ii) climate, (iii) landcover, (iv) geology and soils, and (v) anthropogenic.

To mitigate collinearity, variables within each category were consolidated into composite variables using Robust Principal Components Analysis (ROBPCA) using the R package ROSPCA V.1.0.4 (Reynkens, 2018). We retained the first PC from each of the five ROBPCAs with skew = TRUE and stand = FALSE. These five principal components represented the aforementioned environmental categories and served as our environmental predictors of genetic variation.

The relationships between loci and the composite environmental variables were evaluated using Redundancy Analysis (RDA), a robust multivariate extension of multiple linear regression (Capblancq and Forester, 2021). Adaptive loci were inferred as deviating more than ±3 standard deviations (p<0.003) from the mean loading of all loci on the canonical axes generated by

RDA—equal to the number of environmental PCs—that predict genotype-environment correlations (Forester et al., 2018). Notably, most loci are anticipated to be neutral concerning selection as they predominantly occur in non-coding regions of the genome or result in synonymous mutations that do not alter amino acids (Kimura, 1991).

Our next objective was to discern if distinct populations or areas within the Neosho River system exhibited variations in local adaptation (i.e., different adaptations to different environments). To this end, we revisited population structure analysis, employing solely the adaptive outlier loci deduced from the aforementioned RDA. We again utilized t-SNE (van der Maaten & Hinton, 2008) to examine the population structure of this subset, given its proficiency in rapidly and accurately depicting hierarchical structure (Li et al., 2017). We anticipated that clusters separating in the t-SNE ordination space would indicate geographical regions experiencing differences in local adaptation. We used the RTSNE (as above) with the same parameters: max iterations = 1000, perplexity = maximum ((N individuals -1)/3), initial dimensions = 5, pca = TRUE, and theta = 0.0.

Functional annotation of the NMT genome (Whitacre et al., 2022) was done by mapping functional elements from the Channel Catfish (*Ictalurus punctatus*) genome (IpCoco_1.2; Liu et al., 2016) using an iterative alignment algorithm in LIFTOFF (Shumate and Salzberg, 2021). Accuracy for cross-species mapping (as herein) was improved by requiring an additional 10% of gene length in flanking sequence to align (*-flank 0.1*) and by re-aligning exon/CDS sequences to restore reading frame in cases of start/stop codon loss due to misalignment (*-polish*). Functional classifications and gene associations for single nucleotide variants were additionally predicted using SNPEFF (Cingolani et al., 2012) using default settings. Gene ontologies were collated for gene symbols as annotated in the IpCoco_1.2 assembly using the MYGENE Python API to access QuickGO (Binns et al., 2009).

Appendix 4: Discussion of Local Adaptation

GEA Analysis Caveats

Genotype-environment association analysis (GEA) is a standard way of inferring a signal of locally adapted genotypes along environmental gradients (Forester et al., 2018). However, spurious results caused by factors other than selection can confound interpretation. Therefore, plausible alternative explanations must be carefully considered to avoid erroneous conclusions about local adaptation.

Original GEA methods did not account for population structure, which may have elevated false positives. The approach relies on 'outlier loci' associated with environmental gradients, but this relationship is not necessarily driven by selection (i.e., adaptation). However, it could be a coincidence due to genetic drift or isolation by distance along the environmental gradient. Naturally, correcting for population structure results in fewer false positives but at the cost of losing power to detect genuine associations (false negatives) (Lotterhos and Whitlock, 2015).

There is no 'correct' answer concerning whether population structure should be factored out of GEA analyses. Instead, it is based on the tolerance for false positives and removal of statistical signal (Capblancq and Forester, 2021). Simple RDA, the approach we employed in this study for GEA, has a 'superior combination' of false-positive and false-negative rates when using a conservative threshold for selecting outlier loci (3 sd ± mean) and relatively weak global population structure ($F_{ST} \approx 0.05$; Forester et al., 2018), as was the case in NMT ($F_{ST} = 0.054$). Furthermore, the recovery of environmental-linked outlier loci consistently associated with similar genetic functions makes it unlikely that these are spurious loci and underscores the signal of genuine local adaptation.

We note GEAs are also biased toward detecting large-effect loci, so they are more likely to accurately detect monogenic and oligogenic architectures over polygenic architectures composed of many small-effect loci. This bias is problematic because many adaptive traits are polygenic, including traits likely to be important in shifting environments (Quigley et al., 2020). Thus, while we detect a strong signal of local adaptation, our data may only allow us to glimpse a fraction of the adaptation across the genome. Experimental associations between phenotype and candidate loci can also validate genomic predictions, e.g., via common garden approaches, but this is especially difficult for species of conservation concern with small population sizes in the wild (Schmidt et al., 2023). Further, non-lethal observational studies of NMT phenotypic characteristics informed by these candidate loci and biological functional relationships could provide further validation through quantitative genomics approaches aimed at estimating the narrow-sense heritability of traits that vary along the longitudinal gradient of the Neosho River system (Schmidt et al., 2023).

Adaptive Genes

An overarching concern with GEAs is whether outlier loci are indeed biologically relevant, and thus, candidate adaptive markers need further validation to increase confidence in the results (Schmidt et al., 2023). We provide a first step in validating the candidate adaptive markers by establishing consistent connections between genotypes and phenotypes, i.e., biological functions. The candidate adaptive markers we detected were associated with amino acid changes in protein-coding regions and modifications within non-coding regions that may affect upstream and/or downstream genes and gene complexes. These markers were collectively associated with *N*=30 genes (Table 6) whose known biological functions showed consistent patterns in that they are associated with anatomical development (head, eyes, brain), response to stimuli (stress, chemical, hypoxia), and metabolism (autophagy, ceramide, nitrogen, ubiquitin). Most of the putatively adaptive genes found here have been implicated in the evolution and adaptation of other fishes. See Appendix 4 for more discussion about these genes.

Most of the putatively adaptive genes found herein have been implicated in the evolution and adaptation of other fishes. The gene aim1, for instance, has been associated with pigment evolution, which might be associated with predator avoidance in teleost fish (Streelman et al., 2007). Moreover, prkar1a has been associated with pigment-related adaptation in lizards (Corl et al., 2018) and stress-related lipid metabolism (Ji et al., 2020).

On a metabolic front, immt has been linked to metabolic adaptation in fish (Lou et al., 2020), while genes such as fabp6, cers1, lpcat1, cerk, and gramd4 are associated with various metabolic pathways like fatty acid metabolism, metabolism linked to swimming performance, and rapid metabolic adaptation in response to environmental changes (Venkatachalam et al., 2017; Feng et al., 2021; Rimoldi et al., 2016; Willoughby et al., 2018; Jiang et al., 2023). The gene atg13 is implicated in cell recycling (autophagy) due to nutrient restriction, aiding in nutrient utilization (Balmori-Cedeño et al., 2019).

Performance and physiological resilience in extreme conditions also have genetic underpinnings. The gene bmx has been associated with tolerance to extreme temperatures in reef fishes (Veilleux et al., 2018), while atp13a2 and slc2a9 are implicated in swimming performance (Zheng and Li, 2021; Li et al., 2023). The genes usp28, eftud2, fmn2, rac3, and fgb have been linked to responses to hypoxic and temperature stress (Suo et al., 2022; Yuan et al., 2021; Han et al., 2021; Buckley and Somero, 2009; Dietrich et al., 2018).

Additionally, samd11 is associated with photoreception, which may relate to spawning (Ogawa and Corbo, 2021). Genes such as rnf103 and cpped1 are associated with adaptation to heavy metal pollution (Calboli et al., 2021; Vieira et al., 2023). Finally, harbi1 is another gene linked to adaptation induced by transposable elements (Etchegaray et al., 2022).

XII. SUPPLEMENTAL MATERIAL

SUPPLEMENTAL TABLES

Table S1: Information on 24 dams located in the Neosho River system. Included are: the dam's name, the river it is on, county, latitude, and longitude of its location. Also indicated is whether the dam is a large federal/state or smaller low head dam, its approximate construction start date, and the source of information. The complete Kansas Department of Agriculture, Division of Water Resources database was shared with the authors as an Excel spreadsheet and is available upon request. The earliest date found across sources was used because some dams have been reconstructed.

Dam Name	River	County	Lat	Long	Type	Approx construction date	Source of data
Marion	Cottonwood	Marion	38.36853	-97.08421	Federal	1964	Kansas Department of Agriculture, Division of Water Resources database
Cottonwood Falls	Cottonwood	Chase	38.37515	-96.54169	Low head	1860	Fencl et al. 2015
Cedar Point	Cottonwood	Chase	38.26174	-96.81963	Low head	1875	Kansas Historical Society (https://www.kshs.org/km/items/view/215457)
Soden	Cottonwood	Lyon	38.38612	-96.18187	Low head	1860	Fencl et al. 2015
Council Grove	Neosho	Morris	38.67769	-96.50671	Federal	1960	Kansas Department of Agriculture, Division of Water Resources database
Riverwalk	Neosho	Morris	38.66322	-96.48978	Low head	1995	Fencl et al. 2015
Correll	Neosho	Lyon	38.52398	-96.30409	Low head	1920	Fencl et al. 2015
Ruggles	Neosho	Lyon	38.46856	-96.25002	Low head	1920	Fencl et al. 2015
Emporia	Neosho	Lyon	38.43609	-96.20878	Low head	1890	Fencl et al. 2015
John Redmond	Neosho	Coffey	38.24228	-95.75551	Federal	1964	Kansas Department of Agriculture, Division of Water Resources database
Burlington	Neosho	Coffey	38.20589	-95.73064	Low head	1900	Kansas Department of Agriculture, Division of Water Resources database
Neosho Falls	Neosho	Neosho	38.00921	-95.55325	Low head	1935	US Geologic Survey (https://pubs.usgs.gov/wri/1999/4147/wrir19994147.pdf)
Iola	Neosho	Allen	37.92174	-95.42677	Low head	1936	Kansas Department of Agriculture, Division of Water Resources database
Humboldt	Neosho	Allen	37.81019	-95.44788	Low head	1960	Kansas Department of Agriculture, Division of Water Resources database
Barker	Neosho	Neosho	37.71958	-95.45445	Low head	1955	Kansas Department of Agriculture, Division of Water Resources database
Chanute (northeast)	Neosho	Neosho	37.70861	-95.41384	Low head	1952	Kansas Department of Agriculture, Division of Water Resources database
Chanute (east)	Neosho	Neosho	37.67136	-95.41935	Low head	1954	Kansas Department of Agriculture, Division of Water Resources database
Erie	Neosho	Neosho	37.54889	-95.24703	Low head	1957	Kansas Department of Agriculture, Division of Water Resources database
Parsons	Neosho	Labette	37.41433	-95.13705	Low head	1938	Kansas Department of Agriculture, Division of Water Resources database
Oswego	Neosho	Labette	37.17615	-95.10284	Low head	1936	Kansas Department of Agriculture, Division of Water Resources database
Chetopa	Neosho	Labette	37.03612	-95.08076	Low head	1968	Kansas Department of Agriculture, Division of Water Resources database
Riverton	Spring	Cherokee	37.06365	-94.70493	Low head	1910	Kansas Department of Agriculture, Division of Water Resources database
Baxter Springs	Spring	Cherokee	37.02322	-94.72071	Low head	1905	Kansas Department of Agriculture, Division of Water Resources database
Grand Lake	Grand	Mayes	36.46781	-95.04072	State	1940	Kansas Department of Agriculture, Division of Water Resources database

Table S2: Environmental variable correlations (loadings) with robust principal component (ROBPCA) axes. Five ROBPCAs were generated, one for each group: hydro-physiographic, climate, landcover, geology/soil, and anthropogenic factors. The total number of factors is shown below each group, along with the variance explained by the first PC. The top ten most correlated factors with PC1 are shown for each group. ROBPCs were used via RDA to test for locally adapted loci. These five PCs accounted for 87% of the total environmental variance.

Hydro-physiographic factors	PC
discharge_max	0.248
river_area_upstream	0.247
discharge_annual	0.246
discharge_min	0.245
river_volume_upstream	0.241
inundation_upstream_max	0.240
inundation_upstream_longterm_max	0.238
reservoir_volume_upstream	0.229
elevation_reach_avg	-0.211
elevation_reach_max	-0.220

N=27 factors | 57% var exp.

Climate factors	PC
temp_reach_annual_avg	0.128
temp_reach_annual_min	0.128
actual_evapotrans_reach_avg	0.127
precipitation_reach_avg	0.127
aridity_reach_avg	0.123
moisture_index_reach_avg	0.123
actual_evapotrans_upstream_avg	0.104
precipitation_upstream_avg	0.104
precipitation_jul	-0.101
snow_cover_feb	-0.112

N=78 factors | 76% var exp.

Landcover factors	PC
herbaceous_cover_upstream	0.313
grassland_upstream	0.308
pasture_upstream	0.283
grassland_reach	0.197
pasture_reach	0.107
forest_cover_reach	-0.242
cropland_upstream	-0.277
broadleaved_closed_deciduous_upstream	-0.286
forest_cover_upstream	-0.291
cultivated_upstream	-0.308

N=25 factors | 38% var exp.

Geology factors	PC
soil_water_reach_avg	0.248
soil_water_upstream_avg	0.199
erosion_upstream	0.192
clay_soil_reach	0.188
silt_soil_reach	0.129
erosion_reach	0.114
org_carbon_soil_reach	-0.149
org_carbon_soil_upstream	-0.172
sand_soil_upstream	-0.187
sand_soil_reach	-0.189

N=24 factors | 66% var exp.

