8-2022

Riverscape Community Genomics of Ozark Fishes: A Comparative Framework to Infer Ecological and Evolutionary Determinants of Genetic Diversity

Zachery D. Zbinden
University of Arkansas, Fayetteville

Riverscape Community Genomics of Ozark Fishes:
A Comparative Framework to Infer Ecological and Evolutionary Determinants of Genetic Diversity

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biology

by

Zachery D. Zbinden
University of Pittsburgh
Bachelor of Science in Ecology and Evolution, 2013
University of Oklahoma
Master of Science in Biology, 2016

August 2022
University of Arkansas

The dissertation is approved for recommendation to the Graduate Council.

Marlis R. Douglas, Ph.D.
Dissertation Director

Michael E. Douglas, Ph.D.
Dissertation Co-Director

Jeremy M. Beaulieu, Ph.D.
Committee Member

Adam M. Siepielski, Ph.D.
Committee Member

ABSTRACT

Genetic variation is a crucial component of biodiversity and represents the variability and spatial structure of alleles within and among organisms. Evolution modulates this variability over time through mutation, selection, gene flow, and genetic drift. However, our capacity to test foundational theories of population genetics has always been at the mercy of molecular approaches available to quantify patterns of genetic diversity. Initially, techniques for empirical DNA studies were in their infancy and limited by technologies and the price per unit of genetic information. Because of these constraints, our pursuits have generally been limited to investigations of one or a few species simultaneously, hampering our power to draw broadly applicable conclusions. Advances in molecular technologies, e.g., high-throughput sequencing, now provide so much information at so little cost that a multispecies comparative approach to uncovering generalities about evolution is within reach even for applied studies on non-model organisms. Ultimately, genotyping individuals from all species within a community will be feasible and easily replicated across sampling locations and span entire regions. Variability of genetic diversity, within and among species, can be leveraged to explore the relationship between ecology and evolution and between micro- and macroevolution.

For my dissertation research, I employed a multispecies framework to link ecological and evolutionary processes driving spatial patterns of biodiversity through comparative analyses of genotypic variability among sympatric species of freshwater fish that inhabit a large sub-basin of the Mississippi River. First, I quantified the extent of admixed ancestry among species within a community by assessing genomic variability among individuals from many species. My analyses uncovered that fish in nature — particularly minnows — have higher than expected hybridization rates. My data even show evidence of hybrid viability and genetic exchange among

species (i.e., introgression). I interpret these findings of widespread admixture among distinct species as an indicator that admixture plays a critical role in ecology and evolution – more so than previously considered.

Second, I tested for general mechanisms that define spatial genomic variability within species by comparing models of extrinsic drivers of genetic divergence. The river network, or stream hierarchy model, best explained species' genomic variability, as evidenced by the correspondence between genetic divergence and riverine architecture. This general pattern emerged for all species, but the degree of genetic divergence differed widely, indicating that the intrinsic traits of each species may also play an important role.

Finally, I further explored how phenotypic traits may modulate species' genetic diversity and ultimately evolutionary trajectories by comparing relationships between traits and metrics of genetic variability among species within a comparative framework. Significant associations between trait values and genetic patterns emerged, allowing me to develop predictive models of genetic diversity using traits alone, without requiring direct genetic assessments. These trait-based models can be applied to prioritize species for conservation and management.

My dissertation research demonstrates that modern molecular approaches are uniquely positioned to help unite ecology and evolution, bridging the long-standing dichotomy between these two disciplines. I provide a comparative framework for conservation biology that integrates various temporal and spatial scales and demonstrate with an empirical example how it can be applied to assess thousands of informative genetic markers across entire communities of non-model organisms. My dissertation research elevates population genomics to the community level and outlines how to explore new dimensions in our long-standing inquiry: *What drives variation in genetic diversity among species*?

ACKNOWLEDGEMENTS

I want to thank my advisors, Drs. Marlis and Michael Douglas for supporting me throughout my time at the University of Arkansas and trusting me to produce valuable work with their endowment funds. My written work has benefitted from their careful revisions and keen attention to narrative.

Drs. Tyler Chafin, Bradley Martin, Steve Mussmann, and Max Bangs were instrumental in developing the laboratory and bioinformatic logistics required for my work and were always eager to engage in discussion and provide help. I also acknowledge Dr. Chafin for joining me for two field seasons to collect fish across the Ozarks.

I am grateful to the University of Arkansas professors whose courses helped shape my thinking and develop my capacity as a scientist: Drs. William Etges, Jeremy Beaulieu, Adam Siepielski, Andrew Alverson, and Pawel Wolinski.

I thank Mason Flurry, McKenna George, Tim Goodhart, Kevin Hollar, and Mackenzie Reed, who assisted with thousands of DNA extractions.

This work was made possible by funding and resources provided by the University of Arkansas, including the Arkansas High-Performance Computing Center, which provided analytical resources. I thank Biological Sciences staff Viola Phillips, Jan Warren, and Becky Harris for their support. Several fellowships and endowments provided funding: the University of Arkansas Distinguished Doctoral Fellowship, the Harry and Jo Leggett Chancellor's Fellowship, the Bruker Professorship in Life Sciences (Marlis Douglas), and the 21st Century Chair in Global Change Biology (Michael Douglas).

DEDICATION

To my mother, for accepting nothing short of my best

TABLE OF CONTENTS

LIST OF PUBLISHED PAPERS

Chapter I:

Zbinden ZD, Douglas MR, Chafin TK, Douglas ME. 2022. A community genomics approach to natural hybridization. *Proceedings of the Royal Society B: Biological Sciences*. Submitted.

Chapter II:

Zbinden ZD, Douglas MR, Chafin TK, Douglas ME. 2022. Riverscape community genomics: A comparative analytical framework to identify common drivers of genetic structure. *Molecular Ecology*. Submitted.

Chapter III:

Zbinden ZD, Douglas MR, Chafin TK, Douglas ME. 2022. Linking traits with evolutionary drivers within freshwater fish communities of the Ozark Highlands. *Molecular Ecology*. Submitted.

INTRODUCTION

The ongoing *biodiversity crisis* is fueled by widespread habitat alteration and massive over-exploitation of ecosystem services (Díaz et al., 2019; Millenium Ecosystem Assessment, 2005; Singh, 2002), and freshwater ecosystems are at the forefront of this crisis. Despite their tremendous economic, cultural, and scientific value (Albert et al., 2021), humans degrade riverine systems more rapidly than terrestrial counterparts through the compounding impacts of water over-allocation, pollution, habitat fragmentation, and species introductions. Ultimately, negative impacts on our rivers do not bode well for us because *we all live downstream* (Rabalais, 2003).

Freshwater fishes serve as indicators of riverine ecosystem health, and species differ in their physical characteristics, resource requirements, and habitat preferences (Helfman et al., 2009). Fishes are an essential part of riverine ecosystems, which humans rely on for basic needs such as food and water, and for sustaining our modern lifestyle via hydropower (Auerbach et al., 2014). Freshwater fisheries are of immense economic importance (Harper et al., 2013; Lynch et al., 2016). Given our reliance on freshwater fishes and their ecosystems, species loss and ecosystem collapse are not something we should risk. Unfortunately, the current position of freshwater conservation — underfunded and neglected by apathy — requires a more strategic approach to combatting biodiversity loss.

Constraints imposed by limited conservation funding necessitate predictive schemes to prioritize conservation and management targets. Debates on whether resources should be applied towards threatened species — which may be doomed to extinction regardless — or focused instead on maintaining ecosystems and key species underscore the current dilemma (Wilson et

al., 2011). Conservation and management must transition from a reactive science to a more proactive one, i.e., “new conservation” (Marvier, 2014; Sullivan et al., 2014). Proactivity requires the ability to forecast biodiversity changes across time and space (Urban et al., 2016). Thus, a new framework is needed to define long-term conservation goals and identify appropriate management strategies to sustain biodiversity in a changing world. One avenue to provide more management insight is examining the ecological drivers of genetic diversity to discern how contemporary processes influence long-term evolutionary trajectories.

Freshwater fish diversity and river networks

Out of over 30,000 fish species recognized, almost 50% inhabit freshwaters, despite these ecosystems comprising < 1% of the liquid water available (Tedesco et al., 2017). Given the much larger habitat volume of marine ecosystems, a more proportional number of species would be expected (Hubbell, 2001; MacArthur & Wilson, 1967). This discrepancy is deemed the “freshwater paradox” (Tedesco et al., 2017), hereafter “FP.” Unraveling the processes that have led to the FP is central to conserving freshwater diversity because it requires the integration of ecology and evolution: what contemporary processes most influence persistence through time? My work aims to explain why freshwaters are teeming with so many species in hopes that knowledge can help protect them.

The explanation for FP has focused on the higher potential for geographic isolation across freshwater habitats (Grosberg et al., 2012). Most freshwater fish inhabit riverine rather than lacustrine environments (Miller, 2021). Rivers are networks structured fractally, comprising repeating units of small tributaries, mid-sized streams, and large primary outflows. Many smaller river networks combine into more extensive ones and range from local catchments to

continental-scale basins such as the Amazon or Mississippi rivers. Considerable distances and environmental gradients can separate populations and communities in different network branches, making them comparable to island habitats (Tedesco et al., 2012). The network architecture makes riverine ecosystems isolating environments conducive to population divergence (Dias et al., 2013).

Additionally, the structure of basins and networks can be modified through geological processes, such as erosion, glaciation, and uplift, all of which can alter the courses of rivers, change the direction of flow, and re-arrange basins. As a result, isolated streams in adjacent basins become connected, bringing populations or species into contact and promoting genetic mixing that otherwise would not have been possible (Musher et al., 2022). Alternatively, connected stream segments once harboring a single population can be rerouted and separated, leading to two or more isolated populations. These paleohydrology re-arrangements occur iteratively over and leave an imprint on the evolutionary histories of stream-dwelling organisms (Mayden, 1985, 1988; Strange & Burr, 1997).

The isolating nature of river networks may be related to why freshwaters contain so many fish species (Dias et al., 2013; Oberdorff et al., 2011; Tedesco et al., 2012). The process of isolation and resulting demographic independence of populations has two essential outcomes: 1) Populations established through colonizing a river network trade fewer migrants than in other habitats (e.g., landscapes or oceanscapes); and 2) Populations with low gene flow begin to diverge genetically from others due to their independence. Genetic changes in an isolated population are not exchanged with other populations (and *vice versa*), and as a result, isolated populations become increasingly unique over time due to evolutionary independence.

While isolation can lead to evolutionary independence and speciation, it can also result in ecological collapse and extinction (Hanski, 1997); this underscores why unraveling the FP is central to conserving biodiversity. The trajectory of a population towards becoming a new, independent lineage through spatial isolation is precarious. Unraveling the processes that have led to the FP is a task set in river networks concerned with understanding how species' extrinsic and intrinsic ecological properties modulate population persistence and evolutionary change. A more precise understanding of the relationship between ecological factors and evolutionary trajectory can guide conservation efforts but requires a multispecies comparative approach.

Comparative Multispecies Framework

A comparative multispecies framework is required to unravel the FP. Studies of genetic variation in natural populations have primarily been limited to one or a few species. However, the high-throughput capacity of the massive-parallel next-generation sequencing approaches has made it possible to sequence hundreds or even thousands of species in a single study (Singhal et al., 2018). Analyses of such comprehensive data sets would be decisive in linking processes happening at the micro-evolutionary level (e.g., gene flow) to patterns observed at the macro-evolutionary scale (e.g., speciation).

The call for comparative genetic approaches to examine multiple, non-model species has come from various disciplines to gain insights into diverse processes from community ecology to diversification. For example, “community genetics” focuses on the genetic consequences of species interactions (Whitham et al., 2006). “Landscape community genomics” extends this and considers the additional aspects of spatial scale and landscape heterogeneity (Hand et al., 2015). These sub-fields could arguably fall under the much older “comparative population genetics”

(Hoenigsberg et al., 1969). Still, newer terms like “macrogenetics” are being coined for comparative population genetic studies focused on broad spatial or temporal scales (Blanchet et al., 2017).

Nevertheless, multiple areas of evolutionary inquiry have been converging on the same idea, which underscores the potential of such broad-scale, comparative investigations. Ultimately, a comparative approach seeks to yield generalized insight into the ecological and evolutionary drivers of diversity. Here I appropriate “-scape” community genomics (Hand et al., 2015) for my overall framework for studying the genetic diversity of multiple species collected across a basin: ‘Riverscape community genomics.’ I explore hybridization among them, common extrinsic spatial drivers of genetic structure, and how traits unique to each species that mediate their ecology ultimately influence their evolutionary trajectories.

Dissertation Objectives

My dissertation explores what we can infer about ecological and evolutionary processes by simultaneously investigating genetic diversity for dozens of co-distributed freshwater fish within a river basin. I present three examples of how a multispecies approach to genetic diversity can lead to greater insight applied to fisheries management.

Chapter I leverages signatures of genetic diversity unique to species to identify admixed individuals, i.e., hybridization or introgression. While identifying admixed individuals is necessary before further analyses of genetic diversity. Admixture can have negative consequences by reducing fitness (outbreeding depression), and it is thus of interest in the applied context of fisheries management. However, admixture can also provide a source of genetic diversity if hybrids are viable and backcross with pure individuals, thus facilitating the

exchange of genes between distinct species (introgression). This may lead to novel variation and alter evolutionary trajectories (Taylor & Larson, 2019). My research showed hybridization to be widespread among minnows and that genetic exchange among species via introgression is also common.

Chapter II focuses on the isolating nature of river networks. I examine differences in intraspecific genetic diversity among sampling locations, compare models of spatial genetic structure, and repeat this for dozens of fish species. This framework allows concordant patterns of spatial genetic structure to emerge, revealing spatial scales most relevant for managing and maintaining genetic diversity.

Finally, Chapter III explores the link between phenotypic traits that mediate ecological interactions and evolutionary history inferred from genetic diversity. Here I define genetic diversity broadly for each species as total diversity (γ), average within-site diversity (α), and among-site diversity (β) (Whittaker, 1960, 1972). For multiple fish species, I explore how variation in these genetic diversity metrics is associated with phenotypic characteristics and construct predictive models that aim to predict genetic diversity from traits alone. These insights offer avenues for new conservation strategies that leverage generalized ecological traits to forecast diversity metrics relevant to evolutionary trajectories in thousands of freshwater species currently lacking genetic information. These predictions can guide management and prioritize species conservation efforts, particularly in developing countries where funding, infrastructure, and training constraints limit pro-active management programs.

REFERENCES

Albert, J.S., Destouni, G., Duke-Sylvester, S.M., Magurran, A.E., Oberdorff, T., Reis, R.E., … & Ripple, W.J. (2021). Scientists' warning to humanity on the freshwater biodiversity crisis. *Ambio*, *50*(1), 85–94.

Auerbach, D.A., Deisenroth, D.B., McShane, R.R., McCluney, K.E., & Poff, L.N. (2014). Beyond the concrete: Accounting for ecosystem services from free-flowing rivers. *Ecosystem Services, 10*, 1–5.

Blanchet, S., Prunier, J.G., & De Kort, H. (2017). Time to go bigger: Emerging patterns in macrogenetics. *Trends in Genetics, 33*(9), 579–580.

DeWoody, J.A., Harder, A.M., Mathur, S., & Willoughby, J.R. (2021) The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147-4154

Dias, M.S., Cornu, J.F., Oberdorff, T., Lasso, C.A., & Tedesco, P.A. (2013). Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography, 36*(6), 683–689.

Grosberg, R. K., Vermeij, G. J., & Wainwright, P. C. (2012). Biodiversity in water and on land. *Current Biology, 22*(21), R900-R903.

Hand, B.K., Lowe, W.H., Kovach, R.P., Muhlfeld, C.C., & Luikart, G. (2015). Landscape community genomics: Understanding eco-evolutionary processes in complex environments. *Trends in Ecology and Evolution, 30*(3), 161–168.

Hanski, I. (1997). Metapopulation dynamics: from concepts and observations to predictive models. In I. Hanski & M. Gilpin (Eds.), *Metapopulation biology: Ecology, genetics and evolution* (pp. 69–91). Academic Press.

Harper, S., Zeller, D., Hauzer, M., Pauly, D., & Sumaila, U.R. (2013). Women and fisheries: Contribution to food security and local economies. *Marine Policy, 39*(1), 56–63.

Helfman, G.S., Collette, B.B., Facey, D.E., & Bowen, B.W. (2009). *The diversity of fishes: Biology, evolution, and ecology*. John Wiley & Sons.

Hoenigsberg, H.F., Castro, L.E., Granobles, L.A., & Idrobo, J.M. (1969). Population genetics in the American Tropics. II. The Comparative genetics of Drosophila in European and Neo-Tropical environments. *Genetica, 40*, 43–60.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. *Monographs in Population Biology*. Princeton University Press.

Lynch, A.J., Cooke, S.J., Deines, A.M., Bower, S.D., Bunnell, D.B., Cowx, I.G., … & Beard, T.D. (2016). The social, economic, and environmental importance of inland fish and

fisheries. *Environmental Reviews*, *24*(2), 115–121.

MacArthur, R.H., & Wilson, E.O. (1967). *The theory of island biogeography*. Princeton University Press.

Marvier, M. (2014) New conservation is true conservation. *Conservation Biology*, *28*(1), 1–3.

Mayden, R.L. (1985). Biogeography of Ouachita Highland fishes. *The Southwestern Naturalist*, *30*(2), 195–211.

Mayden, R.L. (1988). Vicariance biogeography, parsimony, and evolution in north american freshwater fishes. *Systematic Biology*, *37*(4), 329–355.

Millenium Ecosystem Assessment. (2005). *Ecosystems and Human Well-being : Current State and Trends*. Island Press.

Miller, E.C. (2021). Comparing diversification rates in lakes, rivers, and the sea. *Evolution*, *75*(8), 2055-2073.

Musher, L.J., Giakoumis, M., Albert, J., Del-Rio, G., Rego M., Thom, G., . . . Cracraft, J. (2022). River network rearrangements promote speciation in lowland Amazonian birds. *Science Advances*, *8*(14), eabn1088.

Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr, H.H. (2011). Global and regional patterns in riverine fish species richness: A review. *International Journal of Ecology*, *2011*, 967631.

Rabalais, N. (2003). We all live downstream (and upstream). *Journal of Soil and Water Conservation*, *58*(3), 52–53.

Singh, J.S. (2002). The biodiversity crisis: A multifaceted review. *Current Science*, *82*(6), 638–647.

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

Strange, R.M., & Burr, B.M. (1997). Intraspecific phylogeography of North American highland

fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, *51*(3), 885–897.

Sullivan, B.K., Douglas, M.R., Walker, J.M., Cordes, J.E., Davis, M.A., Anthonysamy, W.J.B., … & Douglas, M.E. (2014). Conservation and management of polytypic species: The little striped whiptail complex (Aspidoscelis inornata) as a case study. *Copeia*, *2014*(3), 519–529.

Taylor, S.A., & Larson, E.L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, *3*(2), 170–177.

Tedesco, P.A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H.H., Beauchard, O., … & Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography*, *21*(10), 977–987.

Tedesco, P.A., Paradis, E., Lévêque, C., & Hugueny, B. (2017). Explaining global-scale diversification patterns in actinopterygian fishes. *Journal of Biogeography*, *44*(4), 773–783.

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., Pe'er, G., Singer, A., … & Travis, J.M.J. (2016). Improving the forecast for biodiversity under climate change. *Science*, *353*, aad8466.

Whitham, T.G., Bailey, J.K., Schweitzer, J.A., Shuster, S.M., Bangert, R.K., Leroy, C.J., … Wooley, S.C. (2006). A framework for community and ecosystem genetics: From genes to ecosystems. *Nature Reviews Genetics*, *7*(7), 510–523.

Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*, 279–338.

Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, *21*, 213–251.

Wilson, H.B., Joseph, L.N., Moore, A.L., & Possingham, H.P. (2011). When should we save the most endangered species? *Ecology Letters*, *14*(9), 886–890.

CHAPTER I

A community genomics approach to natural hybridization

Zachery D. Zbinden, Marlis R. Douglas, Tyler K. Chafin, & Michael E. Douglas

ABSTRACT

Hybridization plays a pivotal role in ecology and evolution, and while once deemed unnatural and uncommon, it is now recognized as surprisingly ubiquitous between species. However, few studies have quantified its prevalence at the community level, where hybridization may modulate species interactions, change ecological processes, or alter evolutionary trajectories. We quantified rates of hybridization across 75 communities of freshwater fish within the White River Basin (AR/MO, USA) by genotyping 33 species (N=2,865 individuals) across thousands of loci using restriction-site associated DNA sequencing (ddRAD). In contrast to more conventional single species-pair evaluations, our approach is less biased on a per-species/per-individual basis because it does not require *a priori* assumptions of putative hybridization. We found widespread hybridization (N=70 individuals; 2.4%) across 18 parental species pairs and involving 73% of species (N=24) analyzed. Most occurrences were in minnows (Leuciscidae): N=15 species and N=66 individuals. Introgression was evident in 24 backcrossed individuals. Literature-based per-individual hybridization estimates predicted only 2-3 hybrids within our dataset. Our community-level assessment suggested hybridization is even more common than now-contemporary considerations and may substantially confound analytical frameworks of ecological and evolutionary studies and thus should not be disregarded.

INTRODUCTION

Hybridization, once considered “exceedingly rare” by zoologists (Hubbs, 1955), is now recognized as relatively common, primarily because of increased scrutiny due to modern DNA sequencing technologies (Cordonnier et al., 2019; Ottenburghs, 2019). A more contemporary paradigm is that hybridization and introgression play an important role in ecology and evolution (Lowe et al., 2015; Pfennig, 2021; Rhymer & Simberloff, 1996), which diverges from a historical perspective rooted in the biological species concept (Harrison & Larson, 2014).

Hybridization studies focus predominantly on single species pairs, emphasizing reproductive barriers and genomic consequences (Taylor & Larson, 2019). Thus, despite a growing list of species known to hybridize, we are still unaware of how frequently hybridization and introgression occur within communities. This deficiency extends beyond the academic: Admixture often complicates conservation and management (Bangs et al., 2020), inflates divergence estimates in population genomic evaluations (Arnold, 1992; Grant & Grant, 2019), and generally violates model assumptions underlying analytical frameworks (McVay et al., 2017). In short, admixture's ecological, evolutionary, and analytical ramifications remain nebulous without understanding how frequently it occurs in communities.

Several issues have clouded estimates of natural hybridization and introgression rates: First, a narrow focus on single species pairs and an *a priori* assumption of hybridization, e.g., based on morphological observation, have promoted ascertainment bias (Allendorf et al., 2001; Justyn et al., 2020), where ‘cryptic’ hybridization beyond a study’s purview often goes unrecognized (Justen et al., 2021; Vallender et al., 2007). This restricted focus is primarily due to the cost-prohibitive nature of multispecies evaluations. Mandates that stipulate funding must go towards select species, e.g., state- or federally-listed species, further limit the taxonomic scope of

hybridization assessments. Furthermore, detecting low-level admixture where backcrossing has diluted the contribution of ancestral genomes per individual requires sufficient sample sizes and a broad genomic scope (Randi, 2008) to contend with reduced detection probabilities of introgressed alleles (Gerard et al., 2011; Meng & Kubatko, 2009).

Broader comparative examinations often utilize meta-analyses (Justyn et al., 2020) to quantify hybridization rates per species. However, studies collated within meta-analyses to detect hybridization are rarely equivalent in statistical power (Schwenk et al., 2008), with negative results often ignored (Kotze et al., 2004). An unbiased examination of hybridization in nature is past due (Folk et al., 2018), and by determining its extent, we can discern its ecological and evolutionary significance.

Freshwater fishes are an excellent study system in that they demonstrate a greater propensity for hybridization than other vertebrates (Hubbs, 1955; Keck & Near, 2009; Scribner et al., 2000; Wallis et al., 2017). This seemingly stems from unequal abundances, secondary contact facilitated by stream capture, and weak reproductive isolating mechanisms, e.g., external fertilization and spatiotemporal overlap in spawning (Corush et al., 2021; Scribner et al., 2000). Furthermore, developmental incompatibilities are also reduced in fish (Thorgaard & Allendorf, 1988).

We quantified intrafamilial hybridization in a regional North American freshwater stream fish community by genotyping genome-wide SNP loci. We addressed the following: (i) What is the frequency of hybridization in nature, i.e., that proportion of admixed individuals across species and families? (ii) What are the rates of introgression among species? And (iii) Is the occurrence of hybridization related to environmental variation?