Anthropogenic factors	PC
population_upstream	0.291
human_footprint09_upstream	0.287
road_density_upstream	0.282
pop_density_upstream	0.268
road_density_reach	-0.180
human_footprint93_reach	-0.195
pop_density_reach	-0.210
urban_extent_reach	-0.214
human_dev_index	-0.252
gross_dom_prod_avg_reach	-0.264

N=17 factors | 50% var exp.

Table S3: Summary table of environmental factors for stream reaches where Neosho Madtom was collected across the Neosho River system. These factors are a subset of the total (*N*=281) reduced using robust principal component analysis. The five ROBPCs produced from these variables explained 87% of the total environmental variance and were used to examine loci for local adaptation. The environmental variables are grouped by category (used for ROBPC), and their units are given along with summary statistics and the original code from the HydroATLAS v.0.1 database (Linke et al., 2019).

Environmental Variable	Category	Units	Min	Median	Max	Mean	Std dev	HydroRiv Code
gross_dom_prod_avg_reach	anthro	US dollars	46422.0	51129.0	51129.0	50341.4	1630.8	gdp_ud_cav
human_dev_index	anthro	index (x1000)	809.0	934.0	934.0	915.6	42.5	hdi_ix_cav
human_footprint09_upstream	anthro	index (x10)	53.0	58.0	118.0	65.1	16.4	hft_ix_u09
human_footprint93_reach	anthro	index (x10)	31.0	114.0	392.0	130.9	104.9	hft_ix_c93
pop_density_reach	anthro	people/km2	0.5	3.3	687.5	89.7	205.4	ppd_pk_cav
pop_density_upstream	anthro	people/km2	3.5	6.7	41.4	9.6	9.3	ppd_pk_uav
population_upstream	anthro	thousands	5.2	50.5	267.2	60.9	63.1	pop_ct_usu
road_density_reach	anthro	m/km2	0.0	122.0	2323.0	389.3	629.5	rdd_mk_cav
road_density_upstream	anthro	m/km2	86.0	148.0	549.0	175.4	110.0	rdd_mk_uav
urban_extent_reach	anthro	percent cover	0.0	0.0	54.0	6.8	16.4	urb_pc_cse
actual_evapotrans_reach_avg	climate	mm	731.0	794.0	919.0	820.0	68.6	aet_mm_cyr
actual_evapotrans_upstream_avg	climate	mm	718.0	743.0	924.0	773.7	66.4	aet_mm_uyr
aridity_reach_avg	climate	index (x100)	71.0	78.0	89.0	80.0	5.6	ari_ix_cav
moisture_index_reach_avg	climate	index (x100)	-29.0	-22.0	-11.0	-20.0	5.6	cmi_ix_cyr
precipitation_jul	climate	mm	79.0	92.0	96.0	90.2	6.2	pre_mm_c07
precipitation_reach_avg	climate	mm	861.0	938.0	1087.0	968.5	82.1	pre_mm_cyr
precipitation_upstream_avg	climate	mm	846.0	876.0	1095.0	913.4	80.1	pre_mm_uyr
snow_cover_feb	climate	percent cover	0.0	10.0	13.0	7.7	4.1	snw_pc_c02
temp_reach_annual_avg	climate	Celsius (x10)	125.0	129.0	144.0	132.7	6.7	tmp_dc_cyr
temp_reach_annual_min	climate	Celsius (x10)	-27.0	-21.0	10.0	-12.8	13.3	tmp_dc_cmn
clay_soil_reach	geology	percent	22.0	23.0	26.0	23.5	1.3	cly_pc_cav
erosion_reach	geology	kg/hectare/year	892.0	2060.0	6218.0	2750.8	1790.9	ero_kh_cav
erosion_upstream	geology	kg/hectare/year	1160.0	1473.0	5350.0	1994.6	1257.4	ero_kh_uav
org_carbon_soil_reach	geology	percent	17.0	18.5	23.0	19.1	1.8	soc_th_cav
org_carbon_soil_upstream	geology	percent	17.0	20.0	22.0	19.8	1.5	soc_th_uav
sand_soil_reach	geology	percent	22.0	29.0	31.0	28.0	2.8	snd_pc_cav
sand_soil_upstream	geology	percent	27.0	29.0	31.0	29.1	1.2	snd_pc_uav
silt_soil_reach	geology	percent	44.0	49.0	51.0	48.5	1.9	slt_pc_cav
soil_water_reach_avg	geology	percent	62.0	68.0	79.0	70.5	5.5	swc_pc_cyr
soil_water_upstream_avg	geology	percent	61.0	65.0	81.0	67.2	6.1	swc_pc_uyr
discharge_annual	hydrophysio	m3/s	8.7	39.3	126.5	49.0	36.5	dis_m3_pyr
discharge_max	hydrophysio	m3/s	17.7	84.3	228.9	94.1	66.2	dis_m3_pmx
discharge_min	hydrophysio	m3/s	4.3	18.9	71.9	25.2	21.2	dis_m3_pmn
elevation_reach_avg	hydrophysio	m a.s.l	241.0	328.0	372.0	312.9	44.9	ele_mt_cav
elevation_reach_max	hydrophysio	m a.s.l	249.0	332.0	409.0	328.3	50.1	ele_mt_cmx
inundation_upstream_longterm_max	hydrophysio	percent cover	0.0	3.0	6.0	2.5	1.9	inu_pc_ult
inundation_upstream_max	hydrophysio	percent cover	0.0	1.0	4.0	1.3	1.2	inu_pc_umx
reservoir_volume_upstream	hydrophysio	million m3	0.0	233.0	1061.0	421.7	452.6	rev_mc_usu
river_area_upstream	hydrophysio	hectares	149.7	847.6	2938.9	1053.3	861.9	ria_ha_usu
river_volume_upstream	hydrophysio	thousand m3	512.6	5368.7	30116.7	8477.4	9087.6	riv_tc_usu
broadleaved_closed_deciduous_upstream	landcover	percent cover	0.0	0.0	7.0	0.7	1.8	glc_pc_u02
cropland_upstream	landcover	percent cover	37.0	45.0	57.0	45.6	5.8	crp_pc_use
cultivated_upstream	landcover	percent cover	13.0	18.0	96.0	33.3	27.7	glc_pc_u16
forest_cover_reach	landcover	percent cover	0.0	0.0	70.0	8.1	17.3	for_pc_cse
forest_cover_upstream	landcover	percent cover	0.0	1.0	8.0	1.4	1.9	for_pc_use
grassland_reach	landcover	percent cover	0.0	6.0	100.0	36.8	42.5	pnv_pc_c10
grassland_upstream	landcover	percent cover	0.0	73.0	100.0	64.3	31.3	pnv_pc_u10
herbaceous_cover_upstream	landcover	percent cover	0.0	79.0	87.0	64.0	29.4	glc_pc_u13
pasture_reach	landcover	percent cover	9.0	21.0	36.0	22.4	7.8	pst_pc_cse
pasture_upstream	landcover	percent cover	20.0	33.0	47.0	33.7	7.9	pst_pc_use

Table S4: Summary table of environmental factors for stream reaches where Neosho Madtom was collected across the Neosho River system. These factors are a subset of the total (*N*=281). Six potential management units: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR).

Pop	Site	discharge annual (m^3)	discharge min (m^3)	inundation reach min (%)	river area reach (hectares)	elevation reach avg (m.a.s.l)	slope reach (degrees x 10)
NRU	NDUN	8.723	4.307	14	7.978	369	9
	NAME	9.89	4.902	0	0.531	357	12
CWR	CCOT	20.953	10.312	2	9.38	372	12
	CCFE	21.399	10.516	0	2.969	367	10
	CESG	30.885	15.073	2	21.546	346	5
NCC	NECW	12.595	6.24	0	5.856	344	6
	NRAP	46.885	22.818	0	7.486	328	2
NRM	NROY	62.069	31.699	100	9.208	304	1
	NIOL	66.338	34.165	29	8.244	305	7
NRL	NSTP	97.85	55.615	27	33.675	267	5
	NLAB	108.577	63.113	60	4.832	249	4
	NMIA	126.547	71.908	29	25.834	241	4
SPR	SMCP	66.188	24.673	0	9.49	254	18
	SWAC	31.814	11.769	1	23.98	272	5

Pop	Site	temp annual avg (°C)	precipitation Avg (mm)	population upstream (thousands)	human footprint09 (index x 10)	cropland Reach (%)	pasture reach (%)
NRU	NDUN	12.5	879	5.179	30	47	34
	NAME	12.6	885	5.356	32	50	31
CWR	CCOT	12.9	861	13.724	145	18	31
	CCFE	12.9	864	14.261	99	20	36
	CESG	12.7	896	30.825	276	51	20
NCC	NECW	12.7	902	16.979	382	60	21
	NRAP	12.7	938	50.526	46	79	17
NRM	NROY	13.2	987	58.382	52	61	15
	NIOL	13.5	1004	59.649	36	47	24
NRL	NSTP	13.8	1044	90.729	52	51	24
	NLAB	14.1	1058	97.682	218	50	9
	NMIA	14.4	1075	118.958	91	57	21
SPR	SMCP	14.3	1087	267.182	60	31	25
	SWAC	14	1078	62.521	123	54	15

Pop	Site	herbaceous upstream (%)	cultivated reach (%)	cultivated upstream (%)	erosion reach (kg/hect/year)	grassland reach (%)	forest cover reach (%)
NRU	NDUN	87	75	13	1118	100	0
	NAME	86	100	14	939	100	0
CWR	CCOT	82	29	16	1505	100	0
	CCFE	81	14	16	1755	100	0
	CESG	79	11	18	1799	6	12
NCC	NECW	86	52	14	892	54	9
	NRAP	81	33	16	5797	6	0
NRM	NROY	78	100	19	4098	5	0
	NIOL	74	86	23	2060	62	0
NRL	NSTP	56	93	41	4990	0	0
	NLAB	52	100	46	6218	0	0
	NMIA	45	74	52	2166	0	23
SPR	SMCP	0	30	90	1547	0	70
	SWAC	0	95	96	3463	0	5

Table S5: Local adaptation of loci was tested for Neosho Madtom using redundancy analysis (RDA). Five principal components representing the major environmental factors below were tested for their ability to explain the individual genetic variation of alleles significantly. The overall model of genetic variation explained by these five predictors was significant, as was each contribution from the individual environmental factors. Note: the minimum *p*-value obtainable based on the number of permutations (*N*=999) is 0.001. The model accounted for 6% of the variation in the individual genetic data—this might seem low, but we expect most loci to be neutral due to synonymous mutations and non-coding regions. That 6% is referred to as the "constrained variance," and the proportion explained by each environmental factor is shown as "Var. Exp." The *p*-value for each environmental factor is Bonferroni adjusted for multiple comparisons.

RDA overall *p*-value = 0.001					
Factor	**Df**	**Variance**	**F**	***p*-value**	**Var. Exp.**
hydrophysio	1	47.82	3.32	0.005	39%
climate	1	80.32	5.57	0.005	27%
landcover	1	43.72	3.03	0.005	16%
geology	1	34.03	2.36	0.005	11%
anthropogenic	1	38.88	2.70	0.005	8%
RDA $R^2 = 0.09$					
RDA adjusted $R^2 = 0.06$					

Table S6: Local adaptation of loci was tested for Neosho Madtom using redundancy analysis (RDA). RDA produces composite predictor axes (PCs) on which the original factors are correlated (loadings). These correlations are shown below for the five environmental composite factors used to test for local adaptation in Neosho Madtom.

Loadings on RDA axes					
Factor	**RDA1**	**RDA2**	**RDA3**	**RDA4**	**RDA5**
hydrophysio	0.264	**0.515**	**0.716**	0.367	0.134
climate	**0.758**	0.169	0.366	**0.507**	0.083
landcover	**-0.942**	-0.085	0.078	-0.24	-0.203
geology	**0.837**	0.074	0.466	0.236	-0.144
anthropogenic	**0.755**	-0.255	0.172	0.226	**0.534**

Table S7: Local adaptation of loci was tested for Neosho Madtom using redundancy analysis (RDA). Locally adapted loci were determined based on the distribution of loci loadings on each RDA axis (> ± 3 standard deviations from the mean). The rest are assumed neutral. We note the environmental factor for each locus to which it was most highly correlated.

Axis	N Outlier Loci	Factor	N Loci highest loading
RDA1	14	hydrophysio	20
RDA2	11	landcover	18
RDA3	7	geology	11
RDA4	11	anthropogenic	9
RDA5	18	climate	3
Total=	61		

Table S8: Associated environmental factors for *N*=61 locally adapted loci within Neosho Madtom. Columns include: the locus identifier, SNP ID; the environmental component variable (*N*=5) with the highest correlation with each locus and the value (*r*); the specific environmental factor (*N*=281) with the highest correlation with the locus and the value (*r*).

SNP ID	Component	*r*	Main Factor	*r*
loc10680_pos97	anthropogenic	0.18	broadleaved_closed_deciduous_reach	0.24
loc19822_pos71	anthropogenic	0.15	precipitation_jun	0.29
loc6005_pos68	anthropogenic	0.38	broadleaved_closed_deciduous_reach	0.48
loc10344_pos4	anthropogenic	0.38	broadleaved_closed_deciduous_reach	0.44
loc13356_pos33	anthropogenic	0.31	forest_cover_upstream	0.41
loc19585_pos6	anthropogenic	0.38	broadleaved_closed_deciduous_reach	0.48
loc4432_pos9	anthropogenic	0.28	broadleaved_closed_deciduous_reach	0.35
loc8100_pos94	anthropogenic	0.27	temp_may	0.31
loc8428_pos7	anthropogenic	0.14	cultivated_reach	-0.25
loc10145_pos17	climate	0.28	potential_evapotrans_oct	0.33
loc1164_pos70	climate	0.68	temp_sep	-0.74
loc4874_pos1	climate	0.27	snow_cover_feb	0.31
loc4736_pos21	geology	0.32	erosion_upstream	0.35
loc14172_pos82	geology	0.36	erosion_upstream	0.46
loc3028_pos42	geology	0.39	herbaceous_cover_reach	0.44
loc1421_pos123	geology	0.27	sand_soil_reach	-0.35
loc14568_pos18	geology	0.16	precipitation_aug	-0.46
loc20740_pos76	geology	0.28	erosion_upstream	0.41
loc3665_pos20	geology	0.3	erosion_upstream	0.43
loc7221_pos24	geology	0.13	river_area_reach	-0.27
loc7222_pos62	geology	0.16	river_area_reach	-0.28
loc7736_pos16	geology	0.38	herbaceous_cover_reach	0.56
loc8823_pos30	geology	0.28	erosion_upstream	0.41
loc19359_pos6	hydrophysio	0.23	wetland_inclusive_reach	0.46
loc10014_pos12	hydrophysio	0.39	potential_evapotrans_upstream_avg	-0.59
loc10283_pos65	hydrophysio	0.28	lake_volume_upstream	0.37
loc10406_pos90	hydrophysio	0.16	wetland_inclusive_reach	0.38
loc12475_pos17	hydrophysio	0.25	wetland_inclusive_reach	0.32
loc12552_pos53	hydrophysio	0.29	org_carbon_soil_reach	0.5
loc12940_pos3	hydrophysio	0.29	savanna_upstream	-0.5
loc13549_pos67	hydrophysio	0.34	grassland_reach	0.49
loc1408_pos102	hydrophysio	0.31	savanna_upstream	-0.5
loc15878_pos3	hydrophysio	0.34	wetland_inclusive_reach	0.46
loc18008_pos21	hydrophysio	0.3	org_carbon_soil_reach	0.46
loc19318_pos5	hydrophysio	0.43	lake_area_reach	0.62
loc20654_pos103	hydrophysio	0.35	potential_evapotrans_upstream_avg	-0.5
loc21032_pos1	hydrophysio	0.42	potential_evapotrans_upstream_avg	-0.6
loc3186_pos57	hydrophysio	0.29	wetland_inclusive_reach	0.37
loc3351_pos64	hydrophysio	0.27	savanna_upstream	-0.42
loc4679_pos15	hydrophysio	0.29	lake_volume_upstream	0.35
loc4809_pos37	hydrophysio	0.3	wetland_exclusive_reach	0.43
loc5178_pos8	hydrophysio	0.2	precipitation_aug	0.48
loc7635_pos4	hydrophysio	0.44	potential_evapotrans_upstream_avg	-0.59
loc1767_pos35	landcover	0.55	tem_deciduous_upstream	0.66
loc12218_pos99	landcover	0.61	tem_deciduous_reach	0.69
loc13414_pos61	landcover	0.56	nighttime_lights_upstream	0.61
loc13508_pos8	landcover	0.56	erosion_upstream	0.61
loc14964_pos1	landcover	0.58	tem_deciduous_upstream	0.68
loc16738_pos9	landcover	0.5	tem_deciduous_upstream	0.65
loc17339_pos94	landcover	0.26	potential_evapotrans_upstream_avg	0.46
loc19281_pos19	landcover	0.61	tem_deciduous_upstream	0.72
loc19505_pos21	landcover	0.67	tem_deciduous_upstream	0.74
loc3154_pos40	landcover	0.61	tem_deciduous_upstream	0.69
loc3715_pos23	landcover	0.62	tem_deciduous_reach	0.73
loc3731_pos91	landcover	0.56	clay_soil_reach	0.64
loc457_pos25	landcover	0.23	precipitation_aug	0.39
loc5273_pos20	landcover	0.68	tem_deciduous_upstream	0.74
loc5661_pos87	landcover	0.45	forest_cover_upstream	0.47
loc7550_pos61	landcover	0.58	tem_deciduous_reach	0.65
loc8013_pos88	landcover	0.58	road_density_upstream	0.59
loc8366_pos58	landcover	0.57	broadleaved_closed_deciduous_upstream	0.64

Table S9: Biallelic frequencies (1 & 2) for *N*=61 locally adapted loci within Neosho Madtom globally across all populations (G1/G2) and within each of the six populations. Six potential management units: Cottonwood River (CWR), Upper Neosho River (NRU), Neosho-Cottonwood Confluence (NCC), Middle Neosho River (NRM), Lower Neosho River (NRL), and the Spring River (SPR).

SNP ID	G1	G2	NRU1	NRU2	CWR1	CWR2	NCC1	NCC2	NRM1	NRM2	NRL1	NRL2	SPR1	SPR2
loc1767_pos35	0.04	0.96	0	1	0	1	0	1	0	1	0	1	0.28	0.72
loc4736_pos21	0.12	0.88	0.12	0.88	0.02	0.98	0.08	0.92	0.13	0.87	0.08	0.92	0.29	0.71
loc10680_pos97	0.27	0.73	0.27	0.73	0.3	0.7	0.32	0.68	0.26	0.74	0.21	0.79	0.29	0.71
loc14172_pos82	0.05	0.95	0	1	0	1	0	1	0.03	0.97	0	1	0.16	0.84
loc19359_pos6	0.36	0.64	0.62	0.38	0.2	0.8	0.26	0.74	NA	NA	0.29	0.71	NA	NA
loc19822_pos71	0.26	0.74	0.13	0.87	0.16	0.84	0.33	0.67	0.31	0.69	0.37	0.63	0.25	0.75
loc3028_pos42	0.24	0.76	0.17	0.83	0.12	0.88	0.43	0.57	0	1	0.31	0.69	0	1
loc6005_pos68	0.01	0.99	0	1	0	1	0	1	0	1	0	1	0.06	0.94
loc10014_pos12	0.06	0.94	0	1	0.39	0.61	0	1	0	1	0.03	0.97	0	1
loc10145_pos17	0.06	0.94	0.15	0.85	0	1	0.02	0.98	0	1	0.05	0.95	0.12	0.88
loc10283_pos65	0.19	0.81	0.27	0.73	0.08	0.92	0.1	0.9	0.42	0.58	0.17	0.83	0.1	0.9
loc10344_pos4	0.06	0.94	0.06	0.94	0.05	0.95	0.05	0.95	0.02	0.98	0.02	0.98	0.19	0.81
loc10406_pos90	0.25	0.75	0.41	0.59	0.1	0.9	0.2	0.8	NA	NA	0.28	0.72	NA	NA
loc1164_pos70	0.18	0.82	0	1	0.5	0.5	0.33	0.67	0	1	0.24	0.76	0	1
loc12218_pos99	0.07	0.93	0.02	0.98	0.03	0.97	0	1	0.02	0.98	0	1	0.42	0.58
loc12475_pos17	0.02	0.98	0.13	0.87	0	1	0	1	0	1	0.02	0.98	0	1
loc12552_pos53	0.3	0.7	0.14	0.86	0.93	0.07	0	1	0.3	0.7	0.07	0.93	0.25	0.75
loc12940_pos3	0.24	0.76	0	1	0.48	0.52	0	1	0.18	0.82	0.33	0.67	0.08	0.92
loc13356_pos33	0.05	0.95	0	1	0	1	0.1	0.9	0.05	0.95	0.02	0.98	0.16	0.84
loc13414_pos61	0.07	0.93	0.04	0.96	0	1	0	1	0	1	0	1	0.38	0.62
loc13508_pos8	0.14	0.86	0.1	0.9	0	1	0.12	0.88	0.03	0.97	0.12	0.88	0.56	0.44
loc13549_pos67	0.09	0.91	0	1	0.38	0.62	0.05	0.95	0.08	0.92	0.02	0.98	0	1
loc1408_pos102	0.34	0.66	0.12	0.88	0.71	0.29	0	1	0.17	0.83	0.1	0.9	0.31	0.69
loc1421_pos123	0.05	0.95	0	1	0	1	0	1	0.08	0.92	0.05	0.95	0.12	0.88
loc14568_pos18	0.82	0.18	0.92	0.08	0.71	0.29	1	0	0.57	0.43	0.98	0.02	0.75	0.25
loc14964_pos1	0.22	0.78	0	1	0.07	0.93	0	1	0.15	0.85	0	1	0.58	0.42
loc15878_pos3	0.1	0.9	0.4	0.6	0	1	0.13	0.87	0	1	0.02	0.98	0	1
loc16738_pos9	0.05	0.95	0	1	0	1	0	1	0.02	0.98	0.02	0.98	0.29	0.71
loc17339_pos94	0.8	0.2	0.93	0.07	0.33	0.67	1	0	0.77	0.23	0.9	0.1	0.87	0.13
loc18008_pos21	0.19	0.81	0.11	0.89	0.55	0.45	0.1	0.9	0.08	0.92	0.07	0.93	0.23	0.77
loc19281_pos19	0.08	0.92	0.03	0.97	0.03	0.97	0	1	0.02	0.98	0	1	0.46	0.54
loc19318_pos5	0.09	0.91	0	1	0.5	0.5	0.1	0.9	0	1	0.22	0.78	0	1
loc19505_pos21	0.08	0.92	0.02	0.98	0	1	0.04	0.96	0	1	0	1	0.44	0.56
loc19585_pos6	0.01	0.99	0	1	0	1	0	1	0	1	0	1	0.06	0.94
loc20654_pos103	0.1	0.9	0.02	0.98	0.4	0.6	0.02	0.98	0.03	0.97	0.05	0.95	0.08	0.92
loc20740_pos76	0.01	0.99	0	1	0	1	0	1	0	1	0	1	0.08	0.92
loc21032_pos1	0.18	0.82	0.05	0.95	0.65	0.35	0.05	0.95	0.08	0.92	0.13	0.87	0.1	0.9
loc3154_pos40	0.21	0.79	0.16	0.84	0.18	0.82	0.1	0.9	0.05	0.95	0.1	0.9	0.77	0.23
loc3186_pos57	0.01	0.99	0.09	0.91	0	1	0	1	0	1	0	1	0	1
loc3351_pos64	0.09	0.91	0	1	0.33	0.67	0	1	0.1	0.9	0	1	0.12	0.88
loc3665_pos20	0.02	0.98	0	1	0	1	0	1	0	1	0	1	0.12	0.88
loc3715_pos23	0.08	0.92	0	1	0.05	0.95	0	1	0	1	0	1	0.4	0.6
loc3731_pos91	0.21	0.79	0.03	0.98	0.04	0.96	0	1	0.34	0.66	0.03	0.97	0.65	0.35
loc4432_pos9	0.01	0.99	0	1	0	1	0.02	0.98	0	1	0.02	0.98	0.06	0.94
loc457_pos25	0.81	0.19	0.73	0.27	0.65	0.35	0.92	0.08	1	0	0.9	0.1	1	0
loc4679_pos15	0.23	0.77	0.34	0.66	0.17	0.83	0.07	0.93	0.42	0.58	0.28	0.72	0.1	0.9
loc4809_pos37	0.02	0.98	0.1	0.9	0	1	0	1	0	1	0	1	0	1
loc4874_pos1	0.33	0.67	0.2	0.8	0.35	0.65	0.47	0.53	0.24	0.76	0.47	0.53	0.17	0.82
loc5178_pos8	0.12	0.88	0.08	0.92	0.02	0.98	NA	NA	0.27	0.73	0	1	0.04	0.96
loc5273_pos20	0.3	0.7	0.23	0.77	0.05	0.95	0.08	0.92	0.26	0.74	0.2	0.8	0.75	0.25
loc5661_pos87	0.1	0.9	0.15	0.85	0	1	0.13	0.87	0.06	0.94	0.02	0.98	0.28	0.72
loc7221_pos24	0.31	0.69	0.24	0.76	0.37	0.63	0.3	0.7	0.4	0.6	0.32	0.68	0.23	0.77
loc7222_pos62	0.31	0.69	0.23	0.77	0.37	0.63	0.3	0.7	0.4	0.6	0.32	0.68	0.22	0.78
loc7550_pos61	0.04	0.96	0	1	0	1	0	1	0	1	0	1	0.33	0.67
loc7635_pos4	0.07	0.93	0	1	0.37	0.63	0.03	0.97	0	1	0.05	0.95	0	1
loc7736_pos16	0.13	0.87	0	1	0	1	0.35	0.65	NA	NA	0.17	0.83	NA	NA
loc8013_pos88	0.28	0.72	0.33	0.67	0	1	0.12	0.88	0.37	0.63	0.16	0.84	0.73	0.27
loc8100_pos94	0.59	0.41	0.65	0.35	0.42	0.58	0.6	0.4	0.79	0.21	0.4	0.6	0.67	0.33
loc8366_pos58	0.03	0.97	0.02	0.98	0	1	0	1	0	1	0	1	0.22	0.78
loc8428_pos7	0.09	0.91	0.06	0.94	0	1	0.13	0.87	0.08	0.92	0.15	0.85	0.12	0.88
loc8823_pos30	0.01	0.99	0	1	0	1	0	1	0	1	0	1	0.08	0.92

Table S10: Associated genes, mutation types, and effects for *N*=61 locally adapted outlier loci found within Neosho Madtom. The columns include: the identity of each locus based on IPYRAD output (SNP ID); the chromosome and the position that the loci map to on the Neosho Madtom genome; the reference and alternative allele (Ref / Alt); the gene associated with the locus; the type of mutation represented; and mutation effect. LOW = within coding region but no amino acid change (*N*=2); MODERATE = amino acid change within coding region (*N*=6); MODIFIER = non-coding region change (*N*=53).