METHODS

Sampling and data generation

The sampling area spanned the White River Basin (71,911 km^2) of the Ozark Plateau, straddling southern Missouri and northern Arkansas. Sampling procedures were validated by the University of Arkansas Institutional Animal Care and Use Committee (IACUC #17077) and permitted by regional authorities (Supplement S1). Fish were captured using seine nets from June 2017 to September 2018 and euthanized via tricaine methanesulfonate (MS-222) and 95% ethanol. Species diagnosis occurred in the laboratory following Eschmeyer's Catalog of Fishes. For DNA analysis, the right pectoral fin was removed and stored in 95% ethanol at -20°C.

Genomic DNA was isolated (Qiagen Fast kits; Qiagen Inc.) and then quantified by fluorometry (Qubit; Thermo-Fisher Scientific). Individuals were genotyped using a reduced representation of the genome obtained via double-digest restriction site-associated DNA (ddRAD) sequencing, with procedures modified from previous protocols (Chafin et al., 2019; Peterson et al., 2012), and sequenced as pooled batches of 144 individuals per 1x100 lane on the Illumina HiSeq 4000 (Supplement S1).

Data processing and assembly

Before hybrids could be identified, family-level SNP panels had to be generated by demultiplexing, clustering, and filtering sequence data. Raw Illumina reads were demultiplexed and filtered in IPYRAD (Eaton & Overcast, 2020). First, family-level phylogenies were generated to verify that individuals were correctly identified to species using PHYLORAD (Fan et al., 2018) (an alignment and assembly-free method applied to raw reads; Supplement S1). Individuals (*N*=3,042) were subsequently partitioned by family (*N*=6) and processed *de novo* in IPYRAD

(Eaton & Overcast, 2020) to generate family-level assemblies. Adapters and primers were removed and reads with >5 low-quality bases (Phred<20) discarded. Clusters were assembled using an 85% identity threshold, with loci subsequently removed via conditional criteria to ensure high-quality data (Supplement S1). Biallelic SNPs were further filtered and visualized using RADIATOR (Gosselin, 2020; Supplement S1).

Hybrid identification

Our initial objectives were to detect individuals with admixed ancestry and determine rates of inter-familial hybridization (i.e., no hybridization between families assumed). Individuals were screened for admixture using principal components analysis (PCA) and discriminant analysis of principal components (DAPC; Jombart et al., 2010). An initial DAPC (ADEGENET; Jombart, 2008) explained 90% of the variance with *n*-1 discriminant functions, where *n*=number of species. An α-score optimization was implemented to avoid over/underfitting and acquire the optimum number of PCs, with individual assignment probabilities calculated via a second DAPC.

Results were contrasted with a maximum-likelihood clustering approach (SNAPCLUST; Beugin et al., 2018). Three models were tested: (i) F_1 only; (ii) F_1+ first-generation backcross; and (iii) F_1+first- and second-generation backcrosses with AIC as best fit (Beugin et al., 2018).

Admixture analysis employing sparse non-negative matrix factorization (sNMF) was used to refine ancestry coefficients for each family (LEA; Frichot & François, 2015), with 25 repetitions per *K* value (1–25) and regularization parameter (α)=100. Best *K* from each run (via cross-validation) was used to impute missing data (*impute* function, method='mode' in LEA)

and then repeat sNMF (as above). Mixed ancestry was assumed if assignment probability <0.9 for the primary cluster.

Hybrid classification

We contrasted models with and without introgression in SNAPCLUST to identify backcrossed individuals (Beugin et al., 2018; Cordonnier et al., 2019). We confirmed assignments by assigning individuals as F_1, F_2, or backcrosses (NEWHYBRIDS; Anderson & Thompson, 2002). We also implemented HYBRIDDETECTIVE (Wringe et al., 2017) to confirm if our statistical power was sufficient for assigning putative hybrids to genotype frequency classes. We first simulated the hybrids of known classes by randomly sampling two alleles per locus from appropriate parental pools, and three independent replicates assessed convergence. We then determined an optimal posterior probability threshold (=0.70) at which hybrid classes could be reliably assigned. The final MCMC was run for 1,000,000 iterations (250,000 burn-in), using a panel of 200 loci with the greatest among-species differentiation (F_{ST}) and lowest linkage disequilibrium (r^2<0.2).

Spatial analysis of hybrid occurrence

The occurrence of hybridization was evaluated at each sampling location (N=75) across the river network to evaluate spatial autocorrelation and relationships with environmental factors. We acquired a representative environmental dataset for North American rivers (HYDRORIVERSv.1.0; Linke et al., 2019), which we trimmed (ARCGIS; ESRIv.10.8) and then compiled with sampling site coordinates by nearest stream segment (SF; Pebesma, 2018).

Sampling location environmental variables (*N*=281) were partitioned into five classes: (i) Hydrology; (ii) Climate; (iii) Landcover; (iv) Geology; and (v) Anthropogenic. Variables lacking variation were removed, with the remainder being standardized (BESTNORMALIZE; Peterson & Cavanaugh, 2020). Finally, those exhibiting within-class collinearity were removed stepwise until each variation inflation factor <10 (USDM; Naimi, 2013). Dimensionality was reduced via robust PCA (ROBPCA; Hubert et al., 2005; Hubert et al., 2016), and the PCs were then tested for relationships with hybrid occurrence (multiple regression, VEGAN; Oksanen et al., 2020).

We tested spatial autocorrelation using two methods that generated spatial predictor variables: (i) Euclidean distance between all sample locations (latitude/ longitude) decomposed into positively correlated spatial eigenvectors (distance-based Moran's eigenvector maps, function *dbmem*; ADESPATIAL; Dray et al., 2020); (ii) Hydrologic network distance between all sampling locations decomposed as above. Network distances were calculated using RIVERDIST (Tyers, 2017). Both eigenvector sets were tested for relationships with hybrid occurrence using multiple regression (i.e., spatial autocorrelation).

RESULTS

Samples comprised 33 species collected from 75 locations (Figure 1). We examined admixed individuals across six families using SNP genotypes for 2,865 individuals (Table 1). The number of SNPs varied by family but generally was inversely correlated with the number of species (Table 1). Power analyses verified the robustness of panels for detecting hybrid individuals and classes (Supplement S2–S19).

Hybrid identification and classification

We found more evidence of admixture than expected, although most putative hybrids were cyprinids (minnows). Hybrids (*N*=70; 2.4%) were detected within four families (Table 2), with the proportion of hybrid individuals ranging from 0–4.4%. All but four hybrids were cyprinids (Leuciscidae), with exceptions being: (i) *Micropterus dolomieu* x *M. salmoides* (Centrarchidae); (ii) *Etheostoma juliae* x *E. zonale* (Percidae); (iii) *Etheostoma spectabile* x *E. caeruleum* (Percidae); and (iv) a putative multispecific hybrid: *Noturus maydeni* x *N. albater* x *N. exilis* (Ictaluridae). Thus, we recognized *N*=18 hybridizing species-pairs and *N*=8 hybridizing species-triplets, i.e., multispecies hybrids.

Within Leuciscidae, we identified 66 hybrids and 15 hybridizing species-pairs (Supplement S20). The most remarkable occurrence (*N*=29 individuals; 41%) involved *Campostoma anomalum* x *C. oligolepis*. At least one admixed individual was found in *N*=15 cyprinid species, whereas *N*=24 species contributed to at least one hybrid. Species sampled at more sites showed more evidence of hybridization: The number of hybrids per species was significantly related to the number of sample sites (R^2=0.16, F=6.1, p=0.02).

PCA, sNMF, and SNAPCLUST were largely congruent in detecting hybrids (figures 2-3). Both sNMF, and SNAPCLUST designated 36 hybrids but separately identified additional admixed individuals (26 and eight, respectively). DAPC performed poorly (Supplement S21), with only eight admixed individuals detected (admixture results: Supplement S22).

We also found evidence of introgression. SNAPCLUST only designated F_1 hybrids in all cases. However, NEWHYBRIDS identified 39% of hybrids as backcrossed (24/62, excluding eight multispecies hybrids); this corresponds to 55% (10/18) of all putatively hybridizing species-pairs showing evidence of introgressive hybridization (Table 3). Approximately 24% of hybrids

(15/62) were likely F_1 or F_2. Finally, NEWHYBRIDS designated 37% (23/62) of the putative hybrids as pure individuals, although they may be poorly classified introgressed individuals, i.e., late-generation hybrids (Beugin et al., 2018).

Spatial and environmental factors

At least one hybrid occurred at 56% of sampling sites (42/75), yet without significant spatial autocorrelation, based either on Euclidean distance (p=0.546) or river network distance (p=0.736). Likewise, hybrid incidence was not significantly related to environmental variation: Hydrology (p=0.321); Climate (p=0.737); Landcover (p=0.442); Geology (p=0.799); or Anthropogenic (p=0.266).

DISCUSSION

Hybridization plays a pivotal role in ecology and evolution because it can negatively impact the fitness of individuals, spread maladaptations across populations, drive species to extinction, and even generate new lineages through hybrid speciation (Arnold, 1992; Bangs et al., 2020; Lowe et al., 2015; Meng & Kubatko, 2009; Rhymer & Simberloff, 1996). While hybridization is now generally considered a “common” phenomenon, we still expect hybrid individuals to be rare in communities because of reproductive barriers (Cordonnier et al., 2019). Despite our expectations, the prevalence of hybrids at the community level is generally unknown because most hybridization studies focus on single species-pair evaluations or involve only a few closely related species within a specific taxonomic group (Bangs et al., 2020). This study outlines an approach to quantify hybridization at the community level that does not require *a priori* assumptions of putative hybridization and thus lacks a species-specific bias. We demonstrate this

approach by evaluating admixture across freshwater fish communities within the White River Basin of the Ozarks.

Frequency of hybridization in nature

Detection of admixture within 67% of families, involving 73% of species examined across our study underscores the potential relevance of hybridization in ecology and evolution. Hybridization rates were variable among families and species, reinforcing the importance of unbiased community-level estimation. Our conclusions regarding hybridization could drastically differ if only one species pair were randomly evaluated from this regional pool.

Hybridization was encountered more frequently than anticipated, whether per-individual (2.4%) or per-species (73%). Previous estimates of per-individual hybridization rates across plants and animals were lower (0.002–0.06%) (Justyn et al., 2020; Mallet, 2005; Mayr, 1963). Our results delineated substantially higher rates in cyprinids (4%). However, one could ask: Is this a genuine biological phenomenon or simply a sampling artifact? If the actual rate were comparable to that found in other studies (e.g., 0.1%), we would have expected evidence of only 2–3 hybrids. The estimates cited above stemmed primarily from studies relying on morphological identification of hybrids and thus may be less effective at detecting introgressed individuals (Keck & Near, 2009). Our study using genomic signatures was more adept at uncovering cryptic admixture not easily identified via phenotype (Justen et al., 2021).

Hybridization may be higher in fishes (Hubbs, 1955; Wallis et al., 2017). Similar per-individual rates (0–4%) were identified in marine fishes (Burford et al., 2011), with even higher rates (22.5%) among invasive Mississippi River carp (Lamer et al., 2010). However, fish hybridization studies typically focus on single species pairs, often within hybrid zones or

invasion fronts (Dowling & Hoeh, 1991; Matthews et al., 2016), and this may impede a strict comparison by biasing rates higher. Although, a typical study focused on a single species-pair is likely to miss hybrids resulting from one of the study species and a third, unstudied species; perhaps worse, such an individual might be erroneously attributed as a hybrid between the two focal species when it is not (more below).

Published per-species estimates vary from 1–10% across animals and ~25% for plants (Mallet, 2005; Twyford & Ennos, 2012). Although differences in the former are apparent from literature and museum records, one could again ask if it reflects an actual biological signal, or rather, variance in diversity and research effort? For example, Schwenk et al. (2008) found rates mostly homogenous among taxonomic groups once the number of species and records, i.e., research intensity, were accounted for. However, fishes again demonstrated higher rates (Schwenk et al., 2008).

Our per-species rate was highest within minnows (Leuciscidae) (Table 2), yet this was also the most specious group studied. Previous meta-analyses found disproportionately high hybridization rates within Leuciscidae (Hubbs, 1955; Corush et al., 2021), potentially indicating hybrids with greater viability (Avise & Gold, 1977). Not surprisingly, breeding behavior within Leuciscidae is also a significant predictor of hybridization rates (Corush et al., 2021). Many minnow species distribute gametes widely within the water column or affix them onto substrate/nests that are often a shared resource among species (Corush et al., 2021; Peoples et al., 2016). Early observations noted that six different minnow species simultaneously spawned over the same gravel substrate (Hubbs, 1955). In addition, minnows are not only the most diverse and widely distributed of North American stream fish families but are also numerically dominant members within communities (Matthews & Marsh-Matthews, 2017). As such, they encounter

broad environmental heterogeneity, demonstrate a high degree of sympatry, and exhibit uneven abundances among species, all of which may favor more frequent hybridization (Scribner et al., 2000).

Introgression between species

Not only did we find evidence of widespread hybridization among species, but also introgression. The evidence of introgression suggests hybrids can be viable, fertile, and backcross with parental species, thus facilitating inter-specific gene flow (Dowling & Demarais, 1993). This genetic exchange can provide a ready source of novelty for evolutionary forces to act upon (Twyford & Ennos, 2012), and the prevalence found here indicates how probable an extraneous genetic contribution to a species' evolutionary trajectory can be. In some instances, introgression between species has been suspected to result in the generation of new lineages, i.e., hybrid speciation (Meng & Kubatko, 2009; Taylor & Larson, 2019). It is intriguing that one of the most diverse groups of freshwater fish — cyprinids — is gaining a reputation for frequent hybridization and, as shown here, exchange of genes among species. This finding supports the notion that introgressive hybridization has played an important role in cyprinid evolutionary history (Dowling & Demarais, 1993). Questions remain regarding how adaptive such hybridization is and how it may differ among species and environmental contexts (DeSantis et al., 2021). Such data would facilitate our capacity to predict/mitigate species collapse (Gibson et al., 2019) and offer a potential avenue for future conservation and management.

Spatial distribution of hybridization

Hybridization has long been considered "unnatural" (Mallet, 2005), and this legacy interpretation now contrasts with the elevated detection of its occurrence. This recognition also increases concerns that anthropogenic impacts are homogenizing pre-zygotic barriers (Hubbs, 1955; Grabenstein & Taylor, 2018). Environmental extremes, such as those found in modified habitats, often grant selective benefits to hybrids (Arnold et al., 1999; Chafin et al., 2019; Hamilton et al., 2009). In some cases, these advantages promote them over parentals (Carson et al., 2012) because habitat alterations and fragmentations can change temperature, turbidity, and water chemistry such that species barriers weaken or dissipate completely (Chafin et al., 2019; Heath et al., 2010).

However, the incidence of hybridization we found at the community level did not associate with environmental factors. Therefore, we cannot reject the hypothesis of spatially random hybridization within the White River Basin. Alternatively, the underlying determinants may vary along a much finer spatial scale or within a species-specific environmental context. Our study drainage may be less anthropogenically impacted than other regions, an aspect supported by field observations and established indices (i.e., HYDRORIVERS; Linke et al., 2019). If so, a much larger spatial scope may be required to capture anthropogenic drivers of hybridization. Despite these caveats, our results offer a baseline for future investigations that test variability in community-wide hybridization.

Methodological considerations

Despite its wide application, our results suggest a limited capacity for DAPC to detect individuals of mixed ancestry. DAPC results are intuitive in that a discriminant analysis serves to

maximize between-species differences. However, maximizing cluster differences invariably produces a pattern wherein admixed individuals are likely to be assigned with a high probability to either parental cluster, especially when hybridization was not *a priori* suspected (i.e., not reflected in categorical group designations). Hence, its strength in minimizing overlap among species may dampen its utility for identifying admixture. In this sense, we merely document what has been suggested by *in silico* explorations (Miller et al., 2020).

Another methodological observation is the underperformance of SNAPCLUST when compared to NEWHYBRIDS (a conclusion in agreement with other studies [Cordonnier et al., 2019]). Additionally, SNAPCLUST, which must be employed pairwise between each species, could readily detect admixed individuals but not determine which parental pair determined the ancestry. In other words, an A x B hybrid (confirmed by the other methods) might appear admixed no matter the second species tested (A x B, A x C, and A x D). This is problematic for studies assessing a single species pair because SNAPCLUST may erroneously infer an individual as a hybrid between the focal pair when a third species contributed. Thus, we suggest SNAPCLUST be employed only as a first-pass screening tool to identify admixed individuals, from which candidate assignments may be subsequently refined using more robust (and computationally intensive) approaches. Furthermore, this also supports the necessity of a community-level approach for unbiased estimation of hybridization. The more inclusive our datasets are regarding interacting species in natural communities, the better our resolution regarding the landscape of hybridization.

Conclusion

Hybridization happens more frequently than expected. However, fish may not be representative of all animals, so more research into hybridization rates across other taxonomic groups should incorporate a community-level approach as outlined here. The more frequently we encounter hybridization in nature, the more evidence for its influential role in ecology and evolution. While hybridization can be a creative evolutionary force, it is often thought of as maladaptive and threatening to a species' existence. Future work may determine that the reality is more complicated and even show that groups such as cyprinids thrive due to gene exchange among species. Moreover, hybridization is anticipated to increase in frequency due to global environmental change and human-mediated translocation of species (Cordonnier et al., 2019). Therefore, we need baseline estimates to measure the increase in hybridization and detect where it might impact our ecosystems to understand better what those impacts might be.

REFERENCES

Allendorf, F.W., Leary, R.F., Spruell, P., & Wenburg, J.K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*, *16*(11), 613–622.

Anderson, E.C., & Thompson, E.A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217–1229.

Arnold, M.L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology, Evolution, and Systematics*, *23*, 237–261.

Arnold, M.L., Bulger, M.R., Burke, J.M., Hempel, A.L., & Williams, J.H. (1999). Natural hybridization: How low can you go and still be important? *Ecology*, *80*(2), 371–381.

Avise, J.C., & Gold, J.R. (1977). Chromosomal Divergence and Speciation in Two Families of North American Fishes. *Evolution*, *31*, 1–13.

Bangs, M.R., Douglas, M.R., Brunner, P.C., & Douglas, M.E. (2020). Reticulate evolution as a

management challenge: Patterns of admixture with phylogenetic distance in endemic fishes of western North America. *Evolutionary Applications*, *13*(6), 1400–1419.

Beugin, M.P., Gayet, T., Pontier, D., Devillard, S., & Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. *Methods in Ecology and Evolution*, *9*(4), 1006–1016.

Burford, M.O., Bernardi, G., & Carr, M.H. (2011). Analysis of individual year-classes of a marine fish reveals little evidence of first-generation hybrids between cryptic species in sympatric regions. *Marine biology*, *158*(8), 1815–1827.

Carson, E.W., Tobler, M., Minckley, W.L., Ainsworth, R.J., & Dowling, T.E. (2012). Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Molecular Ecology*, *21*(5), 1209–1222.

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

Cordonnier, M., Gayet, T., Escarguel, G., & Kaufmann, B. (2019). From hybridization to introgression between two closely related sympatric ant species. *Journal of Zoological Systematics and Evolutionary Research*, *57*(4), 778-788.

Corush, J.B., Fitzpatrick, B.M., Wolfe, E.L., & Keck, B.P. (2021). Breeding behaviour predicts patterns of natural hybridization in North American minnows (Cyprinidae). *Journal of Evolutionary Biology*, *34*(3), 486-500.

De Santis, V., Quadroni, S., Britton, R.J., Carosi, A., Gutmann Roberts, C., Lorenzoni, M., ... & Zaccara, S. (2021). Biological and trophic consequences of genetic introgression between endemic and invasive Barbus fishes. *Biological Invasions*, *23*(11), 3351-3368.

Dowling, T.E., & Demarais, B.D. (1993). Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature*, *362*, 444–446.

Dowling, T.E., & Hoeh, W.R. (1991). The extent of introgression outside the contact zone between Notropis cornutus and Notropis chrysocephalus (Teleostei: Cyprinidae). *Evolution*, *45*(4), 944–956.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Buenard, G., . . . & Wagner, H.H. (2020). adespatial: Multivariate multiscale spatial analysis. https://cran.r-

project.org/package=adespatial.

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

Fan, H., Ives, A.R., & Surget-Groba, Y. (2018). Reconstructing phylogeny from reduced-representation genome sequencing data without assembly or alignment. *Molecular ecology resources*, *18*(6), 1482-1491.

Folk, R.A., Soltis, P.S., Soltis, D.E., & Guralnick, R. (2018). New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany*, *105*(3), 364-375.

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

Gerard, D., Gibbs, H.L., & Kubatko, L. (2011). Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology*, *11*(1), 1-12.

Gibson, I., Welsh, A.B., Welsh, S.A., & Cincotta, D.A. (2019). Genetic swamping and possible species collapse: tracking introgression between the native Candy Darter and introduced Variegate Darter. *Conservation Genetics*, *20*(2), 287-298.

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

Grabenstein, K.C., & Taylor, S.A. (2018). Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution*, *33*(3), 198-212.

Grant, P.R., & Grant, B.R. (2019). Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences*, *116*(46), 23216-23224.

Hamilton, C.E., Faeth, S.H., & Dowling, T.E. (2009). Distribution of hybrid fungal symbionts and environmental stress. *Microbial ecology*, *58*(2), 408-413.

Harrison, R.G., & Larson, E.L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795-809.

Heath, D., Bettles, C.M., & Roff, D. (2010). Environmental factors associated with reproductive barrier breakdown in sympatric trout populations on Vancouver Island. *Evolutionary Applications*, *3*(1), 77-90.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, *4*(1), 1-20.

Hubert, M., Reynkens, T., Schmitt, E., & Verdonck, T. (2016). Sparse PCA for High-Dimensional Data With Outliers. *Technometrics*, *58*(4), 424-434.

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

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

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.

Justen, H., Kimmitt, A.A., & Delmore, K.E. (2021). Estimating hybridization rates in the wild: Easier said than done? *Evolution*, *75*(8), 2137-2144.

Justyn, N.M., Callaghan, C.T., & Hill, G.E. (2020). Birds rarely hybridize: A citizen science approach to estimating rates of hybridization in the wild. *Evolution*, *74*(6), 1216-1223.

Keck, B.P., & Near, T.J. (2009). Patterns of natural hybridization in darters (Percidae: Etheostomatinae). *Copeia*, *2009*(4), 758-773.

Kotze, D.J., Johnson, C.A., O'Hara, R.B., Vepsäläinen, K., & Fowler, M.S. (2004). Editorial: The journal of negative results in ecology and evolutionary biology. *Journal of Negative Results*, *1*, 1–5.

Lamer, J.T., Dolan, C.R., Petersen, J.L., Chick, J.H., & Epifanio, J.M. (2010). Introgressive hybridization between bighead carp and silver carp in the Mississippi and Illinois rivers. *North American Journal of Fisheries Management*, *30*(6), 1452-1461.

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), 1-15.

Lowe, W.H., Muhlfeld, C.C., Allendorf, F.W. (2015). Spatial sorting promotes the spread of maladaptive hybridization. *Trends in Ecology & Evolution*, *30*(8), 456-462.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in ecology & evolution*, *20*(5), 229-237.

Matthews, W.J., & Marsh-Matthews, E. (2017). *Stream fish community dynamics*. Johns Hopkins University Press.

Matthews, W.J., Turner, T.F., & Osborne, M.J. (2016). Breakdown of a hybrid swarm between two darters (Percidae), *Etheostoma radiosum* and *Etheostoma spectabile*, with loss of one parental species. *Copeia*, *104*(4), 873-878.

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press.

McVay, J.D., Hipp, A.L., & Manos, P.S. (2017) A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854), 20170300.

Meng, C., & Kubatko, L.S. (2009). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical population biology*, *75*(1), 35-45.

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.

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

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2020). vegan: Community ecology package. https://cran.r-project.org/package=vegan

Ottenburghs, J. (2019). Multispecies hybridization in birds. *Avian Research*, *10*(1), 1-11.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446.

Peoples, B.K., Floyd, S.P., Frimpong, E.A. (2016). Nesting microhabitat comparison of Central

stoneroller and Bluehead chub: potential inference for host-switching by nest associates. *Journal of Freshwater Ecology*, *31*(2), 251-259.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, *7*(5), e37135.

Peterson, R.A., & Cavanaugh, J.E. (2020). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*, 2312–2327.