	GENETIC INFORMATION						
SNP ID	**Chromosome**	**Position**	**Ref**	**Alt**	**Assoc. Gene**	**Mutation Type**	**Mutation Effect**
loc1767_pos35	JAFFST010008648.1	14674	G	A	aim1	synonymous_variant	LOW
loc4736_pos21	JAFFST010044978.1	13394	T	C	LOC108270301	5_prime_UTR_premature_start_codon_gain_variant	LOW
loc10680_pos97	JAFFST010049281.1	14636	G	A	immt / rnf103	missense_variant&splice_region_variant	MODERATE
loc14172_pos82	JAFFST010052226.1	48079	T	A	LOC108280674	missense_variant	MODERATE
loc19359_pos6	JAFFST010054742.1	2383	G	C	LOC108277774	missense_variant	MODERATE
loc19822_pos71	JAFFST010054999.1	175362	G	A	LOC108259564 / bmx	missense_variant,downstream_gene_variant	MODERATE
loc3028_pos42	JAFFST010035704.1	11766	T	C	atp13a2 / ca144	missense_variant,downstream_gene_variant	MODERATE
loc6005_pos68	JAFFST010045898.1	1140	G	A	LOC108269087	missense_variant	MODERATE
loc10014_pos12	JAFFST010048778.1	35638	G	C		intergenic_region	MODIFIER
loc10145_pos17	JAFFST010048839.1	19414	C	A		intergenic_region	MODIFIER
loc10283_pos65	JAFFST010049010.1	27513	T	A		upstream_gene_variant,downstream_gene_variant,intron_variant	MODIFIER
loc10344_pos4	JAFFST010049061.1	16031	C	A	usp28	intron_variant,non_coding_transcript_variant	MODIFIER
loc10406_pos90	JAFFST010049086.1	364029	A	G		upstream_gene_variant,downstream_gene_variant,intron_variant	MODIFIER
loc1164_pos70	JAFFST010003372.1	3451	T	C		intergenic_region	MODIFIER
loc12218_pos99	JAFFST010050525.1	126770	G	A		intergenic_region	MODIFIER
loc12475_pos17	JAFFST010050722.1	552772	G	A		downstream_gene_variant,intergenic_region	MODIFIER
loc12552_pos53	JAFFST010050738.1	812144	C	T	samd11	downstream_gene_variant,intergenic_region	MODIFIER
loc12940_pos3	JAFFST010050970.1	142372	T	C		upstream_gene_variant,downstream_gene_variant,intergenic_region	MODIFIER
loc13356_pos33	JAFFST010051448.1	12905	C	A		intergenic_region	MODIFIER
loc13414_pos61	JAFFST010051514.1	5484	A	C		intergenic_region	MODIFIER
loc13508_pos8	JAFFST010051641.1	22266	G	A		intron_variant,non_coding_transcript_variant	MODIFIER
loc13549_pos67	JAFFST010051701.1	34004	A	G		intergenic_region	MODIFIER
loc1408_pos102	JAFFST010006197.1	13649	C	G		intergenic_region	MODIFIER
loc1421_pos123	JAFFST010006223.1	48649	A	C	eftud2	intron_variant,non_coding_transcript_variant	MODIFIER
loc14568_pos18	JAFFST010052435.1	89434	T	A	c5h9orf3	intron_variant,non_coding_transcript_variant	MODIFIER
loc14964_pos1	JAFFST010052698.1	89803	T	A	fabp6 / pwwp2a	upstream_gene_variant,downstream_gene_variant,intergenic_region	MODIFIER
loc15878_pos3	JAFFST010053317.1	15964	G	A		intergenic_region	MODIFIER
loc16738_pos9	JAFFST010053725.1	4567	A	G	heatr5a	intron_variant,non_coding_transcript_variant	MODIFIER
loc17339_pos94	JAFFST010053949.1	310117	A	G		intergenic_region	MODIFIER
loc18008_pos21	JAFFST010054135.1	295290	T	C	cers1	intron_variant,non_coding_transcript_variant	MODIFIER
loc19281_pos19	JAFFST010054710.1	30753	C	A	atg13 / harbi1	upstream_gene_variant,downstream_gene_variant,intron_variant	MODIFIER
loc19318_pos5	JAFFST010054727.1	193662	G	C	lpcat1	intron_variant,non_coding_transcript_variant	MODIFIER
loc19505_pos21	JAFFST010054809.1	83143	T	C		intergenic_region	MODIFIER
loc19585_pos6	JAFFST010054854.1	64225	G	T	efnb2	intron_variant,non_coding_transcript_variant	MODIFIER
loc20654_pos103	JAFFST010055524.1	484382	T	C	fmn2	intron_variant,non_coding_transcript_variant	MODIFIER
loc20740_pos76	JAFFST010055634.1	32010	G	A		intergenic_region	MODIFIER
loc21032_pos1	JAFFST010055760.1	56696	A	T		upstream_gene_variant,downstream_gene_variant,intergenic_region	MODIFIER
loc3154_pos40	JAFFST010040856.1	5305	G	A		intergenic_region	MODIFIER
loc3186_pos57	JAFFST010040894.1	7266	G	A		intergenic_region	MODIFIER

Table S10
Cont'd

SNP ID	Chromosome	Position	Ref	Alt	Assoc. Gene	Mutation Type	Mutation Effect
loc3351_pos64	JAFFST010041121.1	135229	C	G		intergenic_region	MODIFIER
loc3665_pos20	JAFFST010041624.1	56855	C	T	rac3 / prkar1a	upstream_gene_variant,downstream_gene_variant,intergenic_region	MODIFIER
loc3715_pos23	JAFFST010041708.1	20807	T	C		intron_variant,non_coding_transcript_variant	MODIFIER
loc3731_pos91	JAFFST010041732.1	11042	A	C	cpped1	intergenic_region	MODIFIER
loc4432_pos9	JAFFST010044786.1	5565	T	C	fgb	intron_variant,non_coding_transcript_variant	MODIFIER
loc457_pos25	JAFFST010000946.1	80920	A	T		intergenic_region	MODIFIER
loc4679_pos15	JAFFST010044940.1	41797	G	C		downstream_gene_variant,intergenic_region	MODIFIER
loc4809_pos37	JAFFST010045047.1	51242	G	A		intron_variant,non_coding_transcript_variant	MODIFIER
loc4874_pos1	JAFFST010045080.1	102519	A	G	nkd2	intron_variant,non_coding_transcript_variant	MODIFIER
loc5178_pos8	JAFFST010045277.1	14311	T	C	slc52a3	downstream_gene_variant,intergenic_region	MODIFIER
loc5273_pos20	JAFFST010045337.1	84211	C	G	slc2a9	intron_variant,non_coding_transcript_variant	MODIFIER
loc5661_pos87	JAFFST010045657.1	39838	G	A	nptn	intron_variant,non_coding_transcript_variant	MODIFIER
loc7221_pos24	JAFFST010046671.1	16269	T	G		intergenic_region	MODIFIER
loc7222_pos62	JAFFST010046671.1	36648	C	A		intergenic_region	MODIFIER
loc7550_pos61	JAFFST010046947.1	1496	C	T	ptpre	intergenic_region	MODIFIER
loc7635_pos4	JAFFST010047012.1	2266	T	C		intergenic_region	MODIFIER
loc7736_pos16	JAFFST010047116.1	9993	G	A	htr1e	downstream_gene_variant,intergenic_region	MODIFIER
loc8013_pos88	JAFFST010047270.1	15917	A	G	ubap1l	intron_variant,non_coding_transcript_variant	MODIFIER
loc8100_pos94	JAFFST010047357.1	160182	G	T		intron_variant,non_coding_transcript_variant	MODIFIER
loc8366_pos58	JAFFST010047594.1	272111	T	C	cerk / gramd4	intergenic_region	MODIFIER
loc8428_pos7	JAFFST010047619.1	514451	T	C		intergenic_region	MODIFIER
loc8823_pos30	JAFFST010047844.1	57814	C	A		intergenic_region	MODIFIER

SUPPLEMENTAL FIGURES

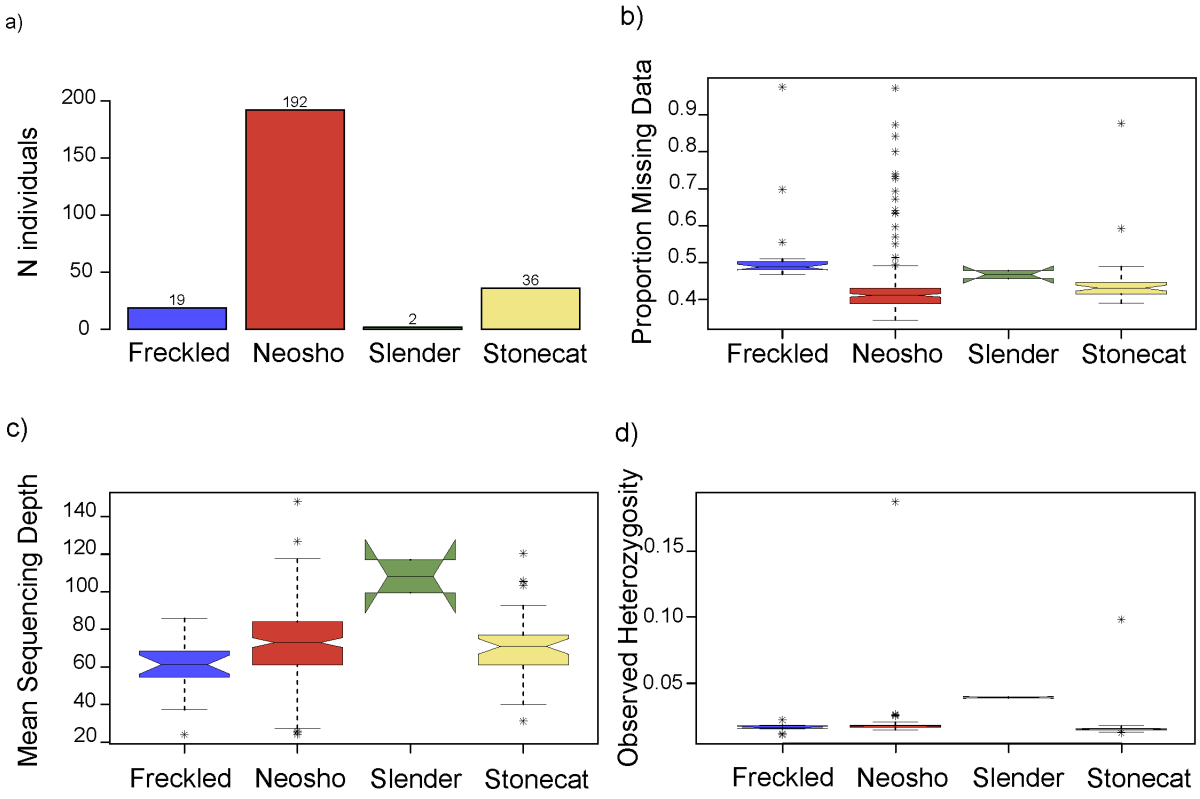

Figure S1: Summary information for four species of madtom evaluated in this study for potential hybridization: Freckled Madtom, Neosho Madtom, Slender Madtom, and Stonecat. (a) Number of individuals collected and genotyped per species (note: the main focus of the study was Neosho Madtom population genetics); (b) Proportion of missing genotype data in the multi-species alignment; (c) Mean sequencing depth per locus per individual; (d) Mean observed heterozygosity per individual.

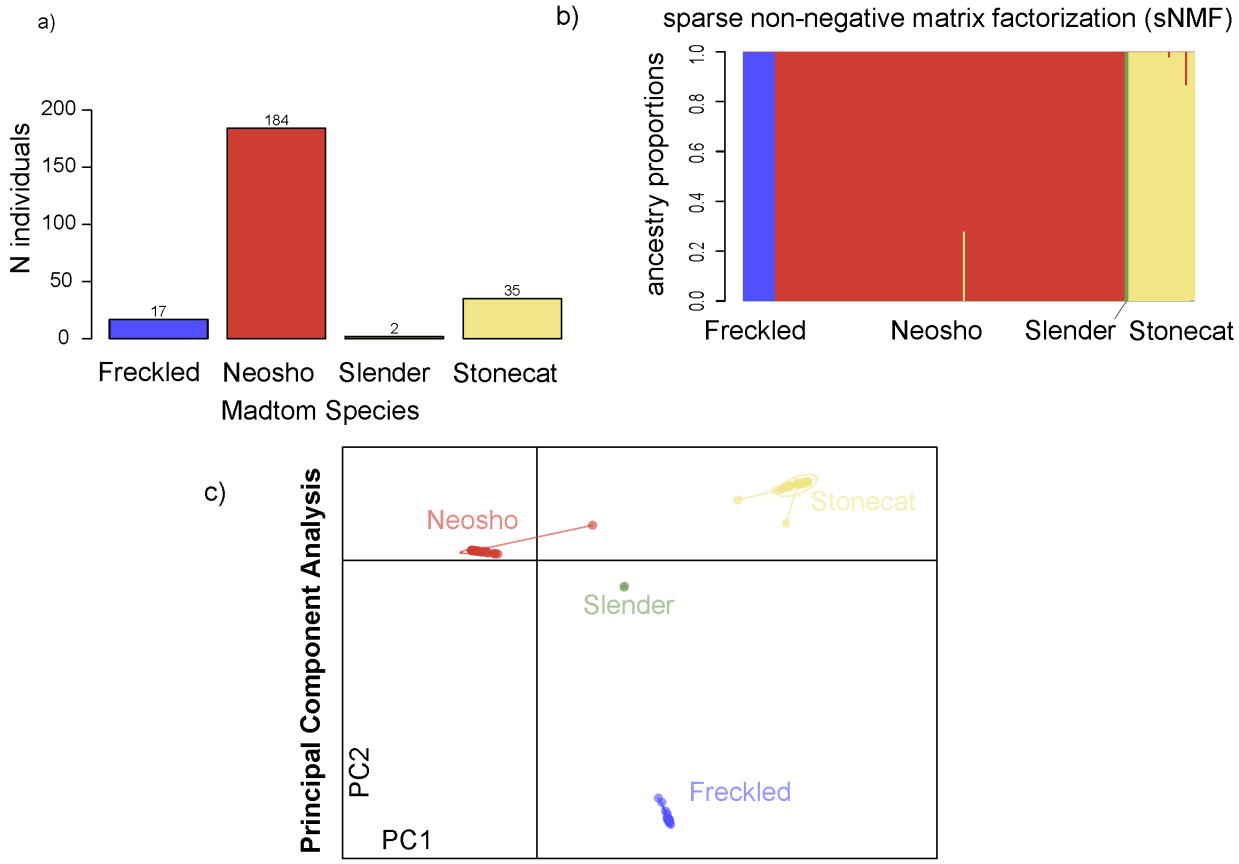

Figure S2: Results from analyses used to screen for admixed individuals (hybrids) in four species of madtom: Freckled, Neosho, Slender, and Stonecat. (a) Number of individuals of each species included in the multi-species loci alignment; (b) Sparse Non-Negative Matrix Factorization (sNMF) showing ancestry proportions of individuals (vertical lines). Admixture is indicated when a vertical line reflects multiple ancestries (colors). Admixture is only deduced if minor ancestry is at least 10%, whereas lower percentages are inferred as 'noise' in the data. (c) Principal Components Analysis (PCA) of genetic variance with individuals plotted as points. In PCA graphs, admixture is indicated when an individual (point) is 'pulled' away from its corresponding species cluster and towards another species cluster.

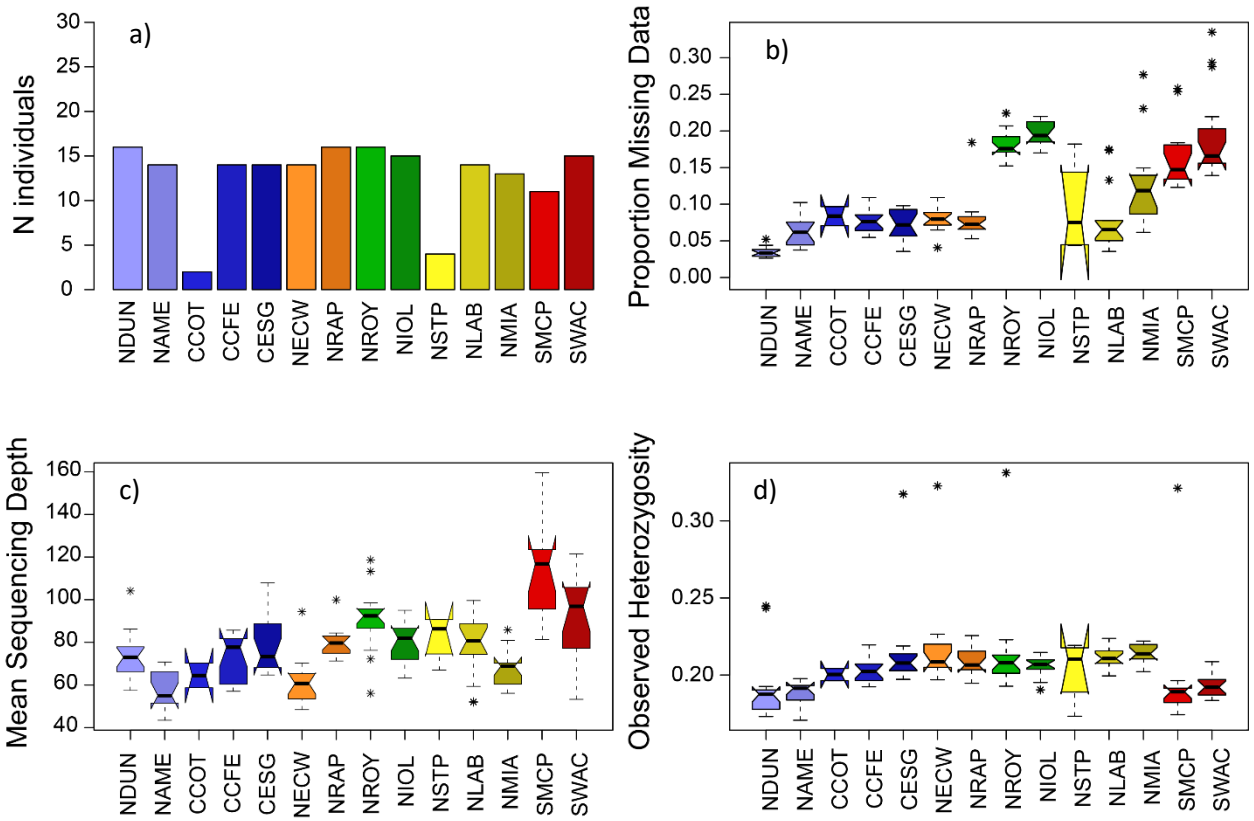

Figure S3: Summary of single nucleotide polymorphism (SNP) information by sampling site following standard data filtering (*N*=2,725 SNPs). These genetic variants/loci were used to assess population structure in Neosho Madtom (NMT). Sites are ordered based on the longitudinal gradient of the river and colored by management unit/population. (a) Number of NMT analyzed per site; (b) Proportion of missing genotype data in the alignment; (c) Mean sequencing depth per locus per individual; (d) Mean observed heterozygosity per individual as an indicator of genetic diversity.

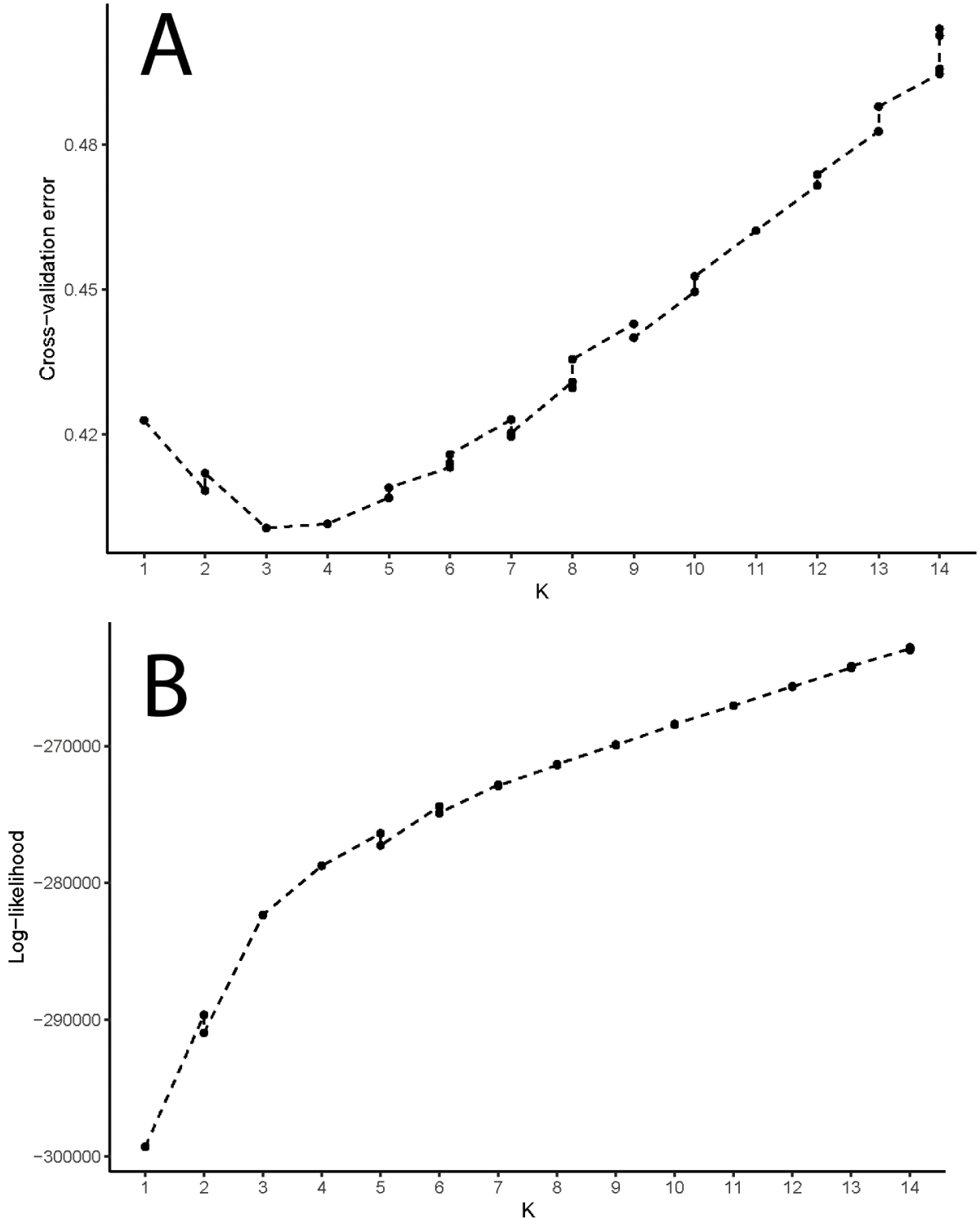

Figure S4: ADMIXTURE analysis (A) cross-validation error among *K* clusters and (B) log-likelihood measures among *K* clusters. The analysis is based on *N*=178 individuals collected at 14 sites, and SNP genotyped (*N*=2,725). The 'best' *K* is interpreted as the 'elbow' in either plot where the error is minimized (A) or log-likelihood is maximized (B).

RiverATLAS Attributes (version 1.0)

(click hyperlinked ID to jump to individual information sheet)