Pfennig, K.S. (2021). Biased Hybridization and Its Impact on Adaptive Introgression. *Trends in ecology & evolution*, *36*(6), 488-497.

Randi, E. (2008). Detecting hybridization between wild species and their domesticated relatives. *Molecular ecology*, *17*(1), 285-293.

Rhymer, J.M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual review of ecology and systematics*, *27*(1), 83-109.

Schwenk, K., Brede, N., Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2805-2811.

Scribner, K.T., Page, K.S., & Bartron, M.L. (2000). Hybridization in freshwater fishes: A review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, *10*(3), 293-323.

Taylor, S.A., & Larson, E.L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature ecology & evolution*, *3*(2), 170-177.

Thorgaard, G.H., & Allendorf, F.W. (1988). Developmental genetics in fishes. In *Developmental Genetics of Animals and Plants* (ed GM Malacinski), pp. 369–391. MacMillan.

Tyers, M. (2017). Riverdist: River Network Distance Computation and Applications. https://cran.r-project.org/package=riverdist.

Twyford, A.D., & Ennos, R.A. (2012). Next-generation hybridization and introgression. *Heredity*, *108*(3), 179-189.

Vallender, R., Robertson, R.J., Friesen, V.L., & Lovette, I.J. (2007). Complex hybridization dynamics between golden-winged and blue-winged warblers (Vermivora chrysoptera and Vermivora pinus) revealed by AFLP, microsatellite, intron and mtDNA markers. *Molecular Ecology*, *16*(10), 2017-2029.

Wallis, G.P., Cameron-Christie, S.R., Kennedy, H.L., Palmer, G., Sanders. T,R,, & Winter, D.J. (2017) Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes. *Molecular Ecology*, *26*(12), 3116-3127.

Wringe, B.F., Stanley, R.R.E., Jeffery, N.W., Anderson, E.C., & Bradbury, I.R. (2017). hybriddetective: A workflow and package to facilitate the detection of hybridization using genomic data in r. *Molecular Ecology Resources*, *17*(6), e275-e284

TABLES AND FIGURES

Table 1: Overview of 33 species sampled across the White River Drainage, USA. N_{SP} =number of individuals/species, N_{SI} =number of collection sites/species, N_{FA} =number of individuals/family, SNPs=single nucleotide polymorphisms in family panel.

Family	Common Name	Scientific Name	N_{SP}	N_{SI}	N_{FA}	SNPs
Centrarchidae (Sunfish)	Bluegill Sunfish	*Lepomis macrochirus*	69	23	375	1926
	Longear Sunfish	*Lepomis megalotis*	239	54		
	Smallmouth Bass	*Micropterus dolomieu*	43	30		
	Largemouth Bass	*Micropterus salmoides*	24	17		
Cottidae (Sculpin)	Banded Sculpin	*Cottus carolinae*†	33	18	75	5344
	Knobfin Sculpin	*Cottus hypselurus*†	42	10		
Fundulidae (Topminnow)	Northern Studfish	*Fundulus catenatus*	105	28	226	2366
	Blackspotted Topminnow	*Fundulus olivaceus*	121	34		
Ictaluridae (Catfish)	Ozark Madtom	*Noturus albater*	10	5	31	2744
	Slender Madtom	*Noturus exilis*	16	12		
	Black River Madtom	*Noturus maydeni*	5	3		
Leuciscidae (Minnow)	Central Stoneroller	*Campostoma anomalum*	128	44	1507	343
	Largescale Stoneroller	*Campostoma oligolepis*	135	41		
	Southern Redbelly Dace	*Chrosomus erythrogaster*	38	10		
	Whitetail Shiner	*Cyprinella galactura*	75	16		
	Steelcolor Shiner	*Cyprinella whipplei*	30	8		
	Striped Shiner	*Luxilus chrysocephalus*	63	18		
	Duskystripe Shiner	*Luxilus pilsbryi*	258	33		
	Bleeding Shiner	*Luxilus zonatus*	100	17		
	Redfin Shiner	*Lythrurus umbratilis*	24	5		
	Bigeye Shiner	*Notropis boops*	226	31		
	Ozark Minnow	*Notropis nubilus*	193	35		
	Carmine Shiner	*Notropis percobromus*	67	15		
	Telescope Shiner	*Notropis telescopus*	83	15		
	Bluntnose Minnow	*Pimephales notatus*	55	24		
	Creek Chub	*Semotilus atromaculatus*	32	14		
Percidae (Darter)	Greenside Darter	*Etheostoma blennioides*	62	26	651	687
	Rainbow Darter	*Etheostoma caeruleum*	348	53		
	Fantail Darter	*Etheostoma flabellare*	26	11		
	Yoke Darter	*Etheostoma juliae*	63	15		
	Orangethroat Darter	*Etheostoma spectabile*	50	10		
	Current Darter	*Etheostoma uniporum*	18	7		
	Banded Darter	*Etheostoma zonale*	84	26		

Table 2: Hybridization summarized within six families sampled across the White River Basin, USA. N Species=number of species analyzed, Unique pairs=number of unique pairs (= n ((n – 1)/2)), N Indiv=number of individuals examined, N Hybrids=number of hybrids detected, Percent Indiv=Hybrid percentage. Unique species-pairs with hybrids only considered for ancestry between two species.

Family	N Species	Unique Pairs	N Indiv.	N Hybrids	Percent Indiv.	Species w/ Hybrids	Percent Species w/ Hybrid	Unique Pairs w/ Hybrid	Percent Pairs w/ Hybrid
Fundulidae	2	1	226	0	0.0%	0	0%	0	0%
Cottidae	2	1	75	0	0.0%	0	0%	0	0%
Ictaluridae	3	3	31	1	3.2%	3	100%	0	0%
Centrarchidae	4	6	375	1	0.3%	2	50%	1	17%
Percidae	7	21	651	2	0.3%	4	57%	2	10%
Leuciscidae	15	105	1507	66	4.4%	15	100%	15	14%
Overall	33	137	2865	70	2.4%	24	73%	18	13%

Table 3: Observed genotype frequency classes of hybrid individuals inferred from NewHybrids analysis for 18 hybridizing fish species-pairs collected across the Whiter River Basin, USA. Note, 8 multispecific hybrid individuals not included. Putative hybrid individuals were assigned to a genotype frequency class [F_1, F_2, Bx(=backcross), pure] based on Bayesian posterior probability > 0.70.

No.	Species A	Species B	N Indiv.	F_1	F_2	Bx	Pure
1	*Campostoma anomalum*	*Campostoma oligolepis*	29	5	-	11	13
2	*Campostoma anomalum*	*Chrosomus erythrogaster*	1	1	-	0	-
3	*Campostoma anomalum*	*Luxilus pilsbryi*	1	-	1	0	-
4	*Campostoma oligolepis*	*Notropis telescopus*	1	1	-	0	-
5	*Cyprinella galactura*	*Cyprinella whipplei*	2	1	-	1	-
6	*Cyprinella whipplei*	*Lythrurus umbratilis*	1	-	-	1	-
7	*Luxilus chrysocephalus*	*Luxilus zonatus*	2	1	-	0	1
8	*Luxilus chrysocephalus*	*Semotilus atromaculatus*	1	-	-	1	-
9	*Luxilus pilsbryi*	*Lythrurus umbratilis*	6	-	1	5	-
10	*Luxilus pilsbryi*	*Notropis percobromus*	2	1	-	1	-
11	*Luxilus pilsbryi*	*Luxilus zonatus*	8	-	-	0	8
12	*Luxilus pilsbryi*	*Luxilus chrysocephalus*	1	-	-	1	-
13	*Luxilus zonatus*	*Pimephales notatus*	1	1	-	0	-
14	*Notropis boops*	*Notropis nubilus*	2	-	-	1	1
15	*Pimephales notatus*	*Semotilus atromaculatus*	1	-	1	0	-
16	*Micropterus salmoides*	*Micropterus dolomieu*	1	1	-	0	-
17	*Etheostoma spectabile*	*Etheostoma caeruleum*	1	-	-	1	-
18	*Etheostoma juliae*	*Etheostoma zonale*	1	-	-	1	-
		Totals	62	12	3	24	23

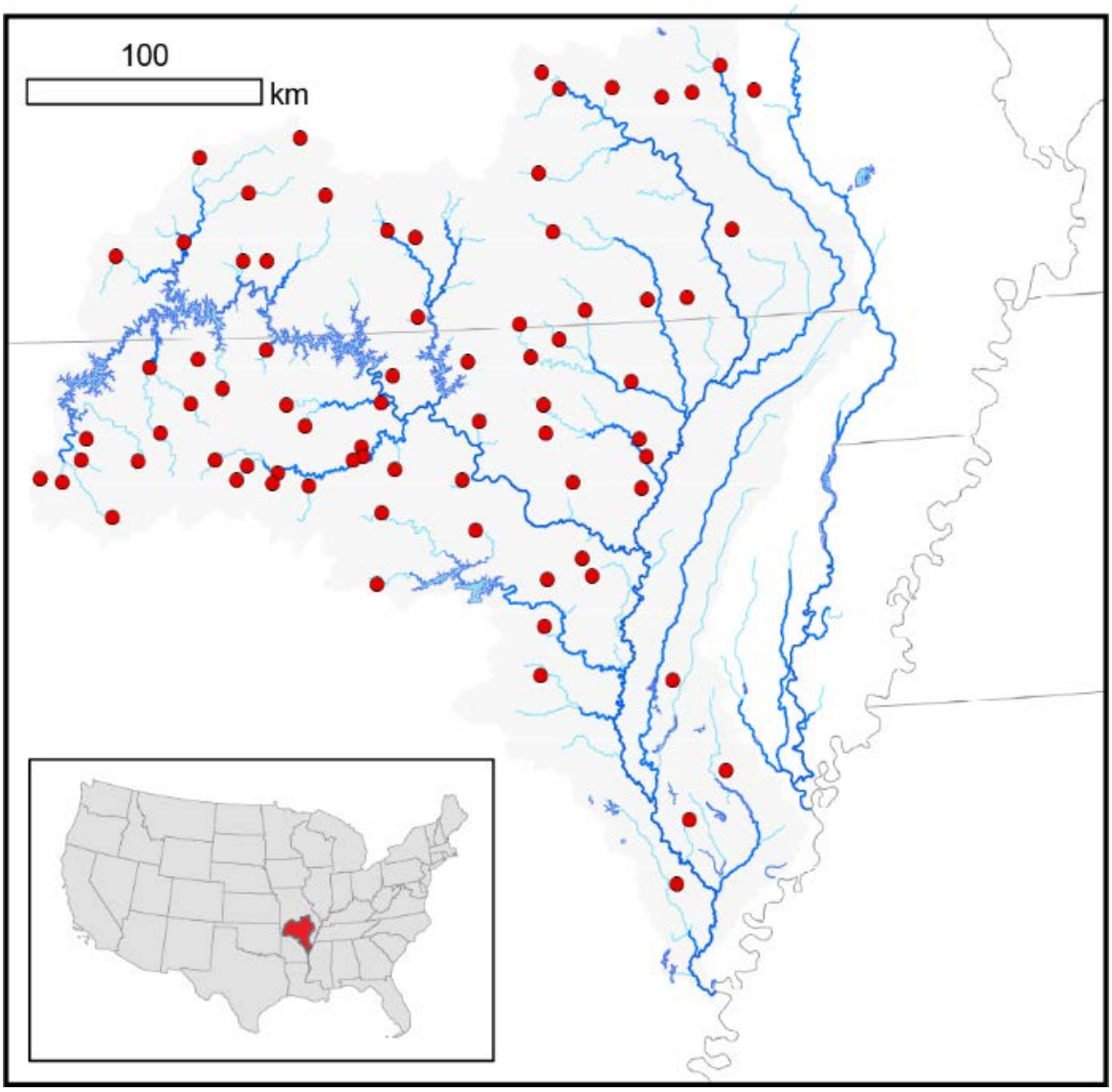


Figure 1: Collection sites (*N*=75; red circles) in the White River Basin (red area of insert; Missouri and Arkansas, USA). Streams/rivers are colored/sized by Strahler Stream Order (larger rivers darker/bolder).

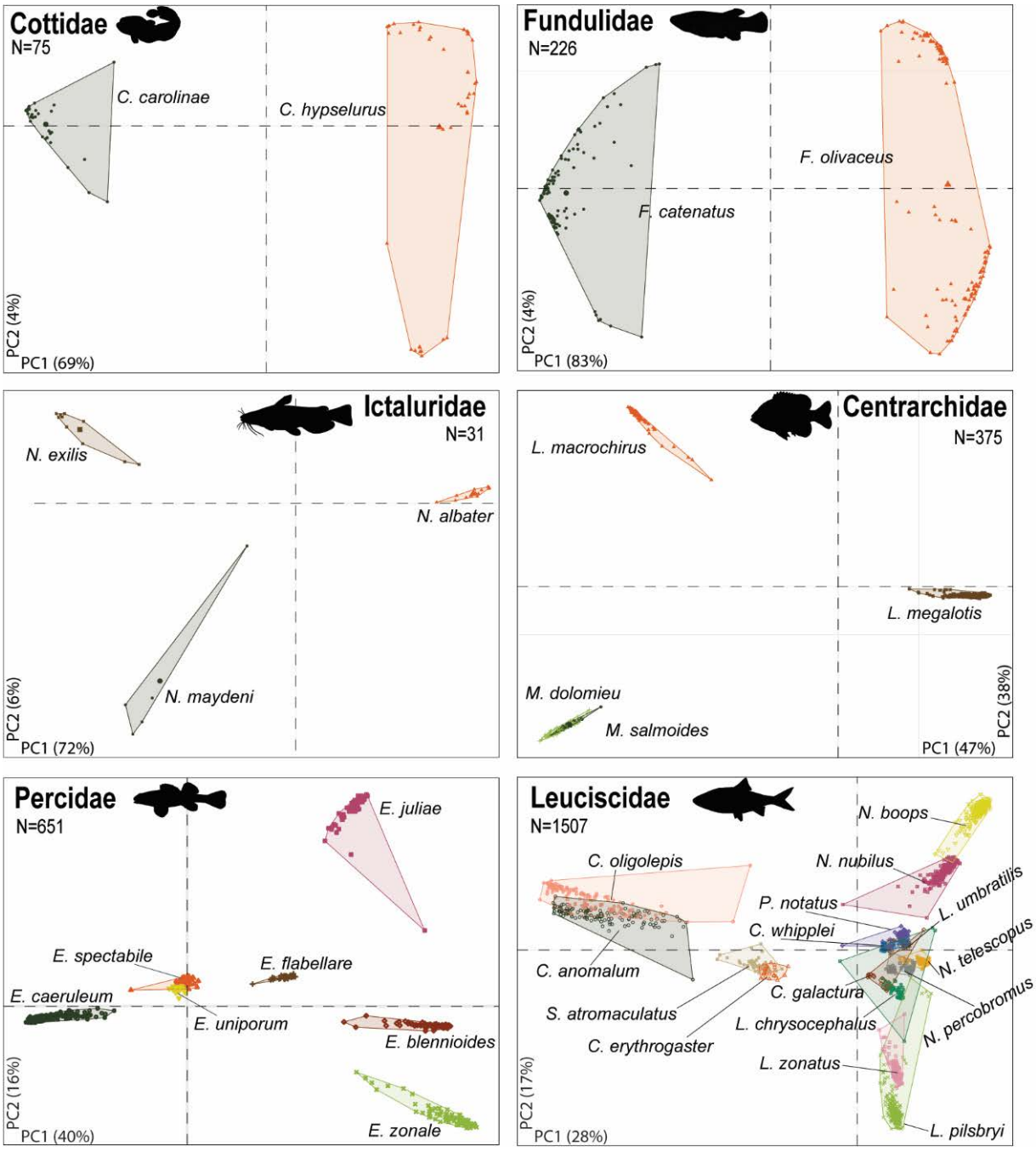


Figure 2: First two principal components derived from SNP genotypes in six stream fish families. Variance explained by each component is in bottom left or right corner of each plot. N=number of individuals/family.

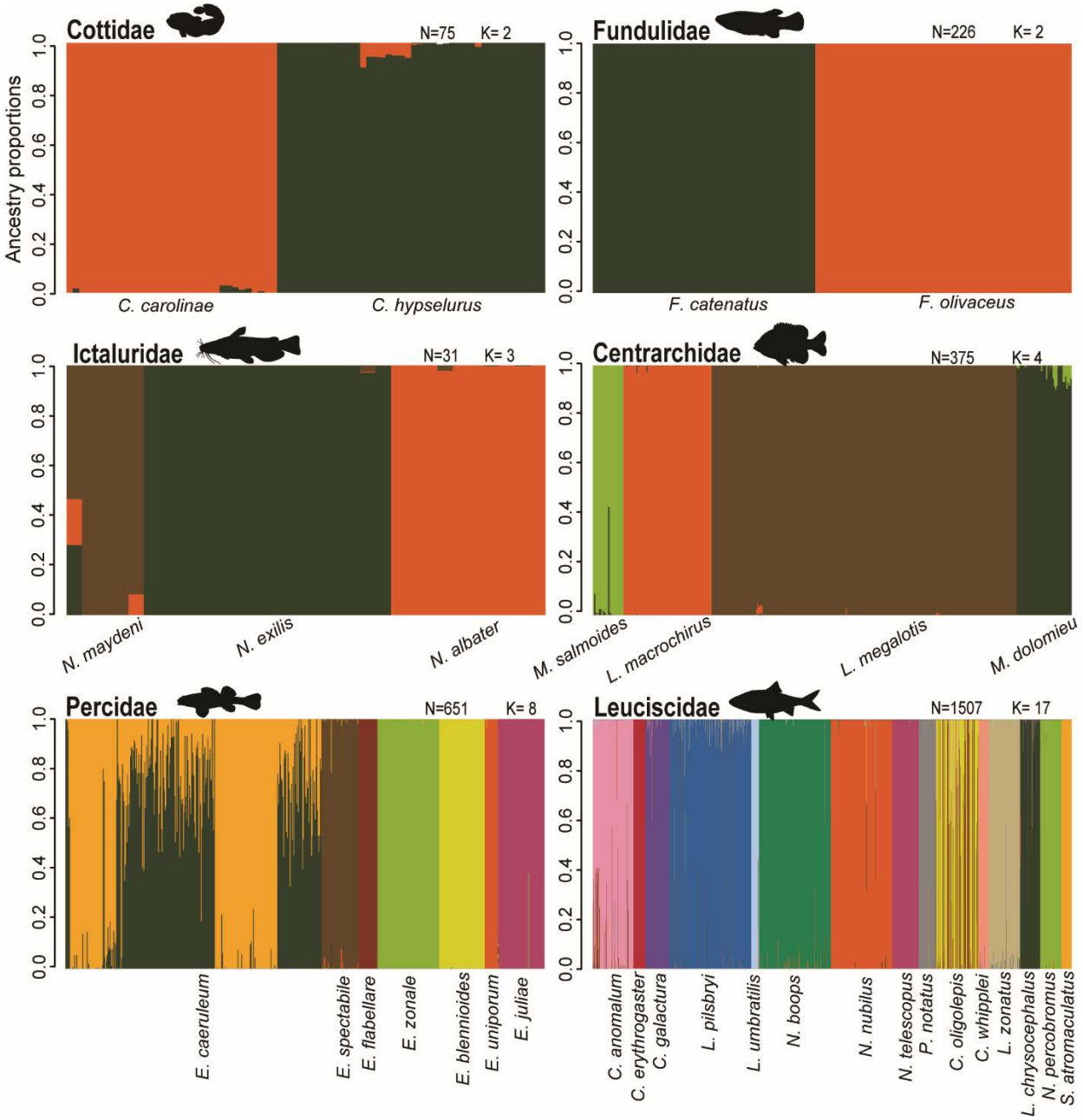


Figure 3: Calculated ancestral proportions for N=2,865 individuals from N=33 fish species grouped by family, where *N*=number of individuals, *K*=optimum number of clusters. When *K*>number of species (Percidae and Leuciscidae) then population structure is present. Individual assignment to parental species <0.90 =putative hybrid.

SUPPLEMENTARY MATERIALS

Supplement S1: Supplementary materials and methods

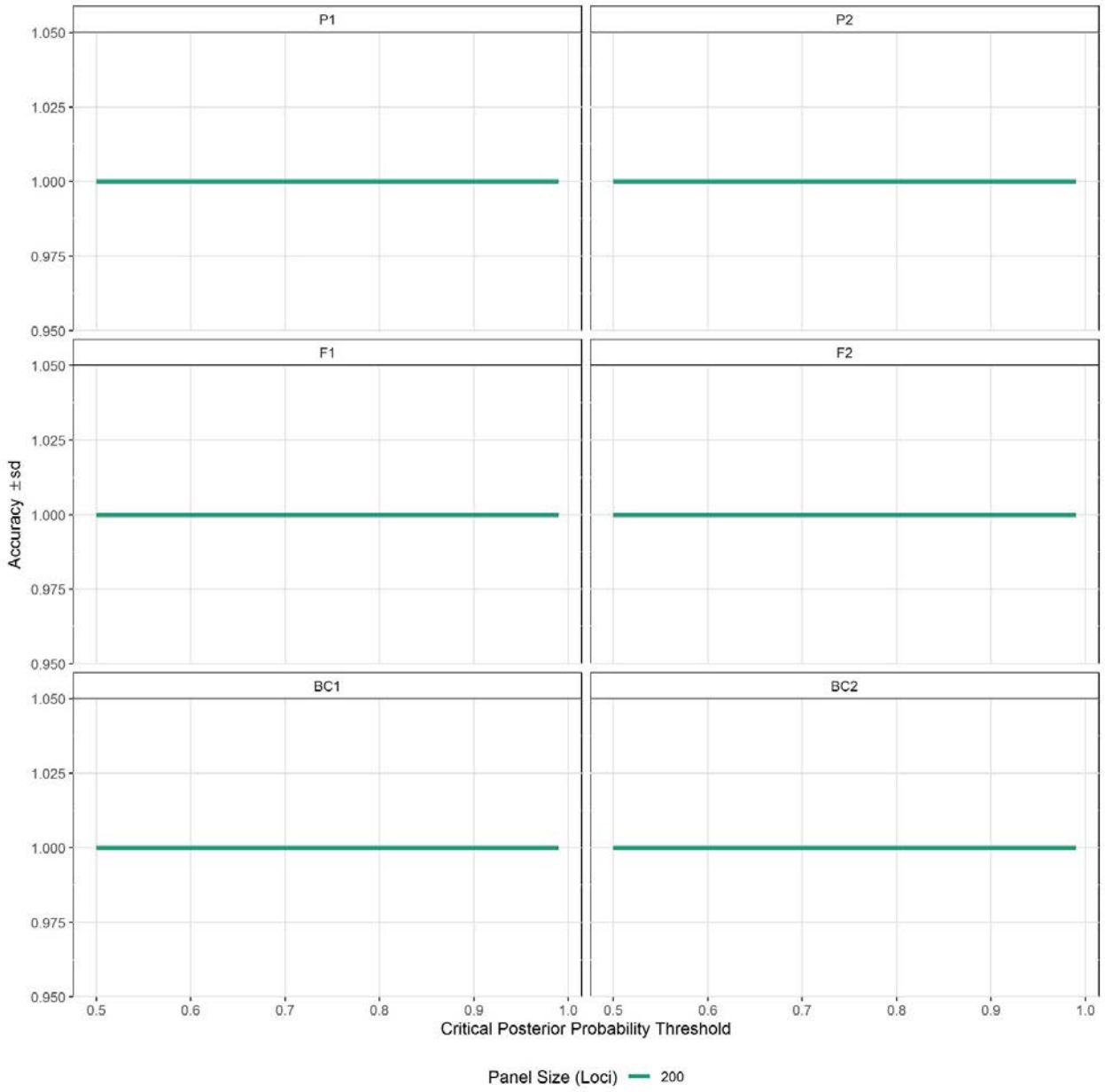
Collecting permits
Collecting permits were provided by: Arkansas Game & Fish Commission (#020120191); Missouri Dept. Wildlife Conservation (#18136); and US National Parks Service (NPS: Buffalo River Permit; BUFF-2017-SCI-0013).

Library preparation
Standardized DNA concentrations (1000 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 100 ng, and then ligated with custom adapters containing in-line identifying barcodes (T4 Ligase; New England Biosciences). Individual samples were pooled in sets of 48 and size-selected from 326 to 426 bp (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 three libraries (3x48=144 individuals/lane) were pooled per-lane and sequenced single-end on the Illumina HiSeq 4000 platform (1x100bp; Genomics & Cell Characterization Core Facility; University of Oregon/ Eugene). Quality control checks, to include fragment analysis and quantitative real-time PCR, were performed at the core facility prior to sequencing.