ID	Category	Attribute	Source Data	Citation	Column(s)	Count
H01	Hydrology	Natural Discharge	WaterGAP v2.2	Döll et al. 2003	dis_m3_---	x3
H02	Hydrology	Land Surface Runoff	WaterGAP v2.2	Döll et al. 2003	run_mm_---	x1
H03	Hydrology	Inundation Extent	GIEMS-D15	Fluet-Chouinard et al. 2015	inu_pc_---	x6
H04	Hydrology	Limnicity (Percent Lake Area)	HydroLAKES	Messager et al. 2016	lka_pc_---	x2
H05	Hydrology	Lake Volume	HydroLAKES	Messager et al. 2016	lkv_mc_---	x1
H06	Hydrology	Reservoir Volume	GRanD v1.1	Lehner et al. 2011	rev_mc_---	x1
H07	Hydrology	Degree of Regulation	HydroSHEDS & GRanD	Lehner et al. 2011	dor_pc_---	x1
H08	Hydrology	River Area	HydroSHEDS & WaterGAP	Lehner & Grill 2013	ria_ha_---	x2
H09	Hydrology	River Volume	HydroSHEDS & WaterGAP	Lehner & Grill 2013	riv_tc_---	x2
H10	Hydrology	Groundwater Table Depth	Global Groundwater Map	Fan et al. 2013	gwt_cm_---	x1
P01	Physiography	Elevation	EarthEnv-DEM90	Robinson et al. 2014	ele_mt_---	x4
P02	Physiography	Terrain Slope	EarthEnv-DEM90	Robinson et al. 2014	slp_dg_---	x2
P03	Physiography	Stream Gradient	EarthEnv-DEM90	Robinson et al. 2014	sgr_dk_---	x1
C01	Climate	Climate Zones	GEnS	Metzger et al. 2013	clz_cl_---	x1
C02	Climate	Climate Strata	GEnS	Metzger et al. 2013	cls_cl_---	x1
C03	Climate	Air Temperature	WorldClim v1.4	Hijmans et al. 2005	tmp_dc_---	x16
C04	Climate	Precipitation	WorldClim v1.4	Hijmans et al. 2005	pre_mm_---	x14
C05	Climate	Potential Evapotranspiration	Global-PET	Zomer et al. 2008	pet_mm_---	x14
C06	Climate	Actual Evapotranspiration	Global Soil-Water Balance	Trabucco & Zomer 2010	aet_mm_---	x14
C07	Climate	Global Aridity Index	Global Aridity Index	Zomer et al. 2008	ari_ix_---	x2
C08	Climate	Climate Moisture Index	WorldClim & Global-PET	Hijmans et al. 2005	cmi_ix_---	x14
C09	Climate	Snow Cover Extent	MODIS/Aqua	Hall & Riggs 2016	snw_pc_---	x15
L01	Landcover	Land Cover Classes	GLC2000	Bartholomé & Belward 2005	glc_cl_---	x1
L02	Landcover	Land Cover Extent	GLC2000	Bartholomé & Belward 2005	glc_pc_---	x44
L03	Landcover	Potential Natural Vegetation Classes	EarthStat	Ramankutty & Foley 1999	pnv_cl_---	x1
L04	Landcover	Potential Natural Vegetation Extent	EarthStat	Ramankutty & Foley 1999	pnv_pc_---	x30
L05	Landcover	Wetland Classes	GLWD	Lehner & Döll 2004	wet_cl_---	x1
L06	Landcover	Wetland Extent	GLWD	Lehner & Döll 2004	wet_pc_---	x22
L07	Landcover	Forest Cover Extent	GLC2000	Bartholomé & Belward 2005	for_pc_---	x2
L08	Landcover	Cropland Extent	EarthStat	Ramankutty et al. 2008	crp_pc_---	x2
L09	Landcover	Pasture Extent	EarthStat	Ramankutty et al. 2008	pst_pc_---	x2
L10	Landcover	Irrigated Area Extent (Equipped)	HID v1.0	Siebert et al. 2015	ire_pc_---	x2
L11	Landcover	Glacier Extent	GLIMS	GLIMS & NSIDC 2012	gla_pc_---	x2
L12	Landcover	Permafrost Extent	PZI	Gruber 2012	prm_pc_---	x2
L13	Landcover	Protected Area Extent	WDPA	IUCN & UNEP-WCMC 2014	pac_pc_---	x2
L14	Landcover	Terrestrial Biomes	TEOW	Dinerstein et al. 2017	tbi_cl_---	x1
L15	Landcover	Terrestrial Ecoregions	TEOW	Dinerstein et al. 2017	tec_cl_---	x1
L16	Landcover	Freshwater Major Habitat Types	FEOW	Abell et al. 2008	fmh_cl_---	x1
L17	Landcover	Freshwater Ecoregions	FEOW	Abell et al. 2008	fec_cl_---	x1
S01	Soils & Geology	Clay Fraction in Soil	SoilGrids1km	Hengl et al. 2014	cly_pc_---	x2
S02	Soils & Geology	Silt Fraction in Soil	SoilGrids1km	Hengl et al. 2014	slt_pc_---	x2
S03	Soils & Geology	Sand Fraction in Soil	SoilGrids1km	Hengl et al. 2014	snd_pc_---	x2
S04	Soils & Geology	Organic Carbon Content in Soil	SoilGrids1km	Hengl et al. 2014	soc_th_---	x2
S05	Soils & Geology	Soil Water Content	Global Soil-Water Balance	Trabucco & Zomer 2010	swc_pc_---	x14
S06	Soils & Geology	Lithological Classes	GLiM	Hartmann & Moosdorf 2012	lit_cl_---	x1
S07	Soils & Geology	Karst Area Extent	Rock Outcrops v3.0	Williams & Ford 2006	kar_pc_---	x2
S08	Soils & Geology	Soil Erosion	GloSEM v1.2	Borrelli et al. 2017	ero_kh_---	x2
A01	Anthropogenic	Population Count	GPW v4	CIESIN 2016	pop_ct_---	x2
A02	Anthropogenic	Population Density	GPW v4	CIESIN 2016	ppd_pk_---	x2
A03	Anthropogenic	Urban Extent	GHS S-MOD v1.0 (2016)	Pesaresi & Freire 2016	urb_pc_---	x2
A04	Anthropogenic	Nighttime Lights	Nighttime Lights v4	Doll 2008	nli_ix_---	x2
A05	Anthropogenic	Road Density	GRIP v4	Meijer et al. 2018	rdd_mk_---	x2
A06	Anthropogenic	Human Footprint	Human Footprint v2	Venter et al. 2016	hft_ix_---	x4
A07	Anthropogenic	Global Administrative Areas	GADM v2.0	University of Berkeley 2012	gad_id_---	x1
A08	Anthropogenic	Gross Domestic Product	GDP PPP v2	Kummu et al. 2018	gdp_ud_---	x3
A09	Anthropogenic	Human Development Index	HDI v2	Kummu et al. 2018	hdi_ix_---	x1
	Total	**Variables: 56**			**Attributes:**	**281**

Figure S5: Copy from the manual of the HydroATLAS v.0.1 database (Linke et al., 2019). Variables from each of the five categories were reduced using robust principal components analysis to test for natural selection.

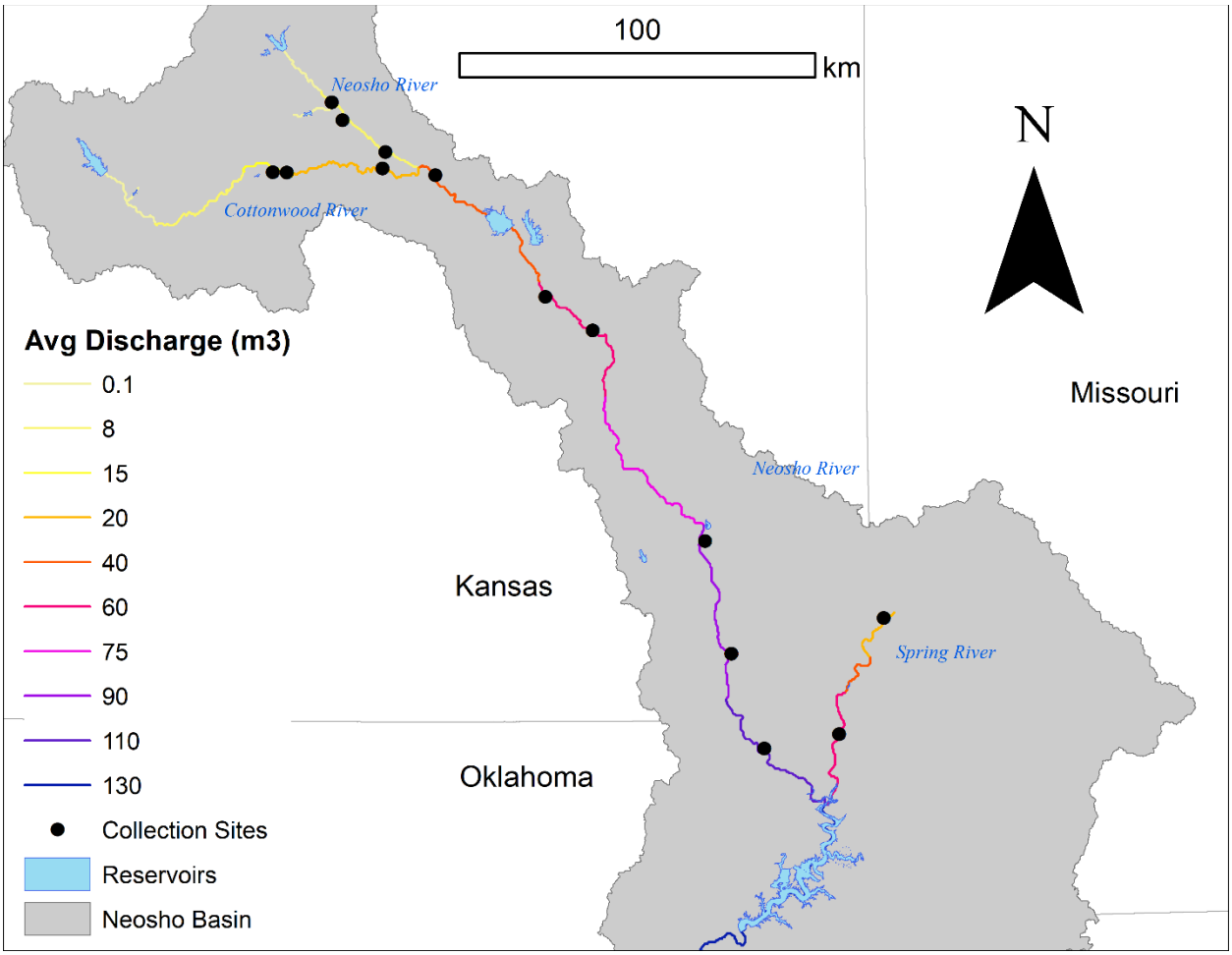

Figure S6: Map of the Neosho River system and the gradient of average annual discharge (cubic meters) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "hydro-physiographic" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Along the longitudinal gradient of a river network, there are many intertwined changes in environmental characteristics (i.e., the river continuum hypothesis). Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

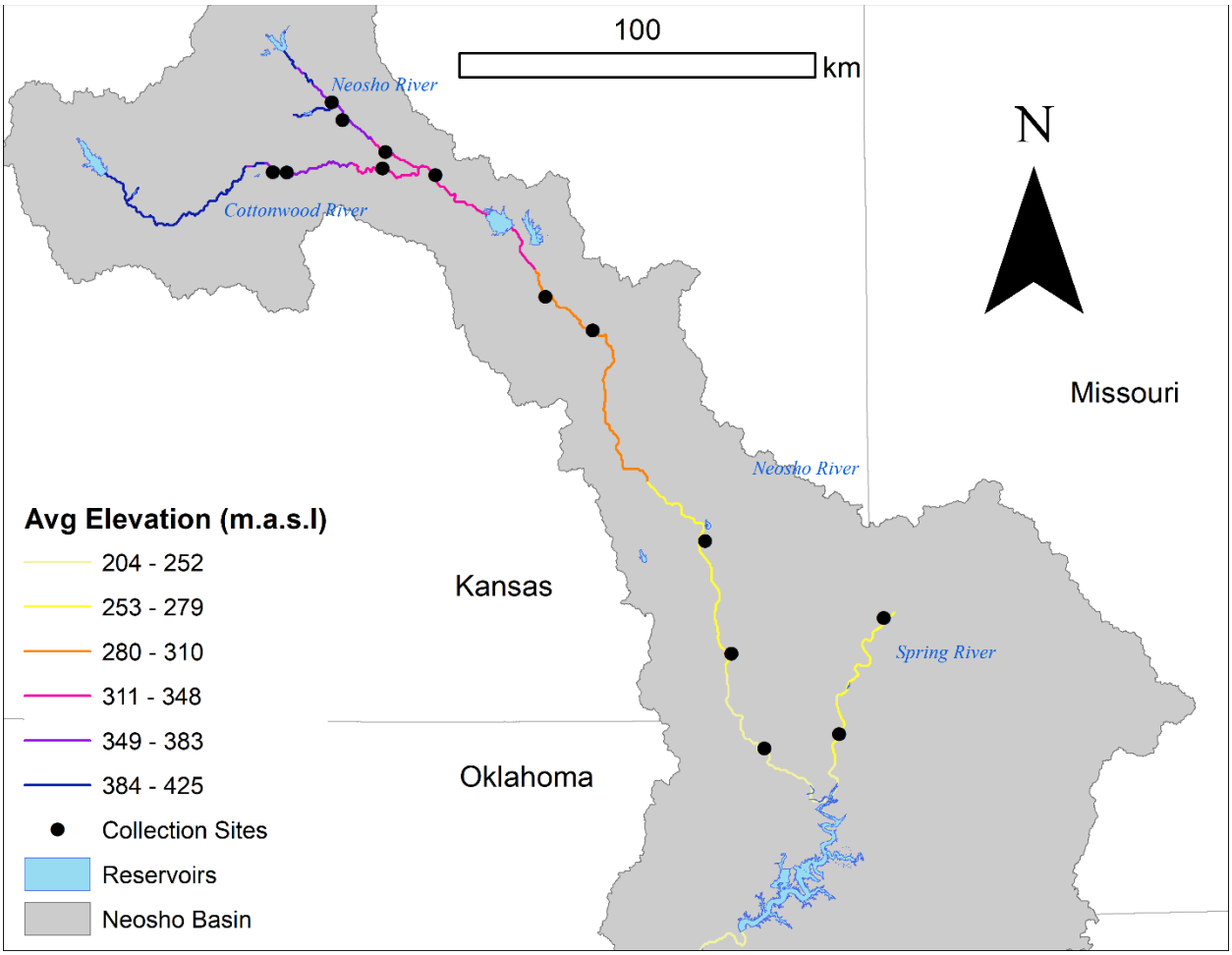

Figure S7: Map of the Neosho River system and the gradient of average elevation (meters above sea level) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "hydro-physiographic" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Naturally, elevation is inversely related to the direction of flow and discharge. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