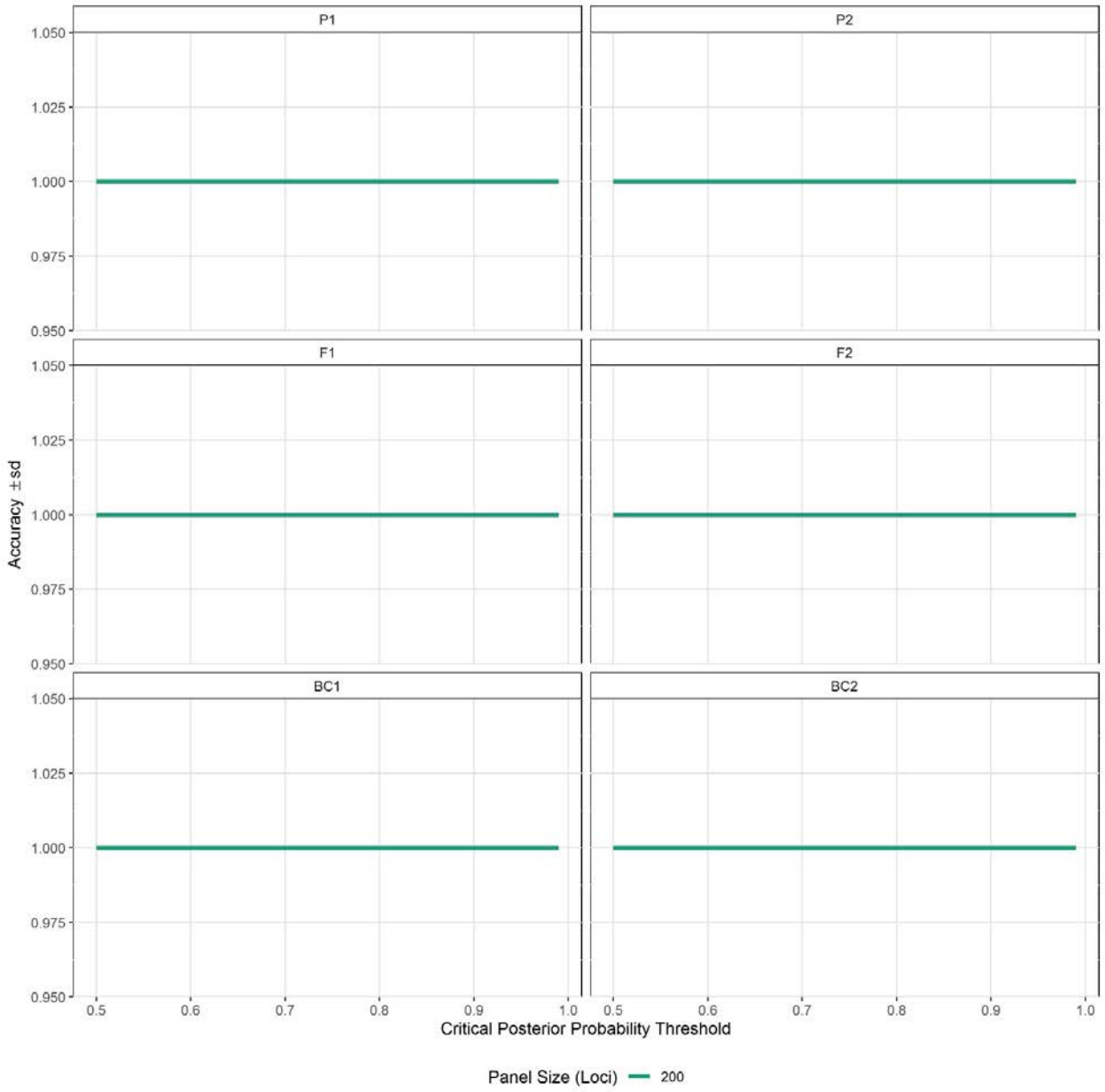
Alignment and assembly free (AAF) phylogenies
Data assembly was performed in multiple stages: Pre-processing; Species validation and screening; and family-level alignment. Reads were first demultiplexed in IPYRAD v.0.9.62 allowing up to one barcode mismatch, yielding individual FASTQ files containing raw reads (*N*=3,101). Individuals averaged >2 million reads, with those extremely low being removed (*N=59*). Species identifications were validated using an alignment and assembly-free (AAF) method (PHYLORAD). This method is comparable to alignment-based methods but with lower computational burdens, as accomplished by computing a pairwise distance matrix as the proportion of DNA substrings of length *k* (=*k*-mer) as a function of total unique identified *k*-mers, followed by phylogenetic reconstruction following the Fitch-Margoliash method. The same *k*-mer length was employed for read selection (*k*s) and reconstruction (*k*) = 21, with a k-mer filtering threshold (*n*) = 2. Phylogenies were inspected manually, with putative mis-identified individuals not clustering with the appropriate species-clade. These individuals were re-evaluated via vouchered specimens (*N*=249; 8%), and a assinged a different species ID due to misidentification.

Data processing and assembly
Individuals (*N*=3,042) were subsequently partitioned by family (*N*=6) and processed *de novo* in ipyrad to generate family-level assemblies. Adapters and primers were removed and reads with >5 low-quality bases (Phred<20) discarded. Clusters were assembled using an 85% identity threshold, with loci subsequently removed via conditional criteria to ensure high-quality data: <20x and >500x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of nucleotides polymorphic; >8 indels present; or presence in <15% of individuals. Putative

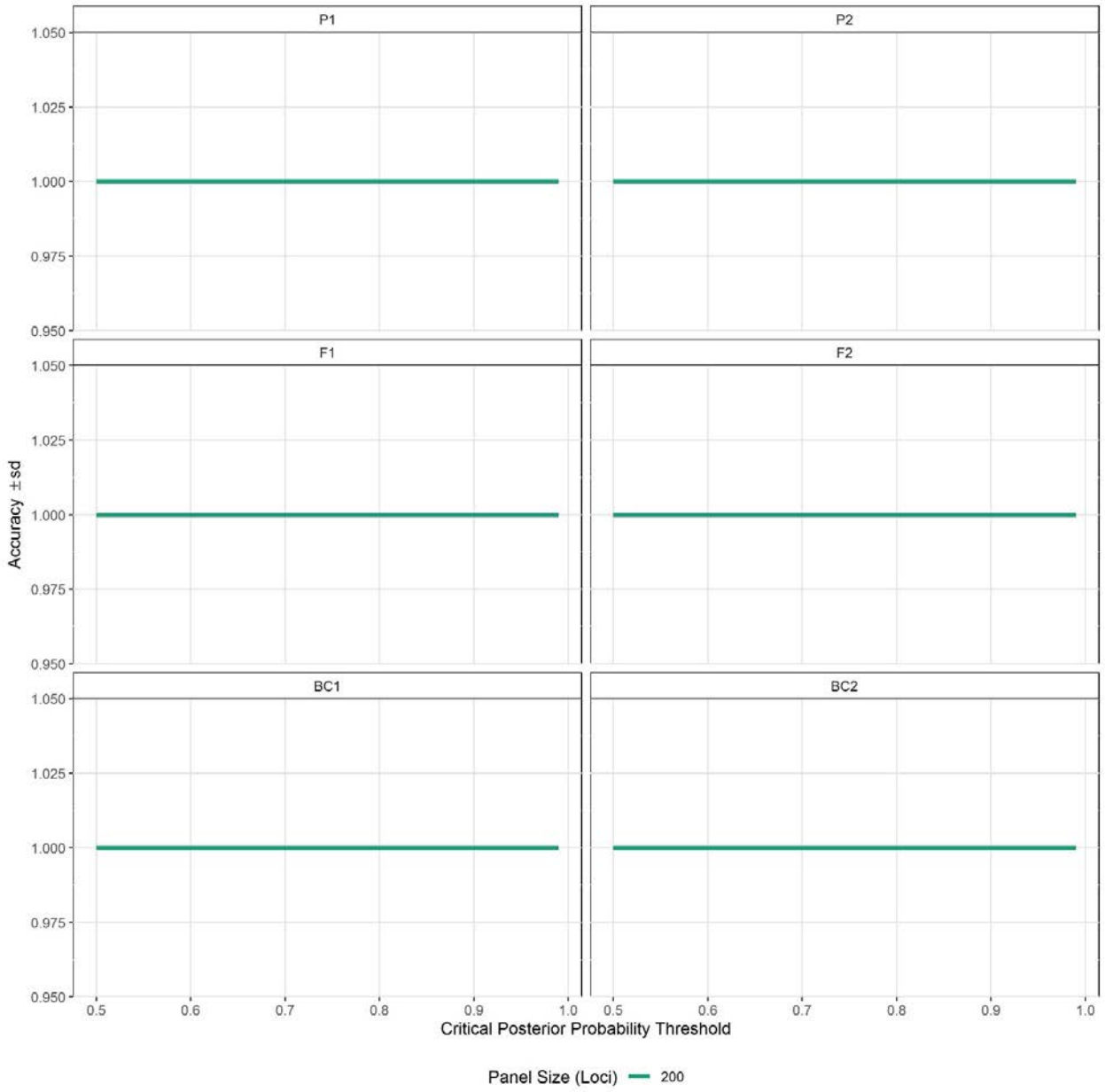
paralogs were removed if clusters displayed >2 alleles per site in consensus sequence or excessive heterozygosity (>5% consensus bases or >50% heterozygosity/site). Biallelic SNPs were further filtered in radiator, and removed if: Monomorphic; minor allele frequency <3%; mean coverage <20 or >200; missing data >30%; SNP position on read >91; and if HWE lacking in one or more species (α=0.0001). Only one SNP was retained per locus (i.e., that which maximized minor allele count). Finally, individuals having >50% missing data were excluded.



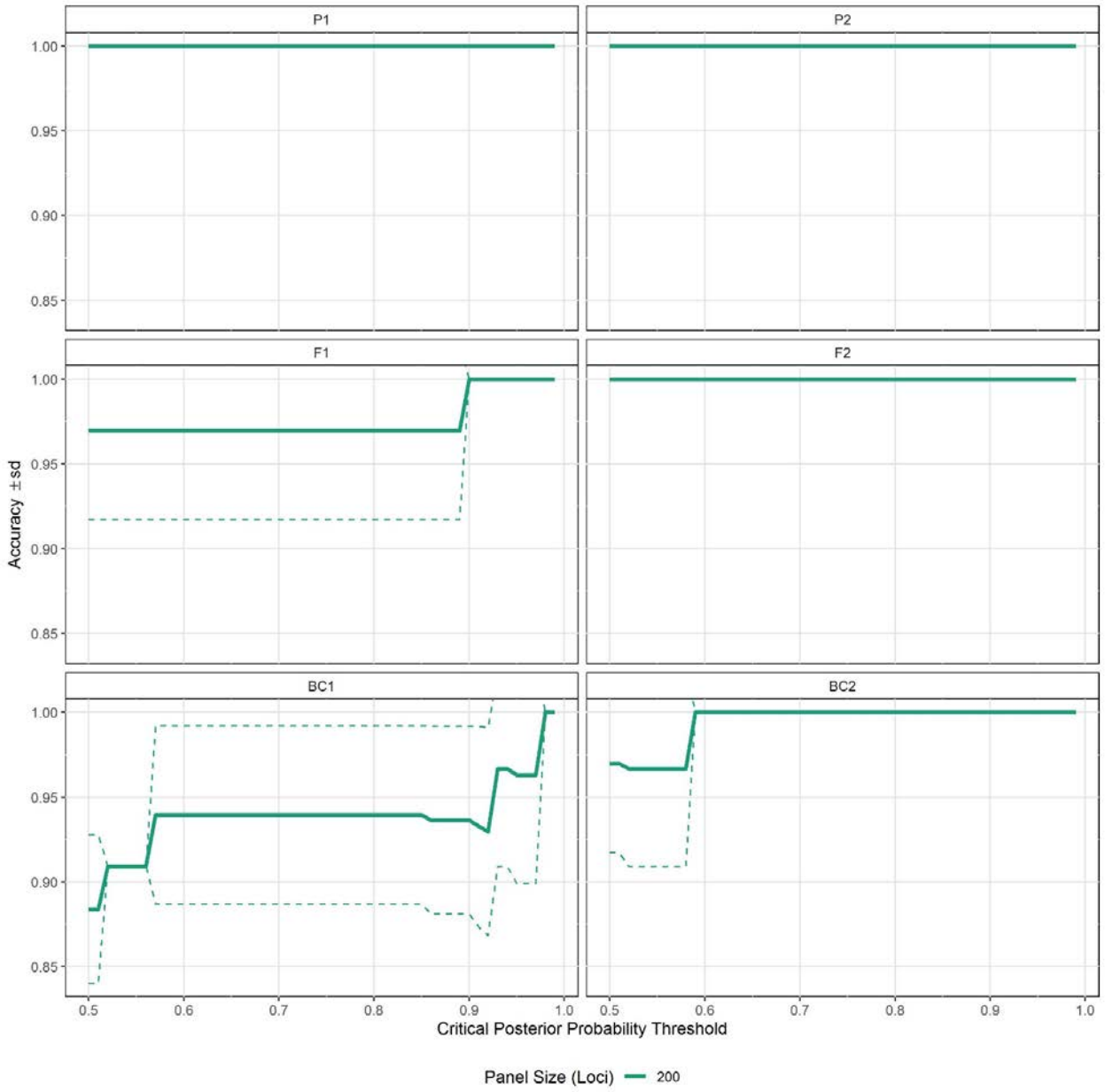
Supplement S2: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Micropterus dolomieu* X *Micropterus salmoides*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



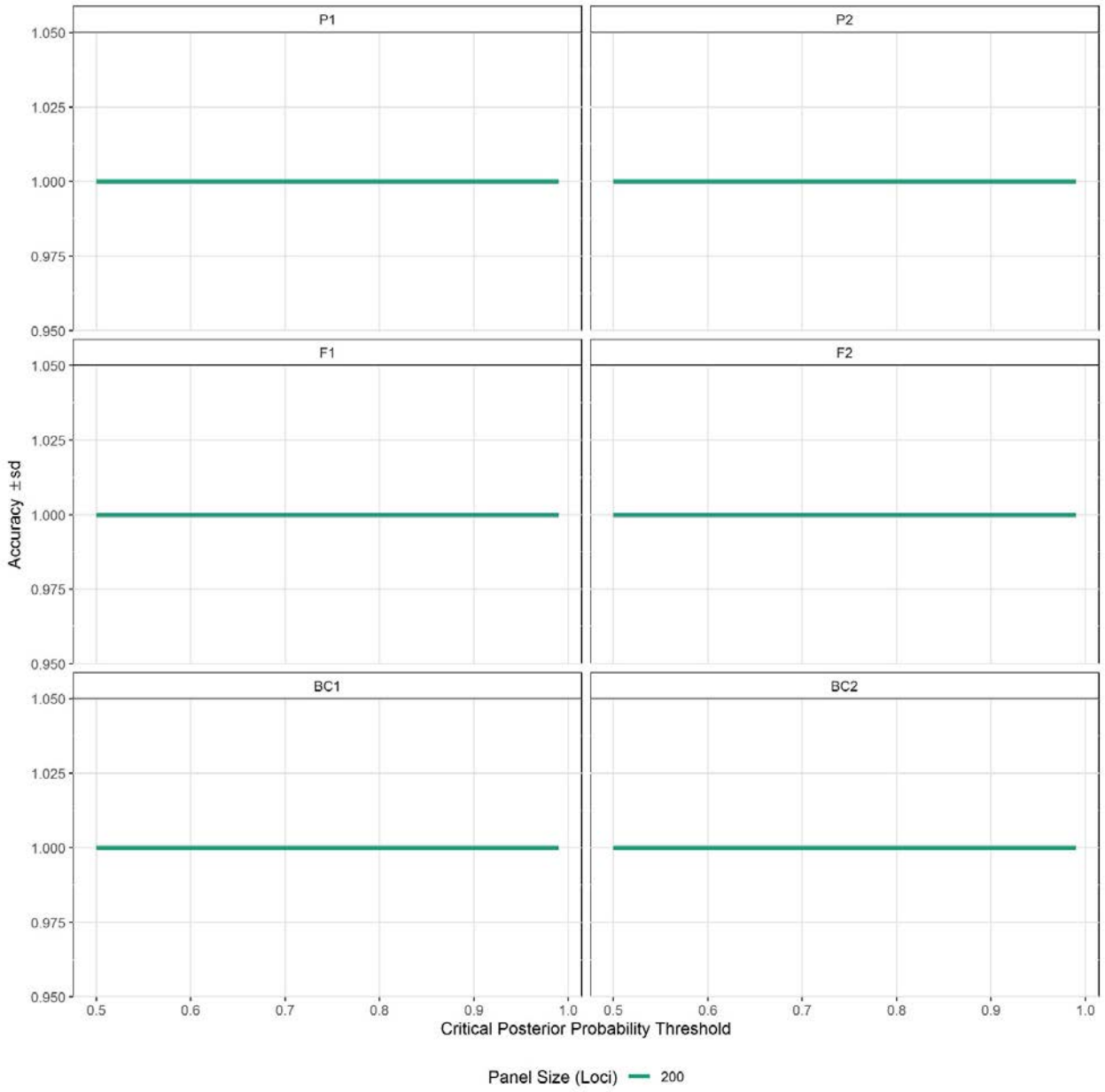
Supplement S3: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Etheostoma caeruleum* X *Etheostoma spectabile*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



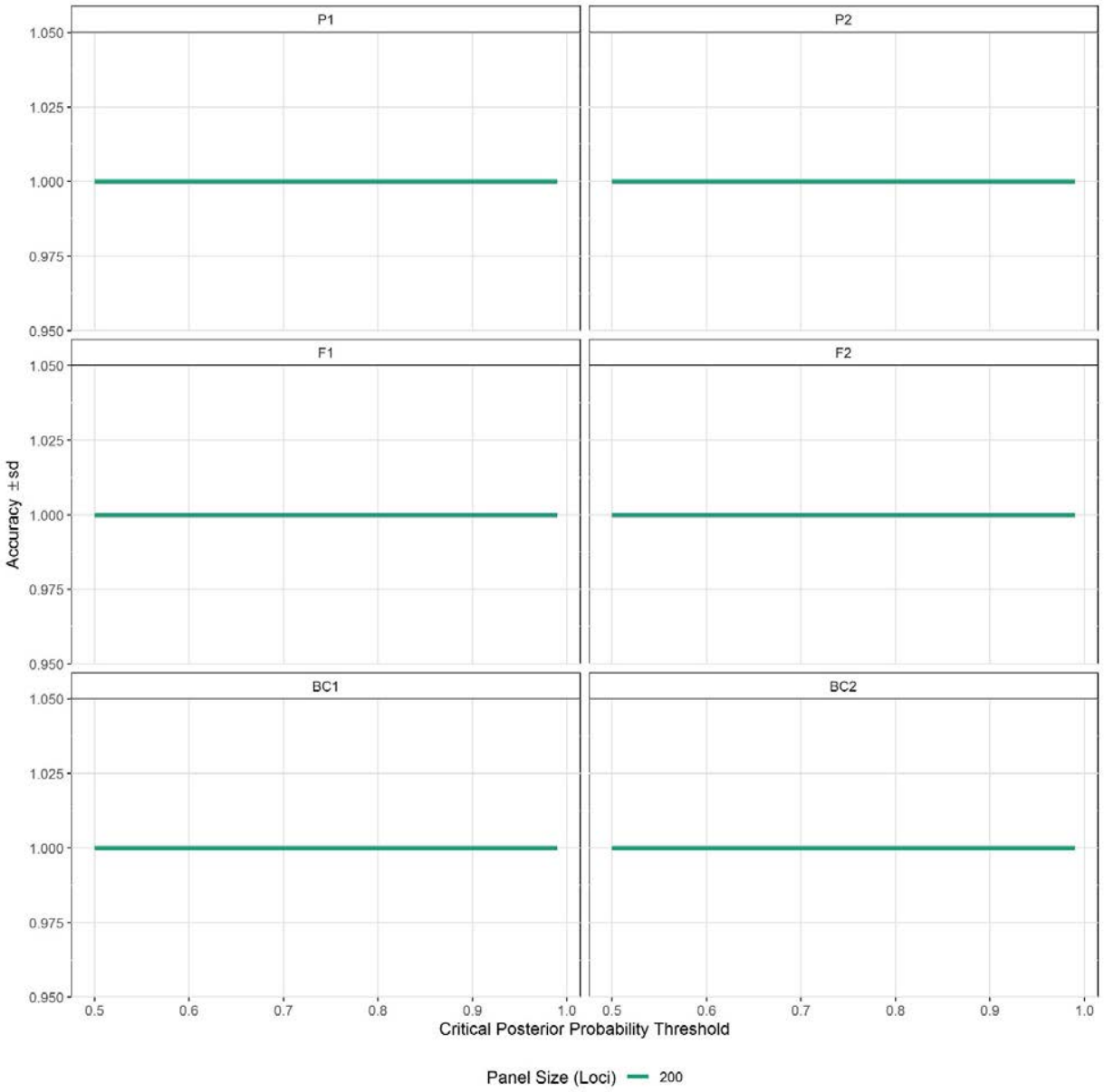
Supplement S4: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Etheostoma juliae* X *Etheostoma zonale*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



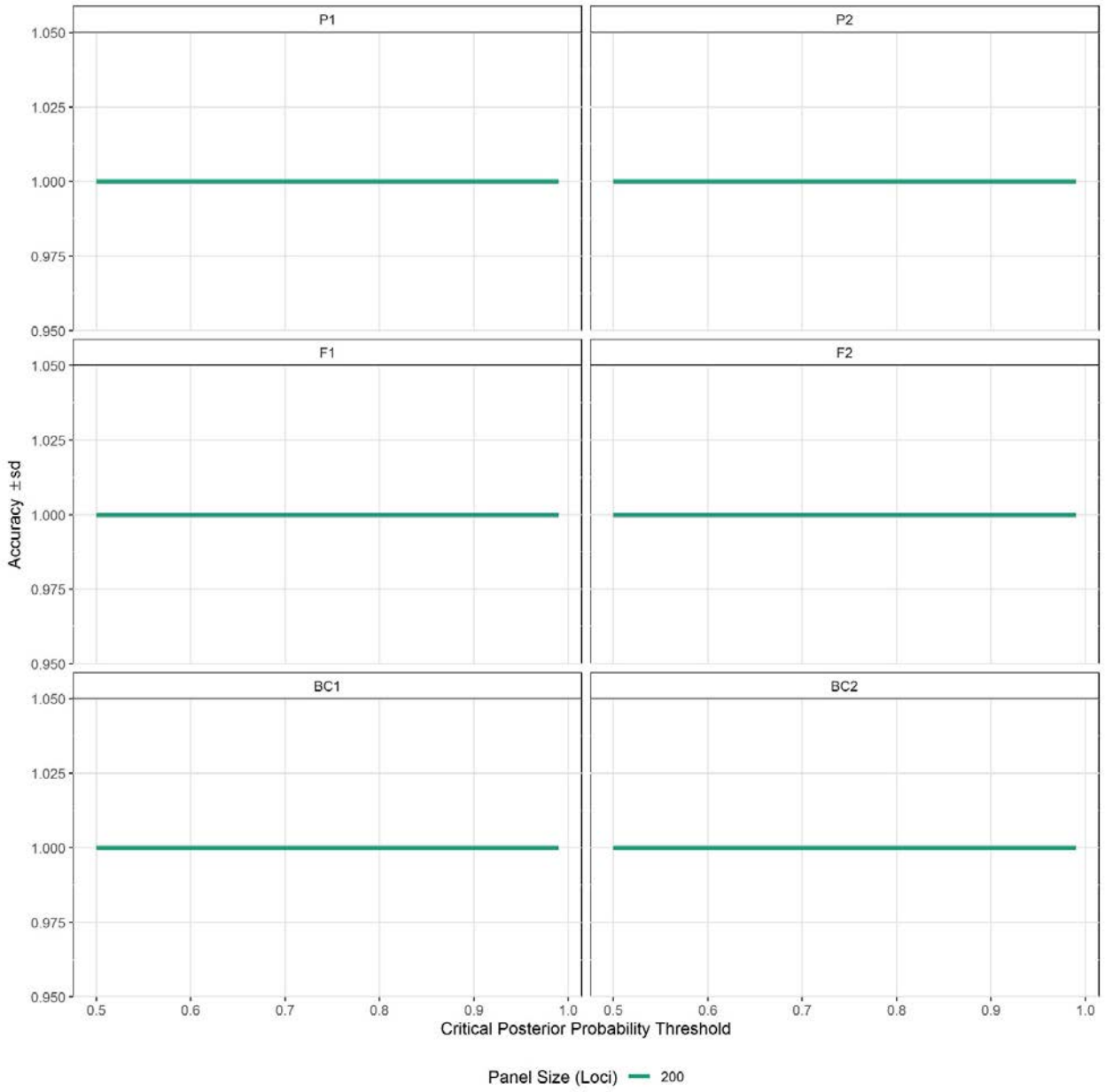
Supplement S5: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Campostoma anomalum* X *Campostoma oligolepis*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



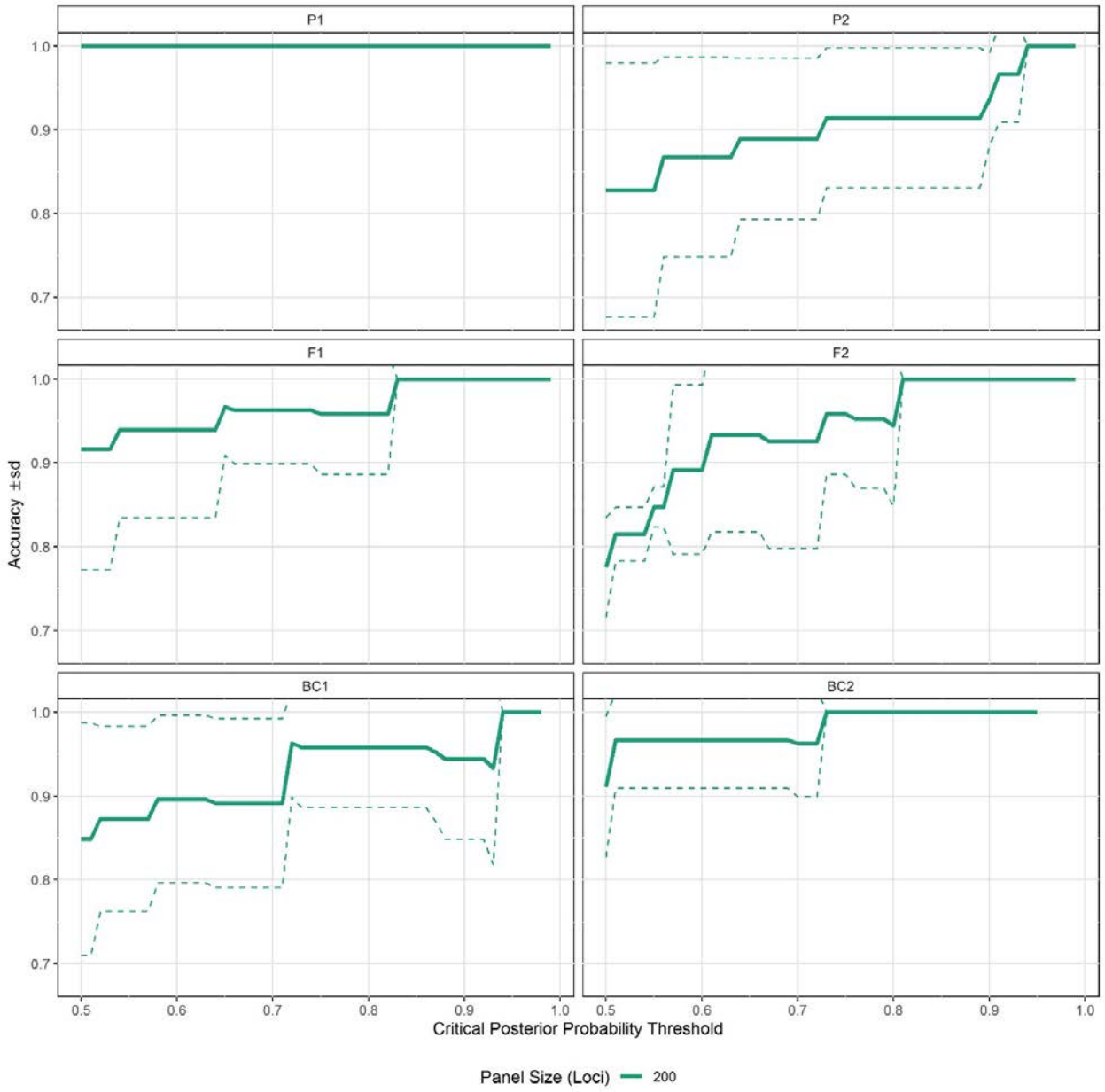
Supplement S6: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Campostoma anomalum* X *Chrosomus erythrogaster*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



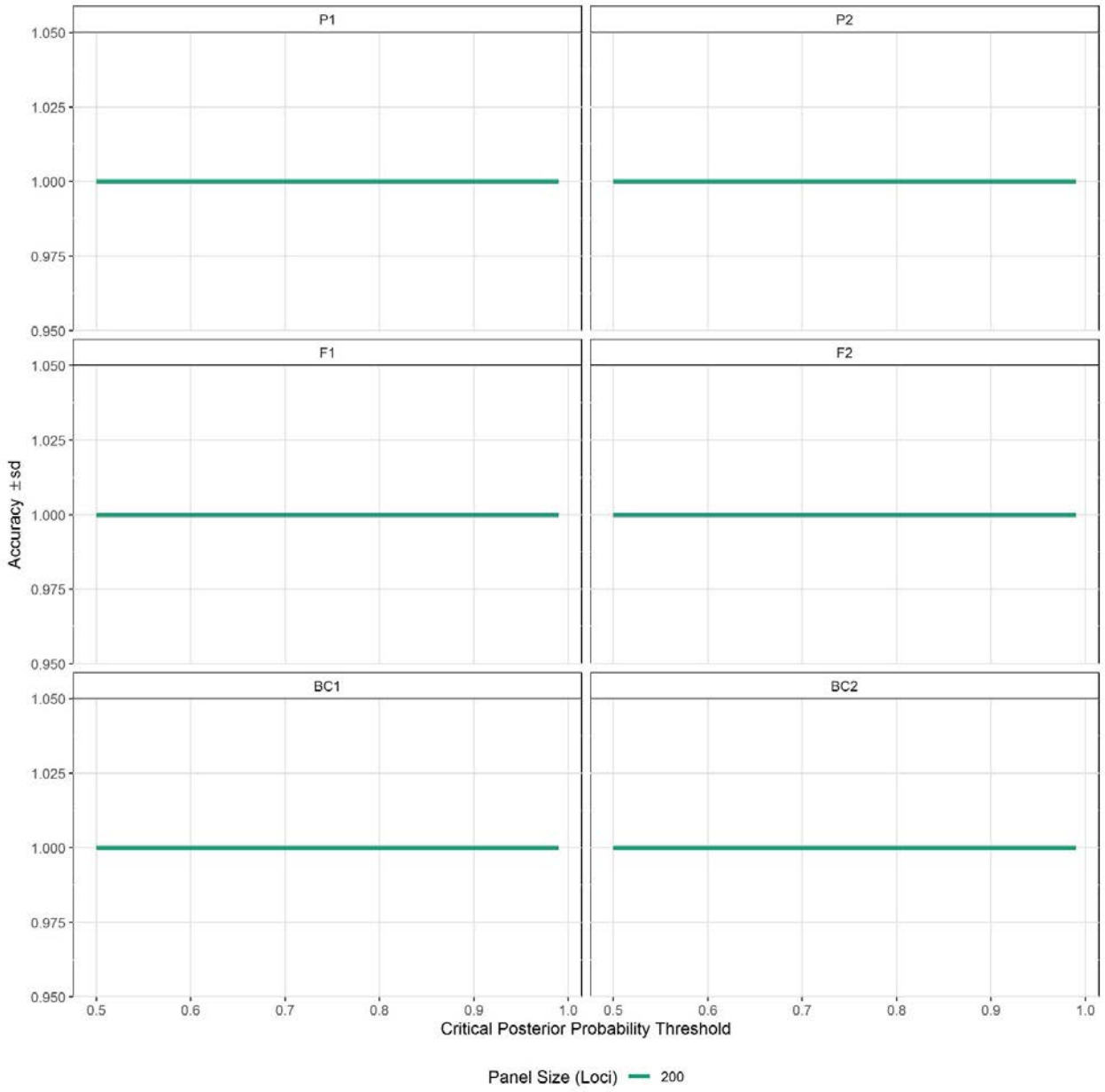
Supplement S7: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Campostoma anomalum* X *Luxilus pilsbryi*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



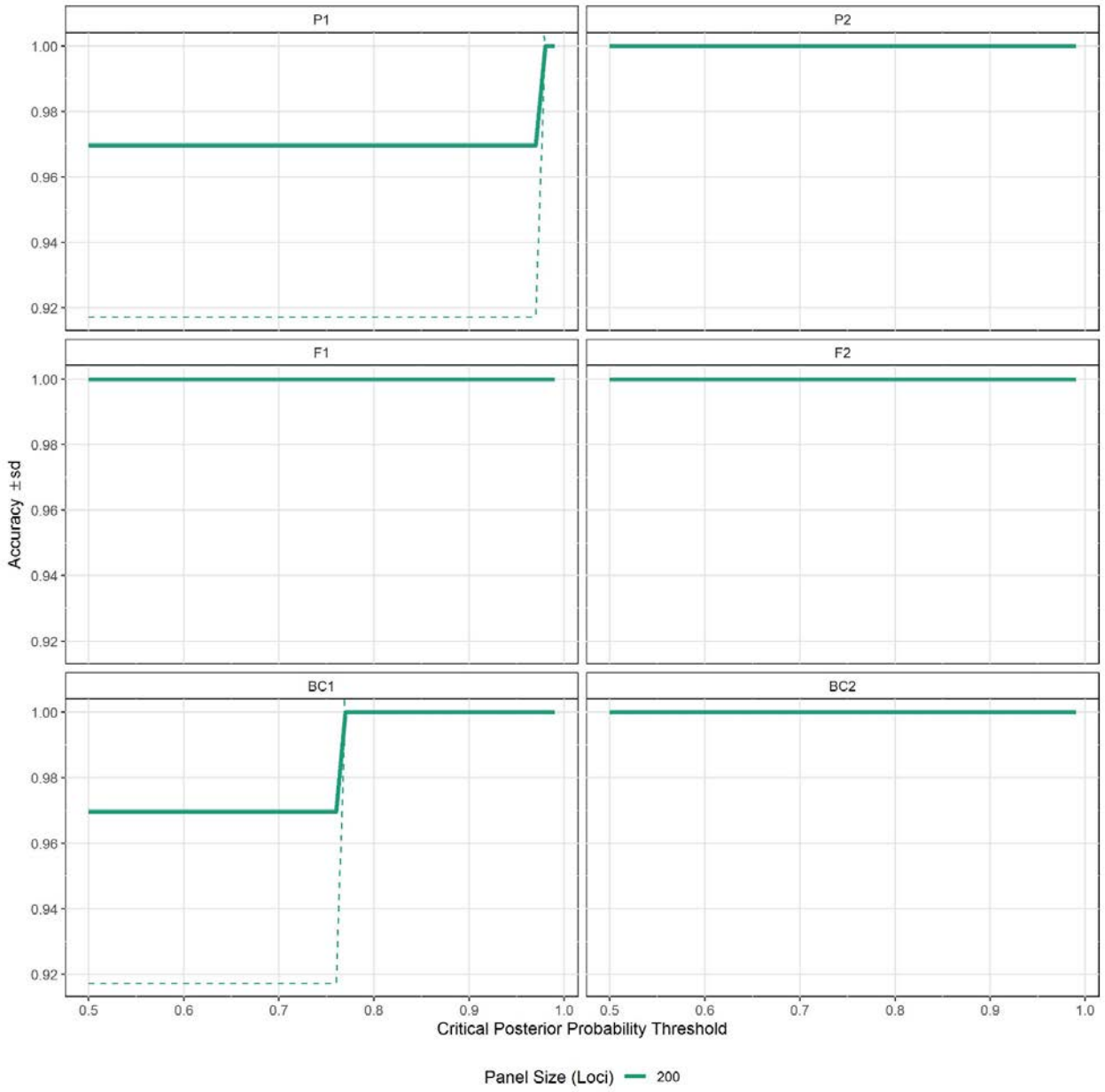
Supplement S8: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Campostoma oligolepis* X *Notropis telescopus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



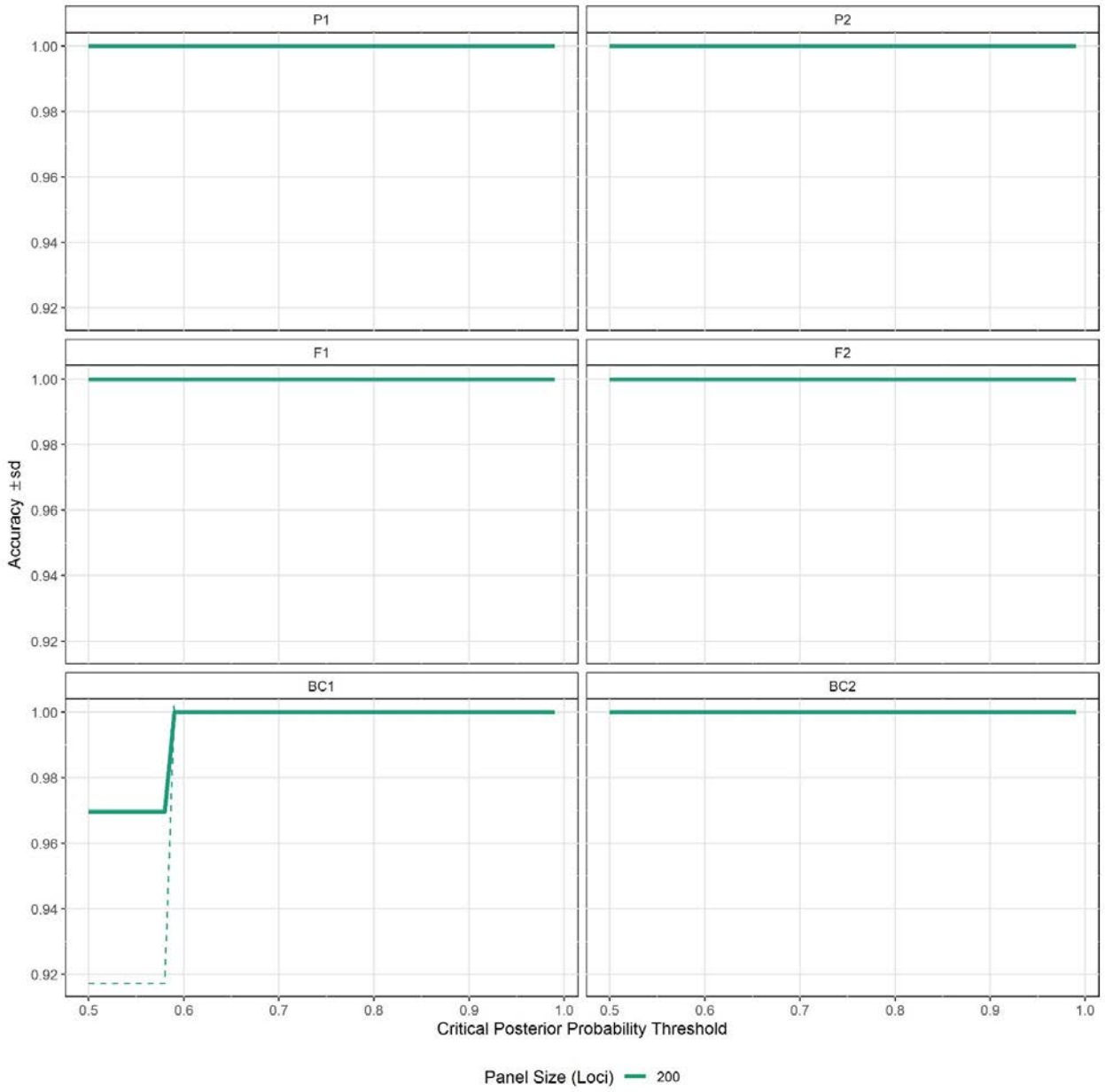
Supplement S9: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Cyprinella whipplei* X *Cyprinella galactura*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



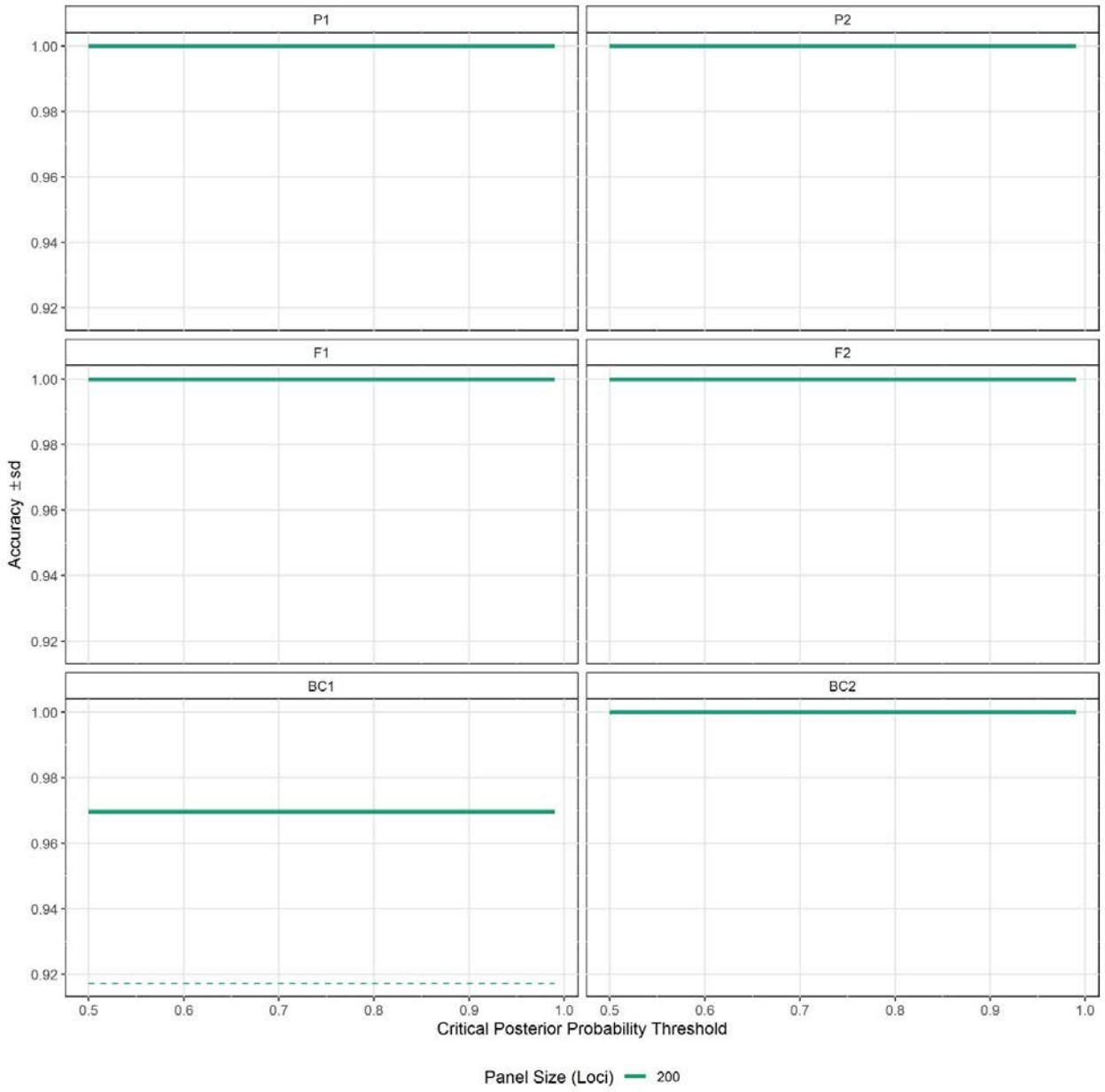
Supplement S10: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Cyprinella whipplei* X *Lythrurus umbratilis*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



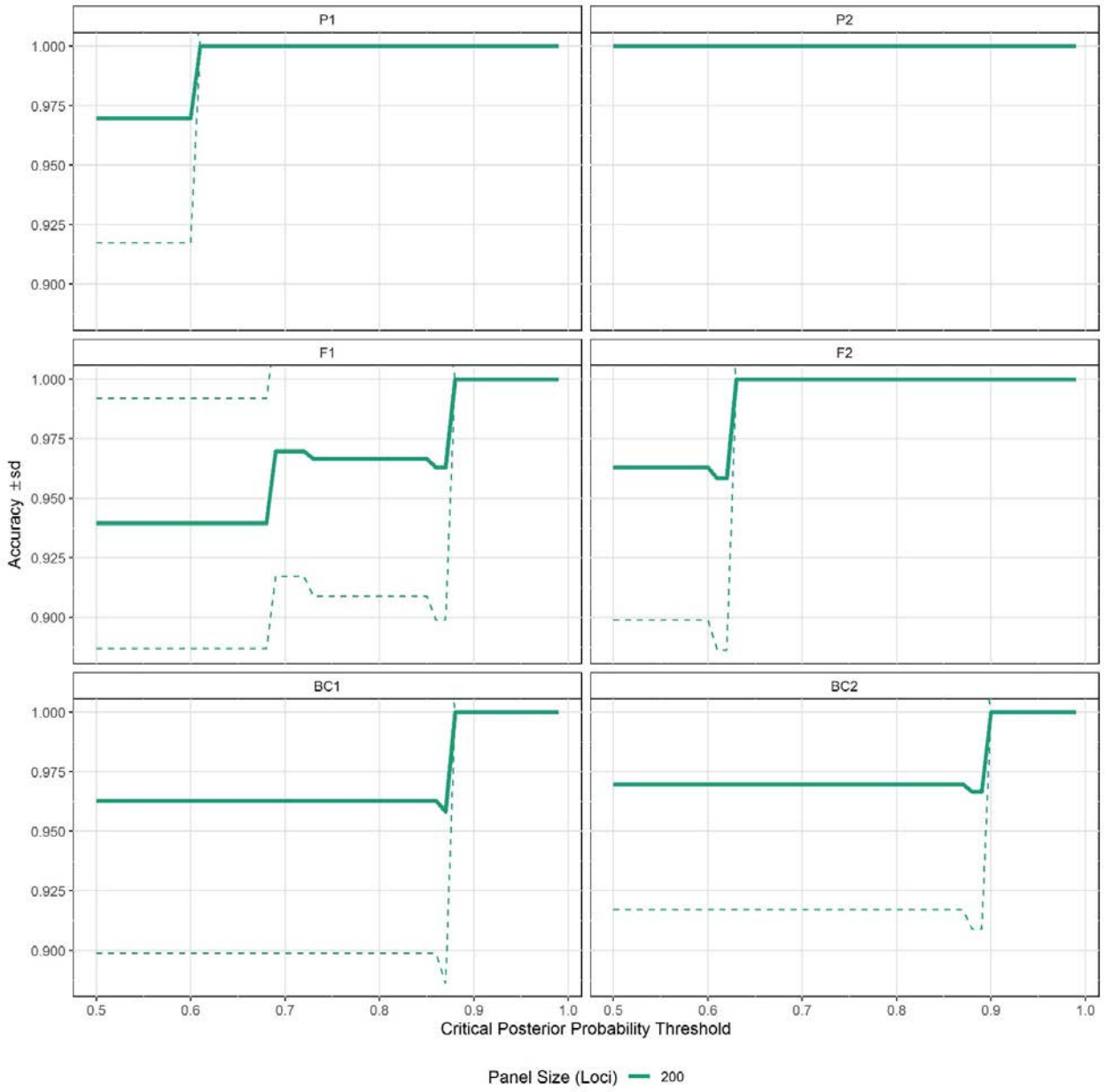
Supplement S11: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus chrysocephalus* X *Luxilus zonatus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