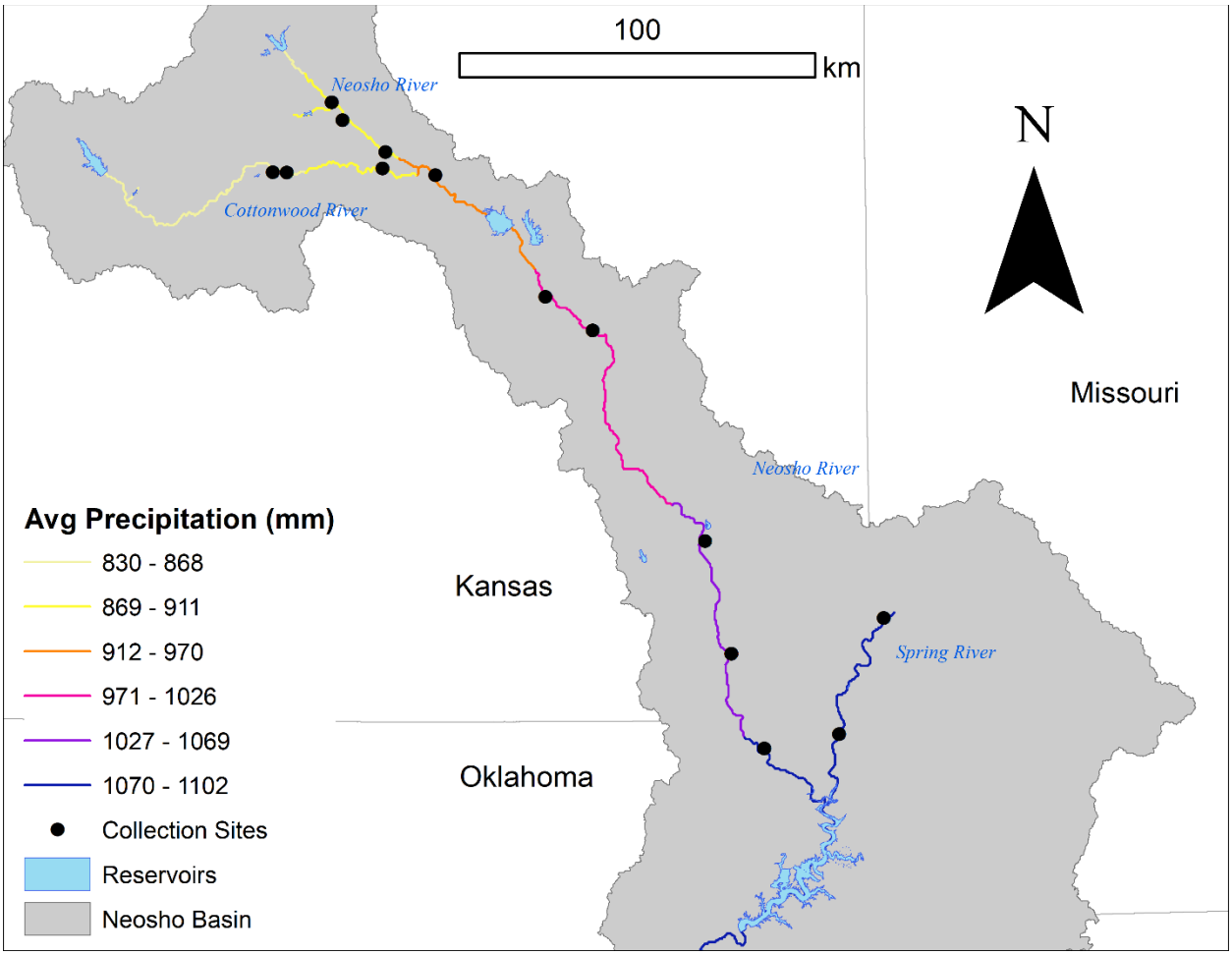

Figure S8: Map of the Neosho River system and the gradient of average annual precipitation (millimeters) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "climate" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Precipitation increases in the southeastern direction. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

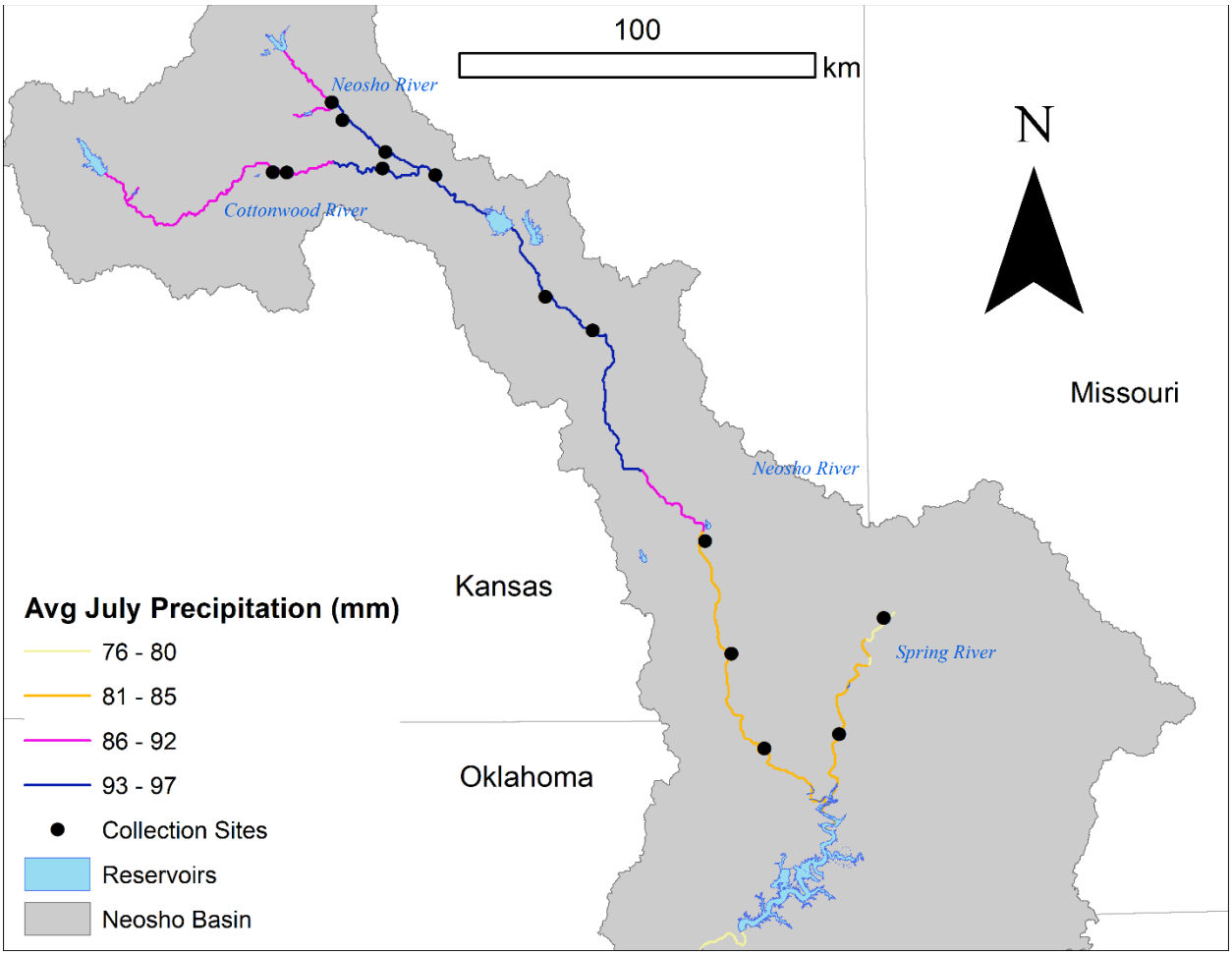

Figure S9: Map of the Neosho River system and the gradient of average July precipitation (millimeters) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "climate" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

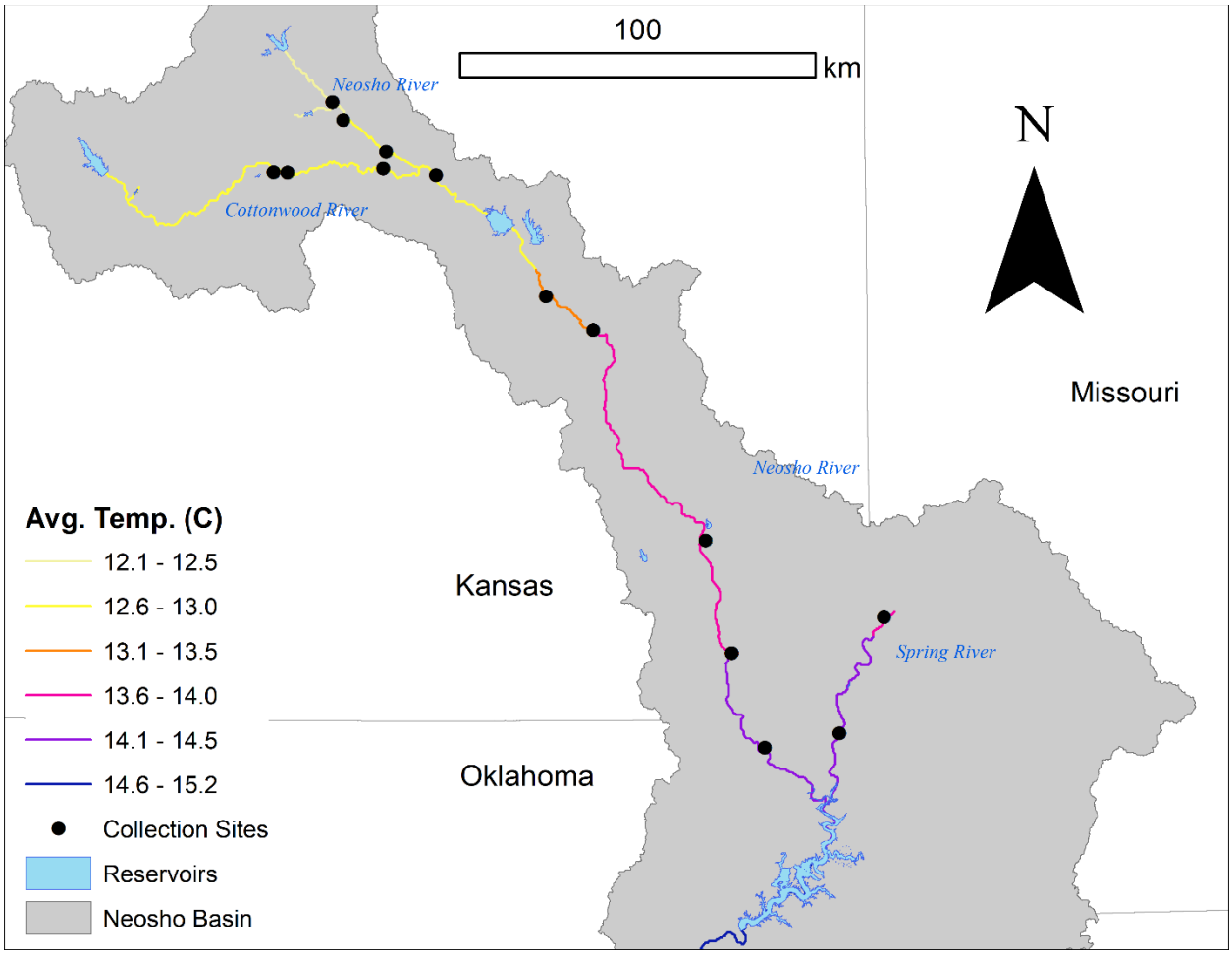

Figure S10: Map of the Neosho River system and the gradient of average annual air temperature (degrees Celsius) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "climate" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Temperature increases in the southeastern direction. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

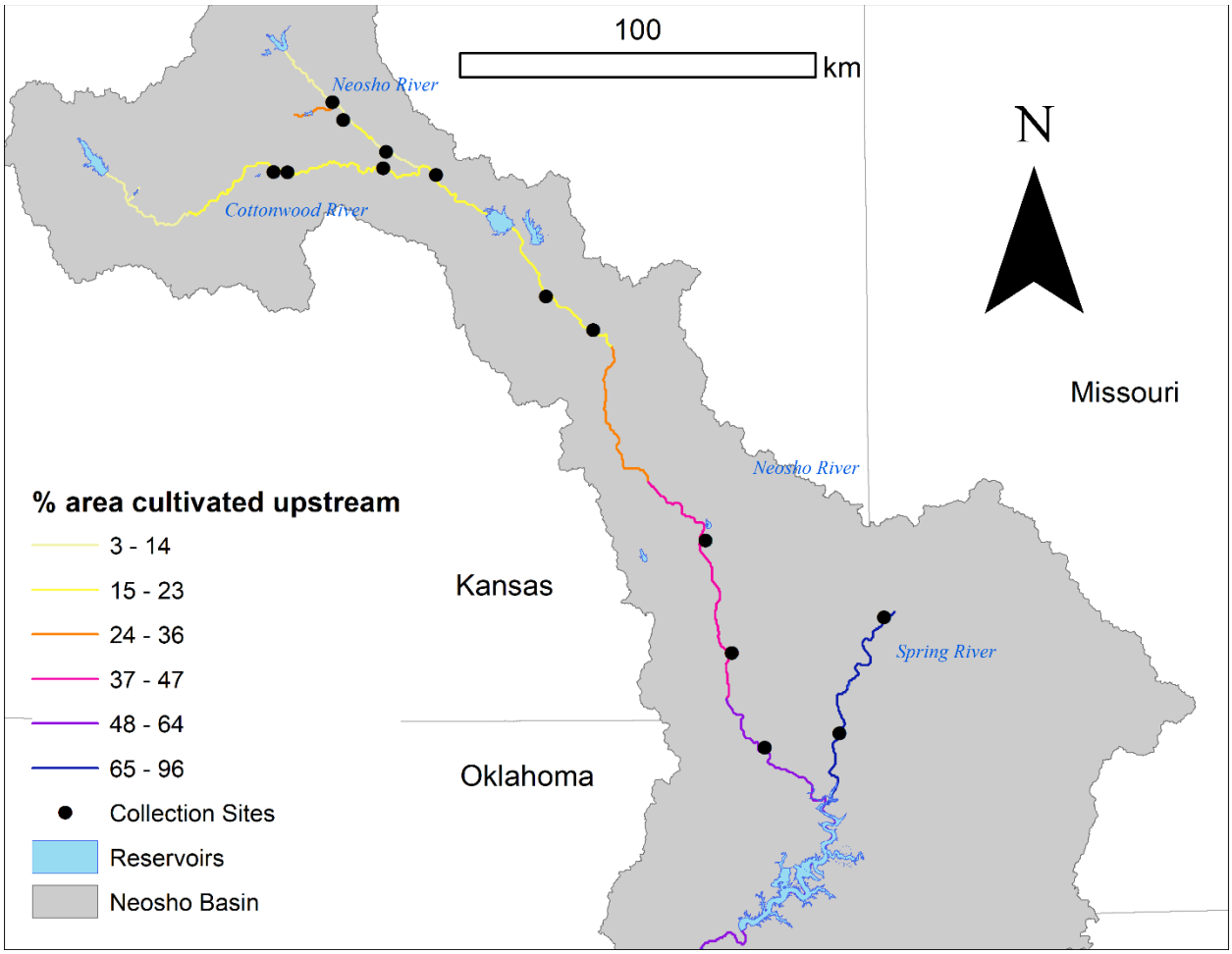

Figure S11: Map of the Neosho River system and the gradient of cultivated area in the total upstream catchment of a stream reach (percent of spatial extent) (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "landcover" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