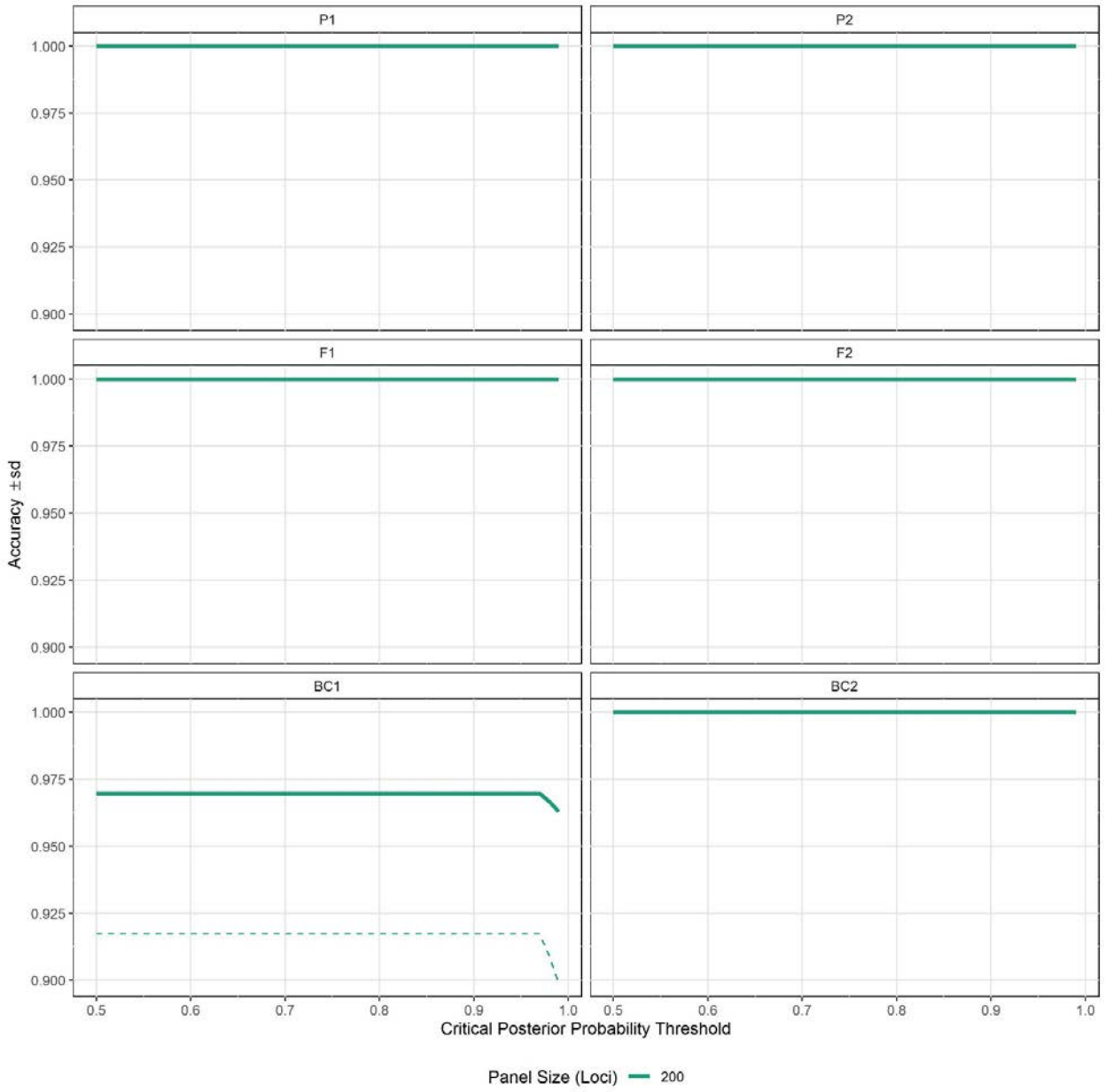
Supplement S12: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus chrysocepalus* X *Semotilus atromaculatus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



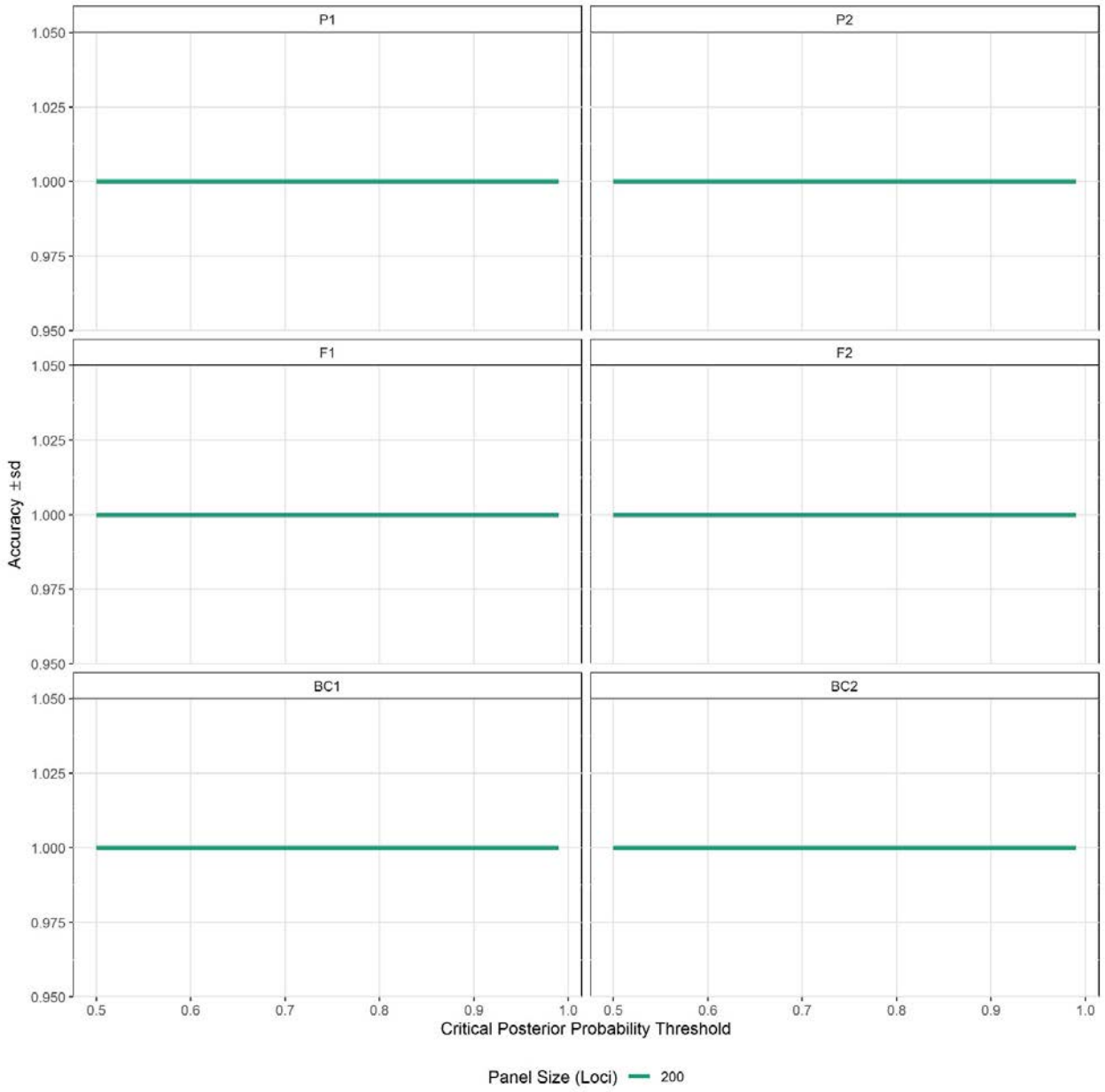
Supplement S13: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus pilsbryi* X *Luxilus chrysocephalus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



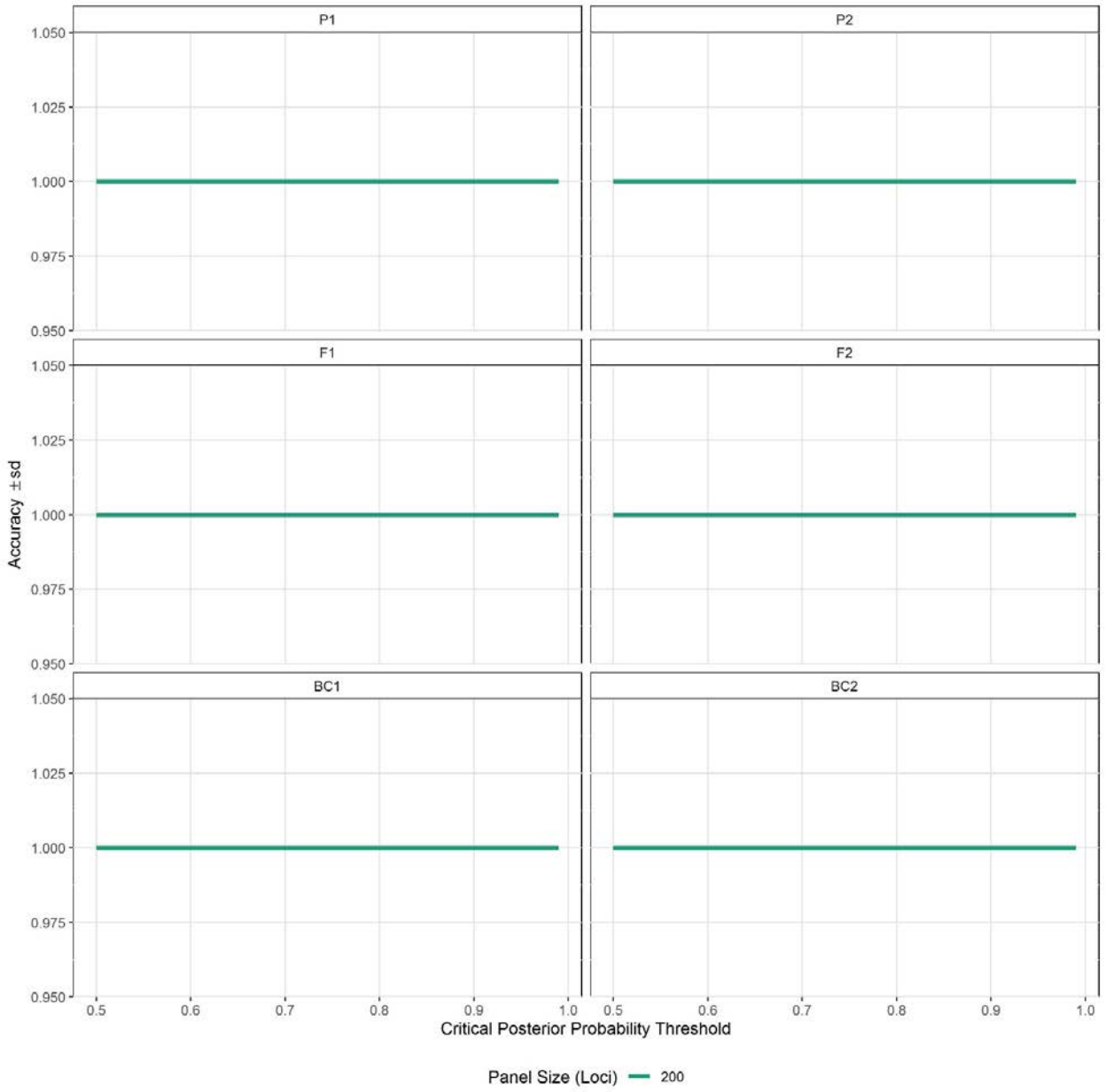
Supplement S14: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus pilsbryi* X *Luxilus zonatus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



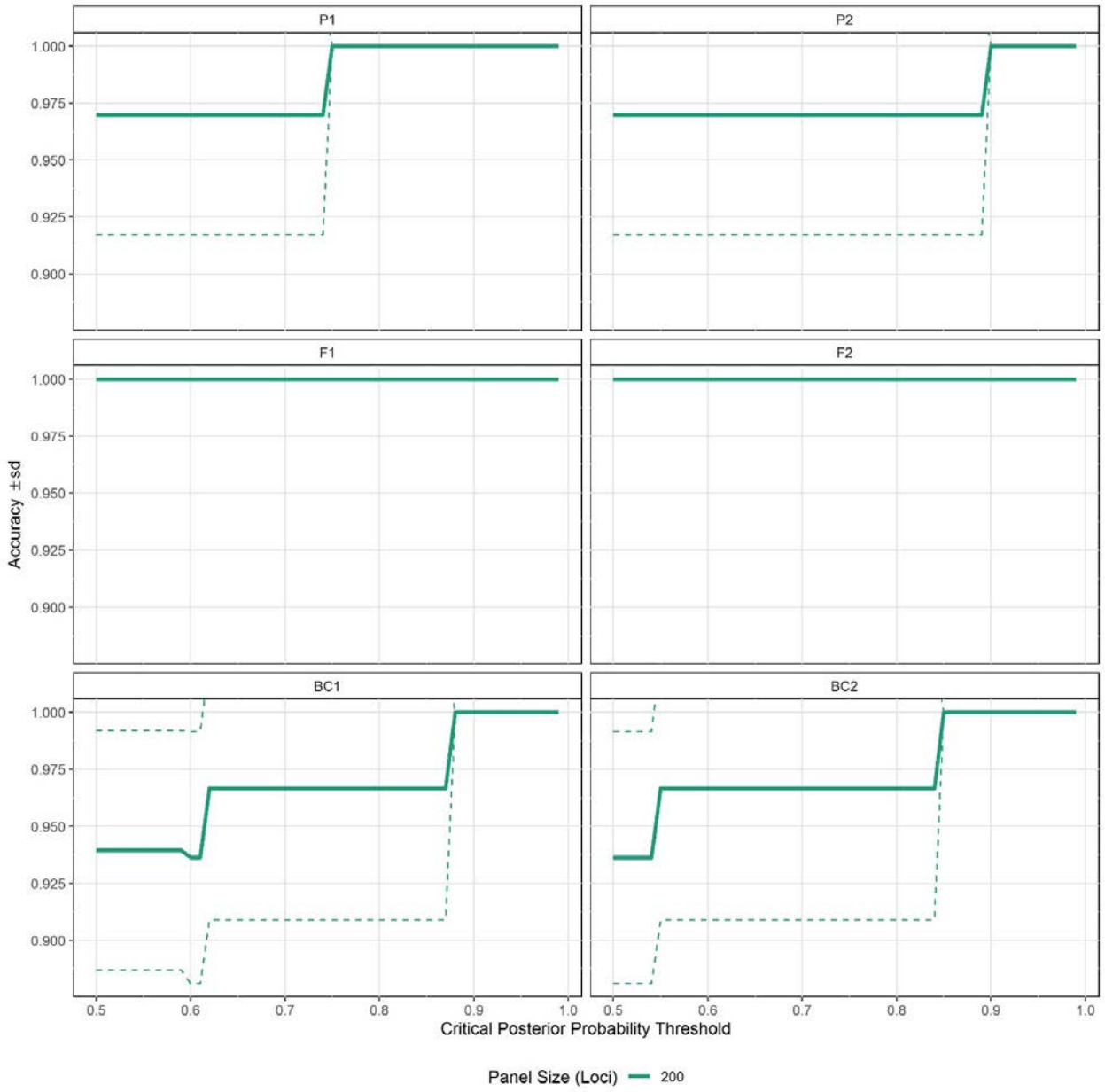
Supplement S15: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus pilsbryi* X *Lythrurus umbratilis*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



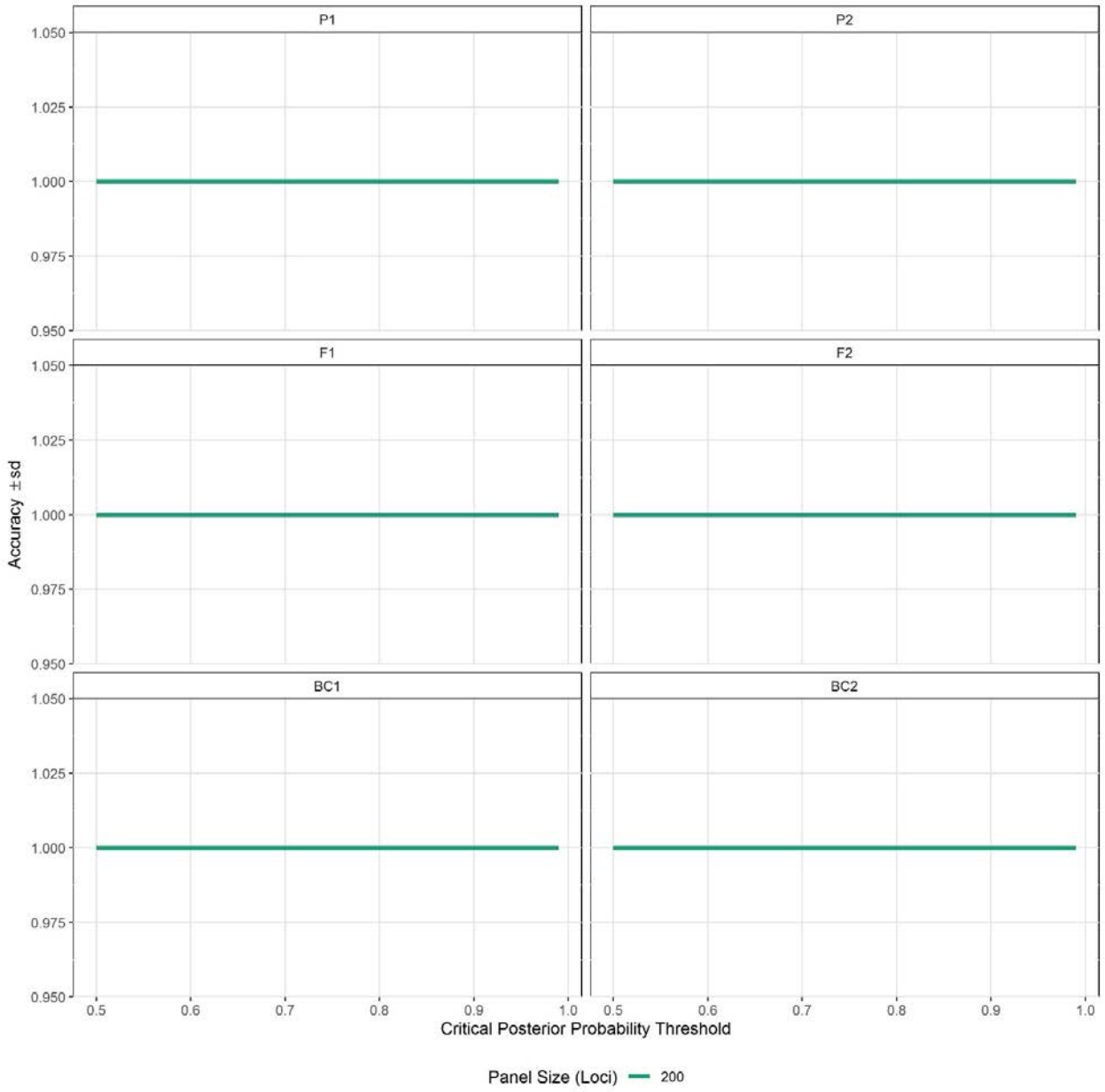
Supplement S16: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus pilsbryi* X *Notropis percobromus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



Supplement S17: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Luxilus zonatusi* X *Pimephales notatus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



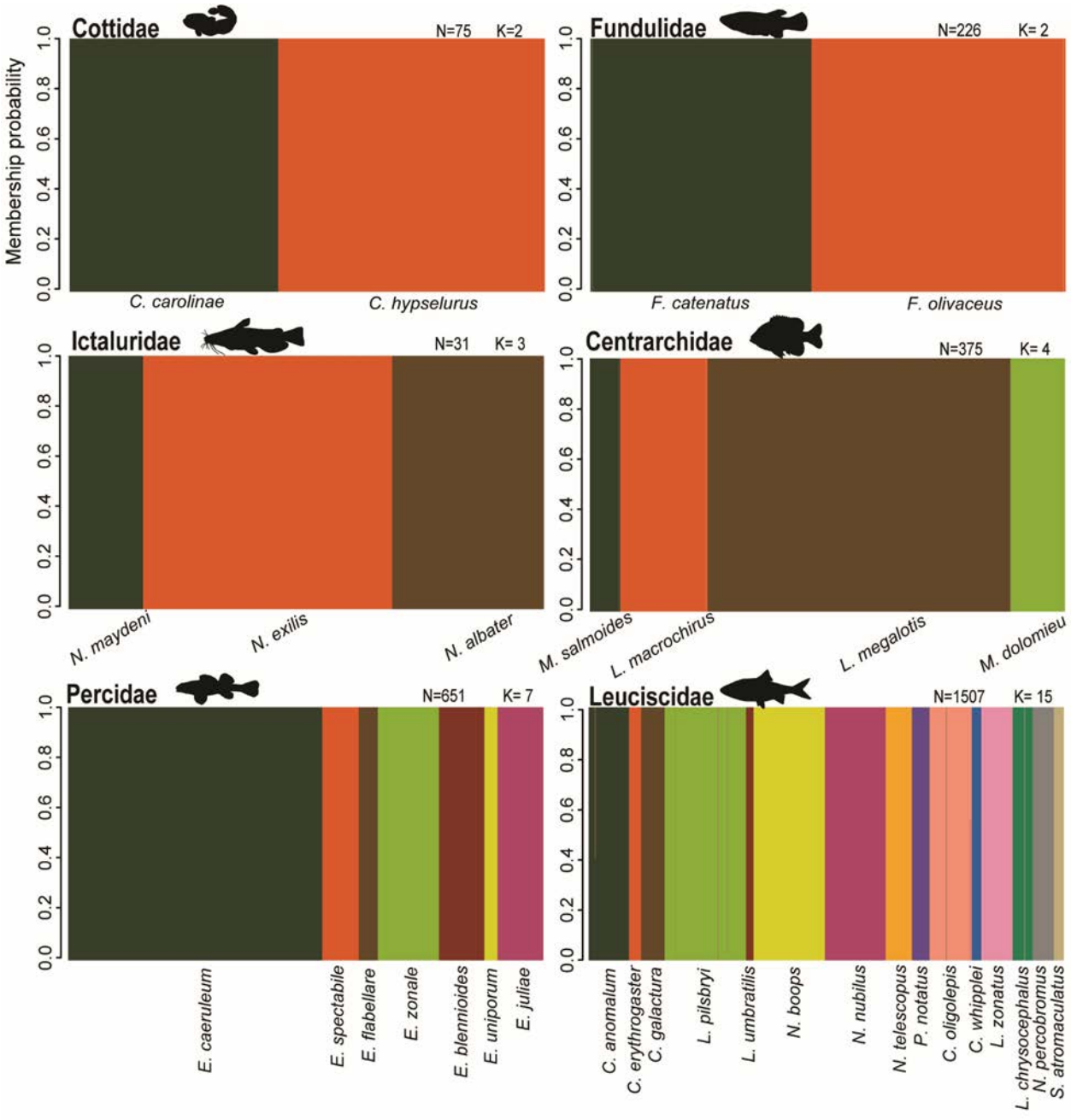
Supplement S18: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Notropis boops* X *Notropis nubilus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.



Supplement S19: Power analysis for detecting genotype frequency classes via HYBRIDDETECTIVE workflow. Results are shown here for *Pimephales notatus* X *Semotilus atromaculatus*. P1 and P2 = pure; F1 and F2 = first filial and second filial hybrids; BC1 and BC2 = backcrosses.

Supplement S20: List of 18 unique hybridizing fish species pairs observed in the White River Basin, USA. Google Scholar was used to search existing literature. Each pair was searched using "hybrid" and the two specific epithets for the pair (e.g., +hybrid* anomalum oligolepis). Searches were also conducted based on related species. Corush et al. (2021) conducted a literature review and compiled known hybrids from museum records and provided an invaluable resource here. Note this list does not include 8 of the 70 total putative hybrid individuals found: multispecies hybrids (more than two ancestral species detected).

No.	Species A	Species B	N Indiv.	Literature Match	Species	Reference
1	*Campostoma anomalum*	*Campostoma oligolepis*	29	yes	*C. anomalum* x *C. oligolepis*	Rakocinski, 1980
2	*Campostoma anomalum*	*Chrosomus erythrogaster*	1	yes	*C. anomalum* x *C. erythrogaster*	Grady & Cashner, 1988
3	*Campostoma anomalum*	*Luxilus pilsbryi*	1	related	*C. anomalum* x *L. chrysocephalus*	Poly, 1997
4	*Campostoma oligolepis*	*Notropis telescopus*	1	no	-	-
5	*Cyprinella galactura*	*Cyprinella whipplei*	2	yes	*C. galactura* x *C. whipplei*	Corush et al., 2021
6	*Cyprinella whipplei*	*Lythrurus umbratilis*	1	no	-	-
7	*Luxilus chrysocephalus*	*Luxilus zonatus*	2	yes	*L. chrysocepalus* x *L. zonatus*	Corush et al., 2021
8	*Luxilus chrysocephalus*	*Semotilus atromaculatus*	1	yes	*L. chrysocepalus* x *S. atromaculatus*	Corush et al., 2021
9	*Luxilus pilsbryi*	*Lythrurus umbratilis*	6	related	*L. chrysocephlus* x *L. umbratilis*	Corush et al., 2021
10	*Luxilus pilsbryi*	*Notropis percobromus*	2	yes	*L. pilsbryi* x *N. rubellus*	Cross 1954
11	*Luxilus pilsbryi*	*Luxilus zonatus*	8	related	*L. zonatus* x *L. chrysocephalus*	Corush et al., 2021
12	*Luxilus pilsbryi*	*Luxilus chrysocephalus*	1	yes	*L. pilsbryi* x *L. chrysocepalus*	Corush et al., 2021
13	*Luxilus zonatus*	*Pimephales notatus*	1	no	-	-
14	*Notropis boops*	*Notropis nubilus*	2	yes	*N. boops* x *N. nubilus*	Corush et al., 2021
15	*Pimephales notatus*	*Semotilus atromaculatus*	1	no	-	-
16	*Micropterus salmoides*	*Micropterus dolomieu*	1	yes	*M. salmoides* x *M. dolomieu*	Barthel et al., 2010
17	*Etheostoma spectabile*	*Etheostoma caeruleum*	1	yes	*E. spectabile* x *E. caeruleum*	Keck & Near, 2009
18	*Etheostoma juliae*	*Etheostoma zonale*	1	no	-	-



Supplement S21: DAPC-based membership probabilities to genomic clusters *K* based on *a priori* species identification for 2,865 individuals from 33 fish species grouped by family. Note the number of individuals per family (*N*) and the number of species (*K*). DAPC excels at discriminating between groups but was largely unable to detect mixed individuals found by other methods (PCA, Snapclust, and sNMF).

Supplement S22: Summary of *N*=70 hybrid individuals form the White River Basin of the Ozarks. Species codes provided below.

INDIVIDUAL	SPECIES	LATITUDE	LONGITUDE	INFERRED PARENTAL PAIR	sNMF	SNAPCLUST	DAPC	NEWHYBRIDS	POSTERIOR
Z01CAMANO01	*Campostoma anomalum*	35.90463	-91.63537	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z02CAMANO01	*Campostoma anomalum*	36.09227	-91.75397	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	BC1	89%
Z03CAMOLI01	*Campostoma oligolepis*	36.20048	-91.759	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO	F1	100%
Z06CAMANO01	*Campostoma anomalum*	36.27342	-91.33447	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z10CAMANO01	*Campostoma anomalum*	35.61367	-91.60728	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	BC1	72%
Z10CAMANO02	*Campostoma anomalum*	35.61367	-91.60728	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC1	100%
Z11CAMANO01	*Campostoma anomalum*	35.5444	-91.56472	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z11CAMANO02	*Campostoma anomalum*	35.5444	-91.56472	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z11CAMANO03	*Campostoma anomalum*	35.5444	-91.56472	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z18CAMANO01	*Campostoma anomalum*	35.54105	-91.77822	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z18CAMANO02	*Campostoma anomalum*	35.54105	-91.77822	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z24CAMANO05	*Campostoma anomalum*	36.14685	-92.07063	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	F1	99%
Z25CAMANO01	*Campostoma anomalum*	36.33653	-92.47492	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z27CAMANO01	*Campostoma anomalum*	36.06698	-92.63615	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z29CAMOLI01	*Campostoma oligolepis*	36.02182	-92.67905	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC2	100%
Z33CAMANO02	*Campostoma anomalum*	36.23702	-92.98725	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC1	100%
Z33CAMOLI02	*Campostoma oligolepis*	36.23702	-92.98725	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	F1	100%
Z34CAMOLI02	*Campostoma oligolepis*	36.15450	-92.90213	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO	F1	100%
Z36CAMANO01	*Campostoma anomalum*	35.94370	-93.06555	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z36CAMOLI01	*Campostoma oligolepis*	35.94370	-93.06555	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	F1	100%
Z37CAMOLI04	*Campostoma oligolepis*	35.96167	-93.23573	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC2	100%
Z37CAMOLI05	*Campostoma oligolepis*	35.96167	-93.23573	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC2	95%
Z42CAMANO09	*Campostoma anomalum*	36.13102	-93.94767	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC1	100%
Z47CAMANO01	*Campostoma anomalum*	36.42222	-93.40465	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC1	100%
Z58CAMANO01	*Campostoma anomalum*	37.45613	-91.69537	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z67CAMANO01	*Campostoma anomalum*	36.88635	-92.47363	CAMANO-CAMOLI	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	BC1	88%
Z68CAMANO01	*Campostoma anomalum*	37.02342	-92.76715	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z71CAMANO01	*Campostoma anomalum*	37.18145	-93.37052	CAMANO-CAMOLI	CAMANO-CAMOLI	PURE	PURE	PURE	100%
Z55CAMANO06	*Campostoma anomalum*	37.35483	-90.97105	CAMANO-CHRERY	CAMANO-CHRERY	CAMANO-CHRERY	PURE	F1	100%
Z52CAMANO02	*Campostoma anomalum*	36.82858	-90.81385	CAMANO-LUXPIL	CAMANO-LUXPIL	CAMANO-CHRERY	PURE	F2	100%
Z69CAMOLI04	*Campostoma oligolepis*	37.04205	-93.13870	CAMOLI-LUXPIL-NOTNUB	CAMOLI-LUXPIL-NOTNUB	LYTUMB-CAMOLI	CAMOLI-NOTNUB	NA	NA
Z04CAMOLI04	*Campostoma oligolepis*	36.38583	-91.80977	CAMOLI-NOTTEL	CAMOLI-NOTTEL	LYTUMB-CAMOLI	PURE	F1	96%
Z06CYPGAL02	*Cyprinella galactura*	36.27342	-91.33447	CYPGAL-CYPWHI	CYPGAL-CYPWHI	CYPGAL-CYPWHI	PURE	BC1	91%
Z06CYPGAL07	*Cyprinella galactura*	36.27342	-91.33447	CYPGAL-CYPWHI	CYPGAL-CYPWHI	CYPGAL-CYPWHI	PURE	F1	99%
Z41CYPWHI02	*Cyprinella galactura*	36.04162	-93.70483	CYPWHI-LYTUMB	CYPWHI-LYTUMB	LYTUMB-CYPWHI	PURE	BC2	100%
Z48ETHJUL05	*Etheostoma juliae*	36.25182	-93.44542	ETHJUL-ETHZON	ETHJUL-ETHZON	ETHJUL-ETHFLA	PURE	BC2	100%
Z01ETHSPE20	*Etheostoma spectabile*	35.90463	-91.63537	ETHSPE-ETHCAE	ETHSPE-ETHCAE	ETHSPE-ETHUNI	PURE	BC2	100%
Z41LUXCHR10	*Luxilus chrysocephalus*	36.04162	-93.70483	LUXCHR-LUXPIL-CYPGAL	LUXCHR-LUXPIL-CYPGAL	LYTUMB-LUXCHR	PURE	NA	NA
Z52LUXCHR03	*Luxilus chrysocephalus*	36.82858	-90.81385	LUXCHR-LUXZON	LUXCHR-LUXZON	LYTUMB-LUXCHR	LUXZON	F1	100%
Z62LUXCHR02	*Luxilus chrysocephalus*	36.55238	-91.53942	LUXCHR-LUXZON	LUXCHR-LUXZON	PURE	PURE	PURE	100%
Z05LUXCHR01	*Luxilus chrysocephalus*	36.44377	-91.67055	LUXCHR-NOTNUB-LUXPIL	LUXCHR-NOTNUB-LUXPIL	LYTUMB-LUXCHR	PURE	NA	NA
Z38LUXPIL08	*Luxilus pilsbryi*	36.03938	-93.33633	LUXPIL-LUXCHR	LUXPIL-LUXCHR	LUXPIL-LYTUMB	PURE	BC1	99%
Z31LUXPIL02	*Luxilus pilsbryi*	35.92747	-92.89285	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z32LUXPIL02	*Luxilus pilsbryi*	36.45005	-93.07547	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z40LUXPIL08	*Luxilus pilsbryi*	36.14380	-93.59412	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z42LUXPIL10	*Luxilus pilsbryi*	36.13102	-93.94767	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z44LUXPIL07	*Luxilus pilsbryi*	35.82850	-93.83217	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z46LUXPIL02	*Luxilus pilsbryi*	35.98301	-94.17291	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z46LUXPIL03	*Luxilus pilsbryi*	35.98301	-94.17291	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z51LUXZON06	*Luxilus zonatus*	36.57930	-91.04640	LUXPIL-LUXZON	LUXPIL-LUXZON	PURE	PURE	PURE	100%
Z24LUXPIL06	*Luxilus pilsbryi*	36.14685	-92.07063	LUXPIL-NOTPER	LUXPIL-NOTPER	LUXPIL-LYTUMB	PURE	BC1	99%
Z31LUXPIL03	*Luxilus pilsbryi*	35.92747	-92.89285	LUXPIL-NOTPER	LUXPIL-NOTPER	LUXPIL-LYTUMB	PURE	F1	100%
Z58LUXZON06	*Luxilus zonatus*	37.45613	-91.69537	LUXZON-PIMNOT	LUXZON-PIMNOT	LYTUMB-LUXZON	PURE	F1	100%
Z52LYTUMB10	*Lythrurus umbratilis*	36.82858	-90.81385	LYTUMB-NOTNUB-LUXPIL	LYTUMB-NOTNUB-LUXPIL	LYTUMB-LUXCHR	PURE	NA	NA
Z15MICSAL03	*Micropterus salmoides*	34.36425	-91.23615	MICSAL-MICDOL	MICSAL-MICDOL	MICSAL-MICDOL	PURE	F1	100%
Z19CAMOLI02	*Campostoma oligolepis*	35.73682	-92.10778	CAMANO-CAMOLI	PURE	CAMANO-CAMOLI	PURE	BC2	99%
Z24LUXPIL01	*Luxilus pilsbryi*	36.14685	-92.07063	LUXPIL-LYTUMB	PURE	LUXPIL-LYTUMB	PURE	BC1	98%
Z29LUXPIL06	*Luxilus pilsbryi*	36.02182	-92.67905	LUXPIL-LYTUMB	PURE	LUXPIL-LYTUMB	NOTNUB	BC1	95%
Z35LUXPIL05	*Luxilus pilsbryi*	35.98273	-93.04035	LUXPIL-LYTUMB	PURE	LUXPIL-LYTUMB	PURE	BC1	100%

Supplement S22 (Cont.)

INDIVIDUAL	SPECIES	LATITUDE	LONGITUDE	INFERRED PARENTAL PAIR	sNMF	SNAPCLUST	DAPC	NEWHYBRIDS	POSTERIOR
Z48LUXPIL04	*Luxilus pilsbryi*	36.25182	-93.44542	LUXPIL-LYTUMB	PURE	LUXPIL-LYTUMB	PURE	BC1	92%
Z58LUXCHR02	*Luxilus chrysocephalus*	37.45613	-91.69537	LUXCHR-SEMATR	PURE	LUXCHR-SEMATR	PURE	BC1	100%
Z67LUXPIL02	*Luxilus pilsbryi*	36.88635	-92.47363	LUXPIL-LYTUMB	PURE	LUXPIL-LYTUMB	NOTNUB	BC1	86%
Z48NOTBOO03	*Notropis boops*	36.25182	-93.44542	NOTBOO-LUXPIL-PIMNOT	NOTBOO-LUXPIL-PIMNOT	LYTUMB-NOTBOO	PURE	NA	NA
Z07NOTBOO09	*Notropis boops*	35.8711	-91.31063	NOTBOO-NOTNUB	NOTBOO-NOTNUB	PURE	PURE	PURE	100%
Z41NOTBOO06	*Notropis boops*	36.04162	-93.70483	NOTBOO-NOTNUB	NOTBOO-NOTNUB	NOTBOO-NOTNUB	PURE	BC1	95%
Z04NOTMAY01	*Noturus maydeni*	36.38583	-91.80977	NOTMAY-NOTEXI-NOTALB	NOTMAY-NOTEXI-NOTALB	PURE	PURE	NA	NA
Z61NOTNUB11	*Notropis nubilus*	36.85519	-91.67677	NOTNUB-LUXPIL-NOTPER	NOTNUB-LUXPIL-NOTPER	NOTNUB-SEMATR	PURE	NA	NA
Z51NOTNUB06	*Notropis nubilus*	36.57930	-91.04640	NOTNUB-LUXPIL-SEMATR	NOTNUB-LUXPIL-SEMATR	NOTNUB-SEMATR	PURE	NA	NA
Z43PIMNOT02	*Pimephales notatus*	36.04882	-93.97490	PIMNOT-SEMATR	PIMNOT-SEMATR	PIMNOT-SEMATR	PURE	F2	93%

Supplement S23: Species codes used in the individual summary table above and throughout code and files.

Family	Code	Common Name	Scientific Name
Centrarchidae (Sunfish)	LEPMAC	Bluegill Sunfish	*Lepomis macrochirus*
	LEPMEG	Longear Sunfish	*Lepomis megalotis*
	MICDOL	Smallmouth Bass	*Micropterus dolomieu*
	MICSAL	Largemouth Bass	*Micropterus salmoides*
Cottidae (Sculpin)	COTCAR	Banded Sculpin	*Cottus carolinae*
	COTHYP	Knobfin Sculpin	*Cottus hypselurus*
Fundulidae (Topminnows)	FUNCAT	Northern Studfish	*Fundulus catenatus*
	FUNOLI	Blackspotted Topminnow	*Fundulus olivaceus*
Ictaluridae (Catfish)	NOTALB	Ozark Madtom	*Noturus albater*
	NOTEXI	Slender Madtom	*Noturus exilis*
	NOTMAY	Black River Madtom	*Noturus maydeni*
Leuciscidae (Minnows)	CAMANO	Central Stoneroller	*Campostoma anomalum*
	CAMOLI	Largescale Stoneroller	*Campostoma oligolepis*
	CHRERY	Southern Redbelly Dace	*Chrosomus erythrogaster*
	CYPGAL	Whitetail Shiner	*Cyprinella galactura*
	CYPWHI	Steelcolor Shiner	*Cyprinella whipplei*
	LUXCHR	Striped Shiner	*Luxilus chrysocephalus*
	LUXPIL	Duskystrip Shiner	*Luxilus pilsbryi*
	LUXZON	Bleeding Shiner	*Luxilus zonatus*
	LYTUMB	Redfin Shiner	*Lythrurus umbratilis*
	NOTBOO	Bigeye Shiner	*Notropis boops*
	NOTNUB	Ozark Minnow	*Notropis nubilus*
	NOTPER	Carmine Shiner	*Notropis percobromus*
	NOTTEL	Telescope Shiner	*Notropis telescopus*
	PIMNOT	Bluntnose Minnow	*Pimephales notatus*
	SEMATR	Creek Chub	*Semotilus atromaculatus*
Percidae (Darters)	ETHBLE	Greenside Darter	*Etheostoma blennioides*
	ETHCAE	Rainbow Darter	*Etheostoma caeruleum*
	ETHFLA	Fantail Darter	*Etheostoma flabellare*
	ETHJUL	Yolk Darter	*Etheostoma juliae*
	ETHSPE	Orangethroat Darter	*Etheostoma spectabile*
	ETHUNI	Current Darter	*Etheostoma uniporum*
	ETHZON	Banded Darter	*Etheostoma zonale*

CHAPTER II

Riverscape community genomics:

A comparative analytical framework to identify common drivers of genetic structure

Zachery D. Zbinden, Marlis R. Douglas, Tyler K. Chafin, Michael E. Douglas

ABSTRACT

Genetic diversity is a key component of population persistence. However, most genetic investigations focus on a single species, overlooking opportunities for multispecies conservation plans to benefit entire communities in an ecosystem. A comprehensive genetic management plan should elevate decision criteria to the community-level, and be informed by multispecies assessments, i.e., genetic studies replicated among several species. We developed a framework to evaluate genetic diversity within and among many species and demonstrate how this *riverscape community genomics* approach can be applied to identify common drivers of genetic structure. Our study evaluated genetic diversity in 31 co-distributed native stream fishes sampled from 75 sites across the White River Basin (Ozark Mountains, U.S.A.) using SNP genotyping (ddRAD). Despite variance in genetic divergence, general spatial patterns were identified corresponding to river network architecture. Most species (N=24) were partitioned into discrete sub-populations (K=2–7). We used partial redundancy analysis to compare species-specific genetic diversity across four models of genetic structure: Isolation by distance (IBD), isolation by barrier (IBB), isolation by stream hierarchy (IBH), and isolation by environment (IBE). A significant proportion of intraspecific genetic variation was explained by IBH ($\bar{x}$ = 62%), with the remaining models generally redundant. Our results indicated that gene flow is higher within rather than between hierarchical units (i.e., catchments, watersheds, basins), supporting the Stream

Hierarchy Model and its generality. We discuss our conclusions regarding conservation and management and identify the 8-digit Hydrologic Unit (HUC) as the most relevant spatial scale for managing genetic diversity across riverine networks.

INTRODUCTION

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

When the concept of genetic diversity is applied in a comparative sense across co-distributed species, it provides a solid framework from which community-wide management and policy can be defined. For example, multispecies assessments can reveal common dispersal barriers (Pilger et al., 2017; Roberts et al., 2013), congruent distributions of genetic diversity (Hotaling et al., 2019; Ruzich et al., 2019), relevant spatial scales for management (Blanchet et al., 2020), and associations among species characteristics and genetic diversity (Bohonak, 1999; Pearson et al., 2014). Despite its potential complexity, a comprehensive management strategy

can emerge, one more appropriately aligned towards managing numerous species, with long-term conservation goals as end-products beneficial to an entire community (Blanchet et al., 2017). In addition, it also tacitly encourages support by stakeholders for an overarching management plan, one representing a consensus across multiple species and ecosystems (Douglas et al., 2020).

The spatial structure of genetic variation is primarily dictated by gene flow and genetic drift within a species (Holderegger et al., 2006), with the uniformity of its distribution (i.e., panmixia; Rosenberg et al., 2005) serving as an implicit null hypothesis. The *de facto* alternative is that genetic variation is spatially autocorrelated (i.e., 'isolation by distance,' IBD; Wright, 1943). For most species, a significant relationship between genetic dissimilarity and geographic distance is the expectation (Meirmans, 2012), yet the strength of this association will vary (Bohonak, 1999; Singhal et al., 2018). For example, genetic divergence may be promoted by environmental dissimilarities (i.e., 'isolation by environment,' IBE; Wang & Bradburd, 2014) or by physical barriers to dispersal (i.e., 'isolation by barrier,' IBB; Cushman et al., 2006; Ruiz-Gonzalez et al., 2015).

For aquatic biodiversity, patterns of genetic divergence will also be governed by the structure and architecture of the riverine network (in contemporary and historic representations). Organisms within such dendritic networks are demonstrably impacted by the physical structure of the habitat (Peterson et al., 2013; White et al., 2020), with genetic relatedness as a surrogate for the underlying structural hierarchy (Hughes et al., 2009). While this is most apparent within the contemporary structure of river networks, their historic structure, i.e., paleohydrology, also serves to bookmark genetic diversity (Mayden, 1988; Strange & Burr, 1997). Moreover, the hierarchical complexity of these networks will likewise dictate population processes, as reflected

within genetic diversities and divergences (Chiu et al., 2020; Hopken et al., 2013; Thomaz et al., 2016). Thus, spatial genetic structuring within such biodiversity should reflect 'isolation by stream hierarchy' (IBH; *sensu* Stream Hierarchy Model (SHM); Meffe & Vrijenhoek, 1988). The initial genesis for the SHM was narrowly-defined and cast within a relatively simplistic ecosystem, e.g., desert stream fishes within the American West (Meffe and Vrijenhoek 1988). As compared to alternative isolating regimes, an assessment of its generality was thus imperative (Brauer et al., 2018; Hopken et al., 2013).

The mechanisms that cause genetic structure can not only be confounding on the one hand (Perez et al., 2018) but also correlated on the other (Meirmans, 2012; Wang & Bradburd, 2014). Different mechanisms can mask the occurrence of major drivers by promoting those more ancillary with regard to single-species assessments. The emerging results are twofold: Potentially erroneous conclusions, which in turn beget ineffective management strategies. However, these issues can be effectively mitigated using replicated multispecies assessments to allow influential major processes to surface, thus effectively categorizing both 'signal and noise' components with the former driving patterns of regional biodiversity (Roberts et al., 2013).

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

This approach also allows the appropriate scale to be defined at which genetic and conservation perspectives can be integrated to optimize benefits across species.

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

METHODS

Study system

Our study system, the White River Basin, is located within the Western Interior Highlands of North America, a previous component of the more extensive pre-Pleistocene Central Highlands extending north and east but subsequently subsumed by numerous glacial advances into two disjunct sub-components: Western Interior Highlands (i.e., Ozark Plateau, Ouachita Mountains), and Eastern Highlands (i.e., Appalachian Plateau, Blue Ridge, Appalachian Highlands) (Mayden,

1985). The Ozark Plateau remained an unglaciated refugium with elevated endemism and diversity (Warren et al., 2000). The White River Basin was established by at least Late Pliocene (>3 MYA; Jorgensen, 1993), but its eastern tributaries were captured by the Mississippi River when it bisected the basin during the Pleistocene (Mayden, 1988; Strange & Burr, 1997). This paleohydrologic signature may remain in contemporary patterns of population divergence in the White River Basin, as manifested by replicated patterns of genetic structure between eastern and western populations.

Sampling

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

Sampling was approved by the University of Arkansas Institutional Animal Care and Use Committee (IACUC: #17077), with collecting permits as follows: Arkansas Game & Fish Commission (#020120191); Missouri Department of Wildlife Conservation (#18136); U.S. National Parks Service (Buffalo River Permit; BUFF-2017-SCI-0013). Fishes were sampled using seine nets in wadable streams during low flow between June 2017 and September 2018. Time spent sampling a site ranged from 30–60 mins, with a target of 5-10 individuals/species encountered. Individuals were euthanized by immersion in tricaine methanesulfonate (MS-222)

at a concentration of 500 mg/L, buffered to pH=7 with subsequent preservation in 95% ethanol. Species diagnosis occurred in the laboratory following Eschmeyer's Catalog of Fishes. The right pectoral fin was removed and stored in 95% ethanol at -20 °C prior to subsequent DNA extraction. Specimens are housed at the Arkansas Conservation and Molecular Ecology Lab, University of Arkansas, Fayetteville.

Genomic data collection and filtering

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

Raw Illumina reads were demultiplexed, clustered, filtered, and aligned in IPYRAD v.0.9.62 (Eaton & Overcast, 2020). Reads were first demultiplexed, allowing up to one barcode

mismatch, yielding individual FASTQ files containing raw reads (*N*=3,060 individual files). Individuals averaged >2 million reads, with those extremely low removed ($< \bar{x} - 2s$) to reduce errors from poor quality sequencing. Individuals were screened for putative hybrids (Zbinden, Douglas, et al., 2022), and those with admixed ancestry were removed. Raw sequence reads were partitioned by species (*N*=31) and aligned *de novo* in IPYRAD (Eaton & Overcast, 2020). Adapters/primers were removed, and reads with >5 bases having Phred quality <20 or read length <35 bases (after trimming) were discarded. Clusters of homologous loci were assembled using an 85% identity threshold. Putative homologs were removed if any of the following were met: <20x and >500x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of nucleotides polymorphic; >8 indels present; or presence in <15% of individuals. Paralogs were identified (and subsequently removed) as those clusters exhibiting either >2 alleles per site in consensus sequence or excessive heterozygosity (>5% of consensus bases or >50% heterozygosity/site among individuals).