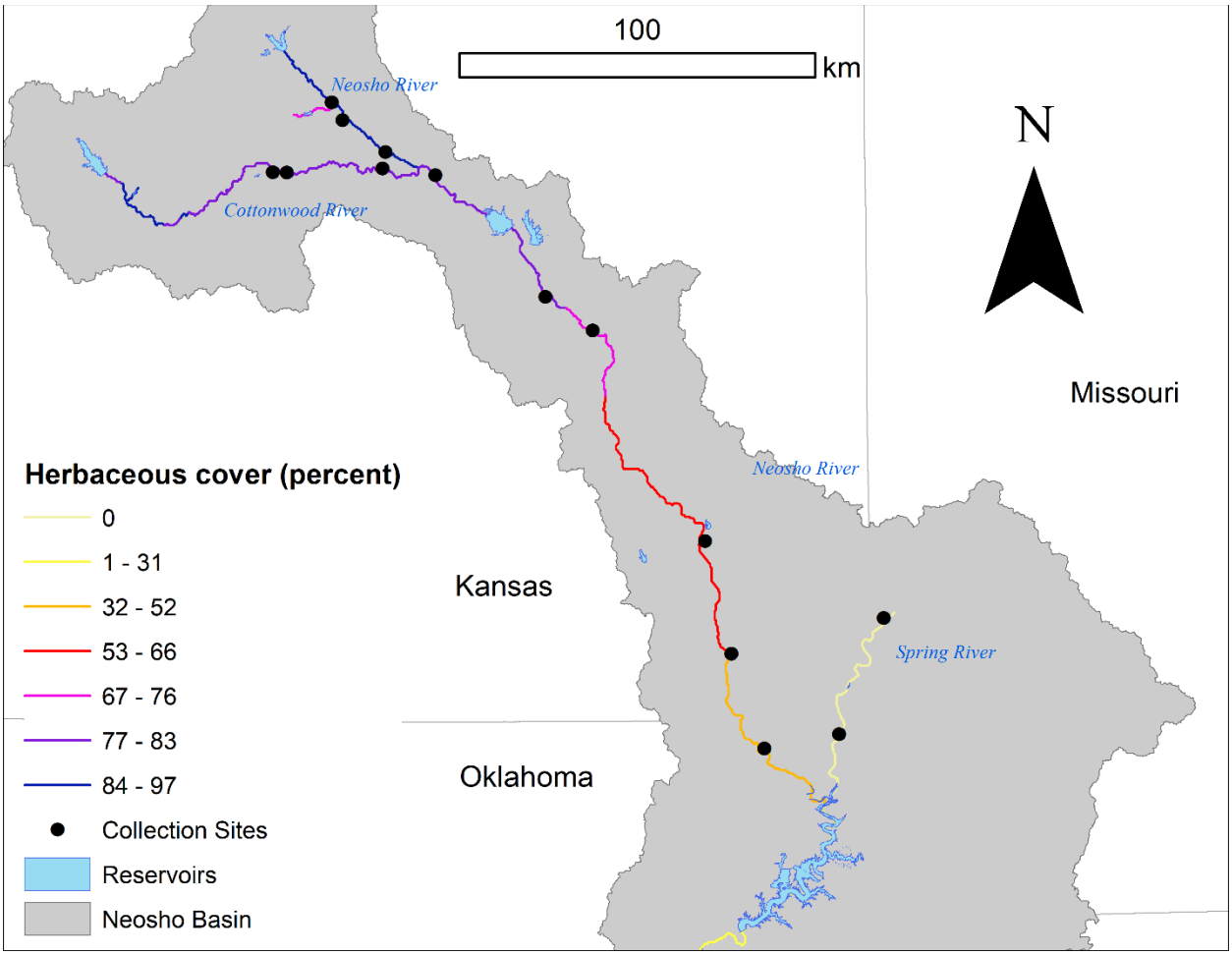

Figure S12: Map of the Neosho River system and the gradient of herbaceous cover (percent spatial extent) at individual river reach catchments (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "landcover" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Herbaceous cover decreases in the southeasterly direction, where landcover is more dominated by forests and cultivated land. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

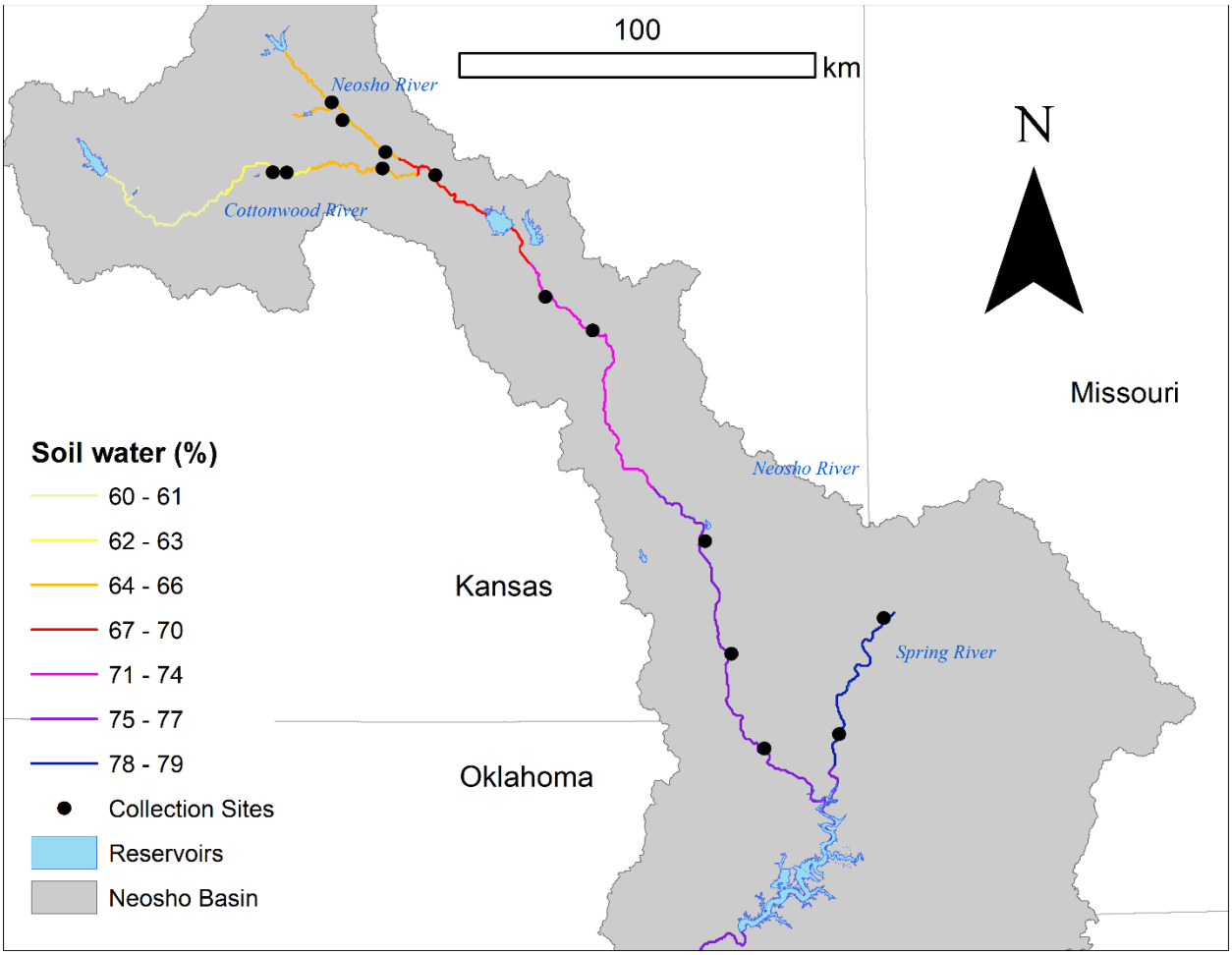

Figure S13: Map of the Neosho River system and the gradient of average water content in soil (percent) at individual river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "soil/geology" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Soil water content corresponds with the gradient in average annual precipitation. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.

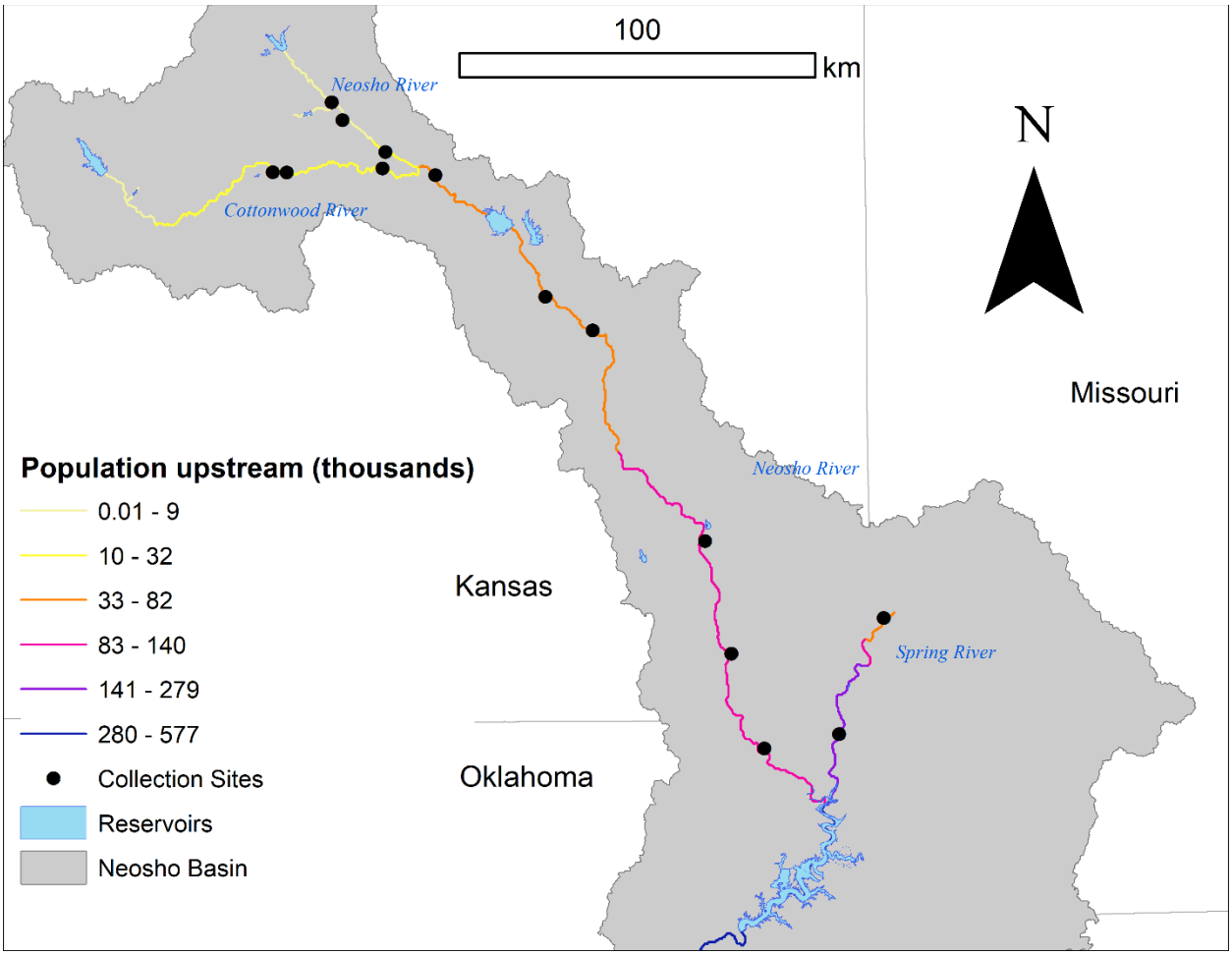

Figure S14: Map of the Neosho River system and the gradient of the human population (in thousands) in the total upstream catchment of river reaches (HydroATLAS v.0.1; Linke et al., 2019). This variable represents the "anthropogenic" composite variable generated via robust principal components analysis and used to screen for adaptive loci candidates. Neosho Madtom was collected from the collection sites (black circles) and SNP genotyped.