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

Genetic structure

Genetic structure was assessed using the resultant SNP genotypes. For each species (N=31), pairwise F_{ST} (Weir & Cockerham, 1984) was calculated among sites (HIERFSTAT; Goudet et al., 2017). Jost's D was also quantified among sites and globally, as it is based on the effective number of alleles rather than heterozygosity and hence less biased by sampling differences (Jost, 2008). Additional global intraspecific F_{ST} analogs were also quantified for comparison: Multi-allelic G_{ST} (Nei, 1973) and unbiased G''_{ST} (Meirmans & Hedrick, 2011) (MMOD; Winter, 2012). We tested for isolation by distance (IBD) using both linearized F_{ST} and Jost's D. Their relationships with river distance (log-transformed) were tested using the Mantel test (Mantel & Valand, 1970) (ECODIST; Goslee & Urban, 2020), then visualized using linear regression (Rousset, 1997).

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

We further assessed among-site genetic variation between Hydrologic Units (HUCs) and discrete population clusters (determined via sNMF) using analysis of molecular variance (AMOVA) (Excoffier et al., 1992). AMOVA was performed for each species at four HUC levels

(4-, 6-, 8-, and 10-digit) to compare the amount of genetic variation among HUCs, among all sites, and among sites within HUCs. The Watershed Boundary Dataset (USGS, 2021) assigned HUC classifications to each site. AMOVA was then performed for each species with genetic clusters $K>1$ to compare the amount of genetic variation among populations, among all sites, and among sites within populations. The variance components were used to estimate Φ-statistics (analogous to *F*-statistics): Φ_{CT} = the genetic variation among groups (either HUCs or discrete populations); Φ_{ST} = the genetic variation among sites across all groups; and Φ_{SC} = the genetic variation among sites within groups. The wrapper R package POPPR (Kamvar et al., 2015) was used to implement the PEGAS (Paradis, 2010) version of AMOVA with default settings.

Modelling genetic structure

We employed a variation partitioning framework (Capblancq & Forester, 2021; Chan & Brown, 2020) to compare four models of genetic structure for each species based on: IBD, IBB, IBH, and IBE. Individual neutral genetic variation within each species was calculated using principal components analysis (PCA) on each SNP panel. The appropriate number of PCs retained for each species was based on observed eigenvalues, Rnd-Lambda (Peres-Neto et al., 2005), implemented in the R package PCDIMENSION (Coombes & Wang, 2019). Individual scores on the retained PCs represented individual genetic variation.

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

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

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

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

In summary: Variables were first tested for a relationship with the response data (individual genetic variation) using redundancy analysis (RDA). If the relationship was significant ($\alpha < 0.05$), a stepwise forward procedure was carried out such that variables were selected if the adjusted R^2 of the model increased significantly ($\alpha < 0.05$) and the adjusted R^2 did

not exceed that of the overall model. This procedure was employed using the *ordiR2step* function in the R package VEGAN (Oksanen et al., 2020). The selected variables from each of the five classes were first combined into a single matrix, then reduced to a set of PCs using robust principal components analysis (ROBPCA; Hubert et al., 2005). The number of PCs retained for each category was determined following Hubert and coworkers (2005), as implemented in the R package ROSPCA (Hubert et al., 2016).

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

RESULTS

Sampling and data recovery summation

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

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

data >75% and those as singletons of their species at a site, the remaining N=2,861 were analyzed for genetic structure (Table 1). The number of individuals analyzed per species ranged from 15–358 ($\bar{x}$=92.3; s=80.8), and the sites at which each species was collected ranged from 5–50 ($\bar{x}$=6.8; s=11.2). Number of individuals/species/site ranged from 2–15 ($\bar{x}$=5.1; s=1.5). Mean number of raw reads/individual/species spanned from 1.65 million to 3.22 million ($\bar{x}$=2,289,230.0; s=341,159.5). Mean N of loci/species recovered by IPYRAD ranged from 14,599–30,509 ($\bar{x}$=20,081.7; s=4,697.6) with a mean sequencing depth/locus of 73.6x (s=12.0x). After filtering loci and retaining one SNP per locus, the panels for each species contained 2,168–10,033 polymorphic sites ($\bar{x}$=4,486.7; s=1,931.1) with mean missing data/species at 12% (s=2%).

Genetic structure

Among-site genetic divergence

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

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

(*Micropterus dolomieu*; Lacepède, 1802) and Largemouth Bass [*Micropterus salmoides*; (Lacepède, 1802)]. Mantel correlation coefficients ranged from 0.15–0.92 ($\bar{x}$=0.51; *s*=0.19). The slope of the linear relationship between Jost's *D* and log river network distance for each species ranged from 0.0007 – 0.28 ($\bar{x}$=0.04; *s*=0.06).

Population structure

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

AMOVA

Discrete genetic structuring was also suggested for many species in our AMOVA analyses. Genetic variation among HUCs was significant for 24 species (Table 4). The genetic variance explained for these species by HUCs ranged from 1–70% ($\bar{x}$=25.0%; *s*=20.7%). For the other seven species, variation among HUCs was ≤ 1%, save for Ozark Sculpin (*Cottus hypselurus*; Robins & Robison, 1985) and Creek Chub [*Semotilus atromaculatus*; (Mitchill, 1818)]. HUC differences for these accounted for >80% of the genetic variance but were non-significant due to a lack of power. Southern Redbelly Dace [*Chrosomus erythrogaster*; (Rafinesque, 1820)] could

not be tested due to a lack of repeated samples within HUC levels. Further evidence of genetic structure among HUCs was revealed in the pattern of Φ_{SC} (genetic divergence among sites within HUCs) < Φ_{ST} (divergence among all sites) found across 26 species. The 8-digit HUC level explained the greatest genetic variance across 21 species (Table 4).

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

Models of genetic structure

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

Combining the four models (IBD, IBB, IBH, IBE) accounted for between 3–100% of the neutral genetic diversity across species ($\bar{x}$=63.0%; s=35.3%; Figure 6). Isolation by stream hierarchy (IBH; $\bar{x}$=62.0%; s=34.7%) and barrier (IBB; $\bar{x}$=49.3%; s=30.0%) contributed most to the total variation explained, while distance (IBD; $\bar{x}$=32.1%; s=25.1%) and environment (IBE; $\bar{x}$=33.0%; s=21.4%) explained less (Figure 6). Variation explained by "pure" models, after

accounting for that explained by the other three, was >0 only for stream hierarchy and barrier (Figure 6), suggesting that distance and environment are encapsulated by the former. Indeed, correlative structure among models revealed most genetic variance was explained by stream hierarchy, with the other models largely redundant (Figure 7).

DISCUSSION

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

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

Drivers of isolation at the basin-wide scale

Isolation by Distance and river networks

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

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

Stream Hierarchy Model

Our results show that individual genetic variation is best explained by the Stream Hierarchy Model (Brauer et al., 2018; Hopken et al., 2013; Meffe & Vrijenhoek, 1988). In other words, the

majority of variation explained by IBD, IBE, and IBB could be accounted for by IBH alone. This was corroborated via variation partitioning, in which IBD, IBE, and IBB models were redundant with IBH. A concordance of population structure with stream hierarchy yielded a similar percentage of among-site genetic variation, as explained by among-HUC and among-population groupings. In short, variance explained by distance and environment was due to differences among HUC drainages. These results highlight the necessity of accounting for population structure prior to exploring the relationship between genotypes and environmental heterogeneity, e.g., within genotype by environment frameworks (Lawson et al., 2020).

Disentangling cumulative effects

Our analyses also revealed complex spatial patterns of genetic diversity. We evaluated competing isolation models using a framework that identified distance and barriers as putative drivers, with strong genetic divergence identified even across short geographical distances (Chan & Brown, 2020; Ruiz-Gonzalez et al., 2015). This interaction can confound analyses that incorporate either alone. For example, if sampling is clustered, discrete genetic groups can be spuriously inferred along an otherwise continuous gradient of genetic variation (Frantz et al., 2009). Furthermore, a continuous pattern can be erroneously extrapolated when the underlying reality is described by distinct clusters separated by geographic distance (Meirmans, 2012). Here we echo the importance of testing various hypotheses concerning genetic structure (Perez et al., 2018). Idiosyncrasies and complex interactions cannot be discerned by testing single models in isolation (e.g., discrete structure or IBD).

Drivers of variation within and among species

The species assayed herein display marked differences concerning dispersal capability (Shelley et al., 2021). Given this, we expected genetic structure to widely vary among species across our study region (Comte & Olden, 2018; Husemann et al., 2012; Pilger et al., 2017). Dispersal-related traits drive gene flow among localities and determine the spatial scale at which patterns of genetic structure emerge (Bohonak, 1999; Riginos et al., 2014). The physical structure of the river network then further modulates these patterns by dictating dispersal pathways of metapopulations and their colonization and extinction probabilities (Falke et al., 2012; Labonne et al., 2008; Fagan, 2002). These superimposed processes promote genetic divergence among distal populations (Thomaz et al., 2016; Chiu et al., 2020). Similar patterns emerge when analyzing community diversity via species composition. Headwater streams tend to have very different communities due to dispersal limitations (Finn et al., 2011; Zbinden & Matthews, 2017; Zbinden, Geheber, Lehrter, & Matthews, 2022). Hence the interaction between traits and environment is an overarching influence that unites ecology and evolution.

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

dynamics, with susceptible species, those smaller and less tolerant, diverging most rapidly (Campbell Grant et al., 2007).

Other life-history traits may also play a role as well. For example, those that directly influence effective population size (Nei & Tajima, 1981) may generate differences among species regarding the rate at which genetic differences arise (Blanchet et al., 2020). Species with 'slow' life histories, characterized by longer generations and delayed maturity, show an increased probability of local extirpation, inflating genetic drift concomitant with global extinction risk (Hutchings et al., 2012; Pearson et al., 2014; Chafin et al., 2019). Similar contingencies exist for other ecological traits, such as highly specialized trophic adaptations, narrow environmental tolerances, or those that follow the same general mechanism by predisposing species to fragmented population structures (Olden et al., 2008). Ecological traits are mirrored by morphology (Douglas & Matthews, 1992), underscoring an interaction of trait effects that are difficult to disentangle. Ultimately, intraspecific genetic divergence is driven by a combination of factors that influence population size, demographic history, and connectivity. Clearly, these complex interactions among drivers require more comparative multispecies assessments as they shape genetic diversity and structure within and among species (microevolutionary scale) and thus ultimately lead to speciation and extinction (macroevolutionary scale). Our analytical framework outlined herein provides a template for such community-genomics studies.

Disentangling historical and contemporary drivers

Paleohydrology in the White River system

In this study, discrete population structure coincides with major topological divides within the White River stream network, such as a consistent east/west divide between Upper White and

Black rivers, mirroring prior community composition studies (Matthews & Robison, 1988; 1998). Similar patterns were observed at smaller scales among drainages within the study region, as reported for White River crayfish (Fetzner & DiStefano, 2008). While the Lower White and Black rivers are certainly contemporary large-river habitats, both would have been much larger pre-Pleistocene when together they represented the main channel of the Old Mississippi River (Mayden, 1988; Strange & Burr, 1997). This large-river habitat would have separated the eastern and western highland tributaries, with inhospitable habitat for upland species. Pronounced limitations regarding historic dispersal induced by the Old Mississippi could explain the greater isolation of the Little Red River and Black River tributary populations compared to those in the Upper White River. Here, additional work should incorporate coalescent perspectives (e.g., Oaks, 2019) that test the role of past geomorphic events in driving co-divergence and co-demographic patterns, such as the Pleistocene incursion by the Old Mississippi into the modern Black River channel.

Contemporary drivers

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

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

Conclusion

The multispecies comparative approach employed here revealed general patterns that could not have been discerned from a singular study of any one species. Additionally, the variability in intraspecific genetic structure among species provides a specific, all-encompassing dimension that single-species studies cannot. While meta-analytic frameworks have some potential, they are limited by confounding effects that stem from differences between studies, such as markers, sample sizes, environmental contexts, and historical contingencies. This necessitates a

community-level approach within a study region. Further work aimed at modeling variables can lead to greater insight, ultimately improving our hypotheses regarding genetic diversity for which contemporary data are unavailable.

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

As previously recognized, independent populations warrant independent management (Hopken et al., 2013). When migration is low or non-existent, management of one population is unlikely to impact another. Genetic variation unique to hydrologic units could allow for adaptation to future environmental change, while on the other hand, isolation of populations could underscore elevated extirpation risks (Harrisson et al., 2014). Furthermore, efforts to propagate populations via stocking or translocation should carefully assess the genetic landscape of the species in question, particularly before co-mingling diversity from different sub-basins (Meffe & Vrijenhoek, 1988). Such uninformed mixing of genetic stocks could promote outbreeding and the erosion of unique genetic diversity within river catchments. However, this must be weighed against the risks of local extirpation (Pavlova et al., 2017).

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

REFERENCES

Alexander, D.H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 1–6.

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

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

Blanchet, F.G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, *89*(9), 2623–2632.

Blanchet, S., Prunier, J.G., & De Kort, H. (2017). Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics*, *33*(9), 579–580.

Blanchet, S., Prunier, J.G., Paz-Vinas, I., Saint-Pé, K., Rey, O., Raffard, A., . . . & Dubut, V. (2020). A river runs through it: The causes, consequences, and management of intraspecific diversity in river networks. *Evolutionary Applications*, *13*(6), 1195–1213.

Bohonak, A.J. (1999). Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology*, *74*(1), 21–45.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045–1055.

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.

Campbell Grant, E.H., Lowe, W.H., & Fagan, W.F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology letters*, *10*(2), 165-175.

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

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

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

Comte, L., & Olden, J.D. (2018). Fish dispersal in flowing waters: A synthesis of movement- and genetic-based studies. *Fish and Fisheries*, *19*, 1063–1077.

Coombes, K.R., Wang, M., & Coombes, M. K. R. (2019). PCDimension: finding the number of significant prinicpal components. https://cran.r-project.org/package=PCDimension

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

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

da Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., . . . & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics, 30*, 3–13.

Davis, C.D., Epps, C.W., Flitcroft, R.L., & Banks, M.A. (2018). Refining and defining riverscape genetics: How rivers influence population genetic structure. *WIREs Water*, *5*(2), e1269.

DeWoody, J.A., Harder, A.M., Mathur, S., & Willoughby, J.R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147-4154.

Douglas, M.R., Anthonysamy, W.J.B. Mussmann, S.M., Davis, M.A., Louis, W. and Douglas, M.E. (2020). Multi-targeted management of upland game birds at the agroecosystem interface in midwestern North America. *PLoS ONE* 15(4), e0230735.

Douglas, M.E., and Matthews, W.J. (1992). Does morphology predict ecology? Hypothesis testing within a freshwater stream fish assemblage. *Oikos*, 65(2), 213-224.

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

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

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

Fagan, W.F. (2002). Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology*, *83*(12), 3243-3249.

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

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

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

Fluker, B.L., Kuhajda, B.R., & Harris, P.M. (2014). The influence of life-history strategy on genetic differentiation and lineage divergence in darters (Percidae: Etheostomatinae). *Evolution*, *68*(11), 3199–3216.

Frantz, A.C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*(2), 493–505.

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

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

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

Gosselin, T. (2020). R package radiator: RADseq Data Exploration, Manipulation and Visualization using R. https://thierrygosselin.github.io/radiator/

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

Guillot, G., Leblois, R., Coulon, A., & Frantz, A.C. (2009). Statistical methods in spatial genetics. *Molecular Ecology*, *18*(23), 4734–4756.

Hansen, M.M., Limborg, M.T., Ferchaud, A.L., & Pujolar, J.M. (2014). The effects of medieval dams on genetic divergence and demographic history in brown trout populations. *BMC Evolutionary Biology*, *14*(1), 1–14.

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

Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, *21*(6), 797-807.

Hopken, M.W., Douglas, M.R., & Douglas, M.E. (2013). Stream hierarchy defines riverscape genetics of a North American desert fish. *Molecular Ecology*, *22*(4), 956–971.

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

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

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

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

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

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

Hutchings, J.A., Myers, R.A., García, V.B., Lucifora, L.O., & Kuparinen, A. (2012). Life-history correlates of extinction risk and recovery potential. *Ecological Applications*, *22*(4), 1061-1067.

Jorgensen, D.J. (1993) Paleohydrology of the Central United States. *U.S. Geological Survey Bulletin*, 1989-D, 1-32.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015–4026.

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

Labonne, J., Ravigné, V., Parisi, B., & Gaucherel, C. (2008). Linking dendritic network structures to population demogenetics: the downside of connectivity. *Oikos*, *117*(10), 1479-1490.

Laikre, L. (2010). Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics*, *11*(2), 349–354.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., . . . & Timpson, N. J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics*, 139, 23-41.

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

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

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), 1–15.

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

Mantel, N., & Valand, R.S. (1970). A Technique of Nonparametric Multivariate Analysis. *Biometrics*, *26*(3), 547.

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

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

Matthews, W.J., & Robison, H.W. (1998). Influence of drainage connectivity, drainage area and regional species richness on fishes of the interior highlands in Arkansas. *American Midland Naturalist*, *139*(1), 1–19.

Mayden, R.L. (1985). Biogeography of Ouachita Highland Fishes. *The Southwestern Naturalist, 30*(2), 195–211.

Mayden, R.L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Biology*, *37*(4), 329–355.

Meffe, G.K., & Vrijenhoek, R.C. (1988). Conservation Genetics in the Management of Desert Fishes. *Conservation Biology, 2*(2), 157–169.

Meirmans, P.G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846.

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

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

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

Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*.

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, *98*(3), 625–640.

Oaks, J.R. (2019). Full bayesian comparative phylogeography from genomic data. *Systematic Biology*, 68(3), 371-395.

Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., . . . & Wagner, H. (2020). vegan: community ecology package. https://cran.r-project.org/package=vegan

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

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

Pavlova, A., Beheregaray, L.B., Coleman, R., Gilligan, D., Harrisson, K.A., Ingram, B.A., . . . & Sunnucks, P. (2017). Severe consequences of habitat fragmentation on genetic diversity of an endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications, 10*(6), 531–550.

Paz-Vinas, I., Loot, G., Hermoso, V., Veyssière, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences, 285*, 20172746.

Pearson, R.G., Stanton, J.C., Shoemaker, K.T., Aiello-Lammens, M.E., Ersts, P.J., Horning, N., . . . & Akçakaya, H. R. (2014). Life history and spatial traits predict extinction risk due to climate change. *Nature Climate Change, 4*(3), 217-221.

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

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

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

Peterson, E.E., Ver Hoef, J.M., Isaak, D.J., Falke, J.A., Fortin, M.J., Jordan, C.E., . . . & Wenger, S.J. (2013). Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters, 16*(5), 707–719.

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

Radinger, J., & Wolter, C. (2015). Disentangling the effects of habitat suitability, dispersal, and fragmentation on the distribution of river fishes. *Ecological Applications, 25*(4), 914–927.

Riginos, C., Buckley, Y.M., Blomberg, S.P., & Treml, E.A. (2014). Dispersal Capacity Predicts Both Population Genetic Structure and Species Richness in Reef Fishes. *The American Naturalist*, *184*(1), 52–64.

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

Robison, H.W., & Buchanan, T.M. (2020). *Fishes of Arkansas*. University of Arkansas Press.

Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., & Feldman, M.W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, *1*(6), 660–671.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228.

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

Ruzich, J., Turnquist, K., Nye, N., Rowe, D., & Larson, W.A. (2019). Isolation by a hydroelectric dam induces minimal impacts on genetic diversity and population structure in six fish species. *Conservation Genetics*, *20*(6), 1421–1436.

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

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

Singhal, S., Huang, H., Grundler, M.R., Marchán-Rivadeneira, M.R., Holmes, I., Title, P.O., . . . & Rabosky, D.L. (2018). Does Population Structure Predict the Rate of Speciation? A Comparative Test across Australia's Most Diverse Vertebrate Radiation. *The American Naturalist*, *192*(4), 432–447.

Strange, R.M., & Burr, B.M. (1997). Intraspecific phylogeography of North American highland fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, *51*(3), 885–897.

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

Turner, T.F., & Robison, H.W. (2006). Genetic diversity of the Caddo madtom, Noturus taylori, with comments on factors that promote genetic divergence in fishes endemic to the Ouachita Highlands. *Southwestern Naturalist*, *51*(3), 338–345.

Tyers, M. (2017). Riverdist: River Network Distance Computation and Applications. https://cran.r-project.org/package=riverdist

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

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

Wang, I.J., & Bradburd, G.S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649-5662.

Warren, M. L., Burr, B. M., Walsh, S. J., Bart, H. L., Cashner, R. C., Etnier, D. A., . . . & Starnes, W. C. (2000). Diversity, Distribution, and Conservation Status of the Native Freshwater Fishes of the Southern United States. *Fisheries*, *25*(10), 7–31.

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

White, S.L., Hanks, E.M., & Wagner, T. (2020). A novel quantitative framework for riverscape genetics. *Ecological Applications*, *30*(7), 1–17.

Willing, E.M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by fst do not necessarily require large sample sizes when using many snp markers. *PLoS ONE*, *7*(8), e42649.

Winter, D.J. (2012). MMOD: An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, *12*(6), 1158–1160.

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

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

Zbinden, Z.D., Geheber, A.D., Lehrter, R.J., & Matthews, W.J. (2022). Multifaceted assessment of stream fish alpha and beta diversity using spatial models. *Hydrobiologia*, *849*(8), 1795–1820.

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

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

TABLES AND FIGURES

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

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

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

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

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

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

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

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

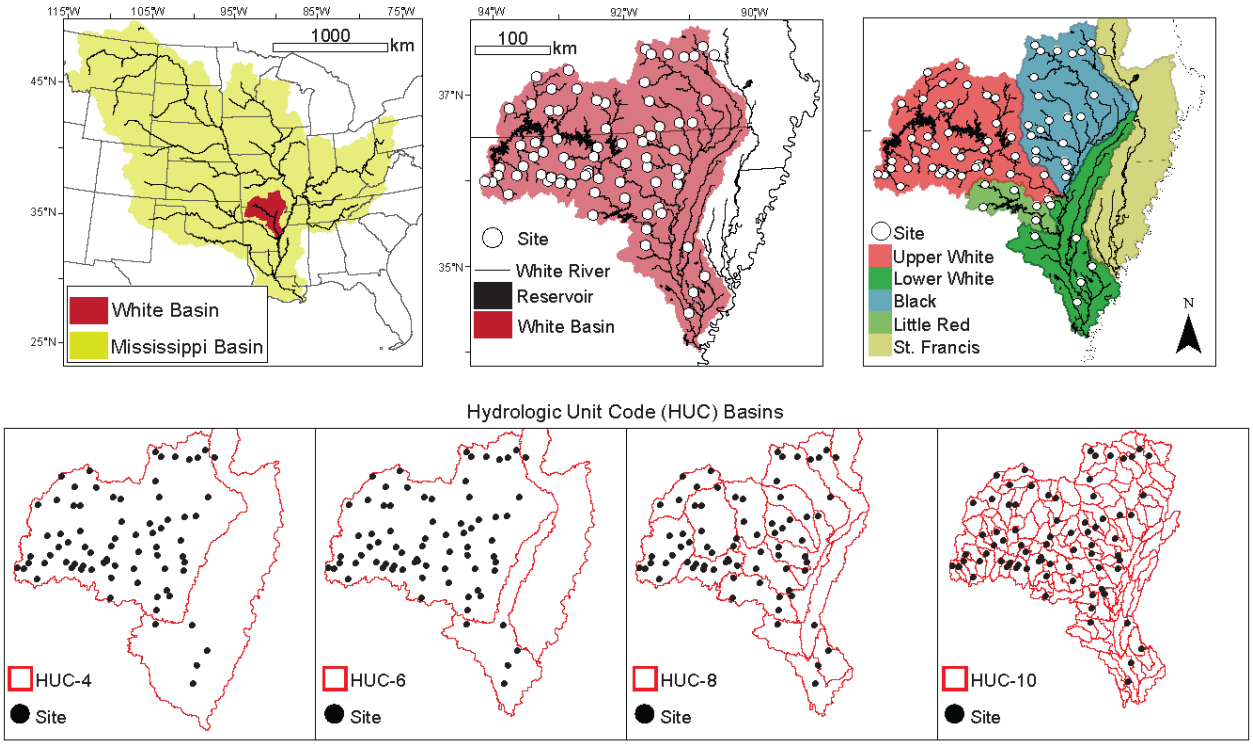


Figure 1: Fish were sampled at N=75 locations across the White River Basin (Ozark Mountains, U.S.A.). The study basin is contained within the larger Mississippi River Basin, and is a direct tributary to the mainstem Mississippi. The study region is subdivided into five subbasins: Upper White, Lower White,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

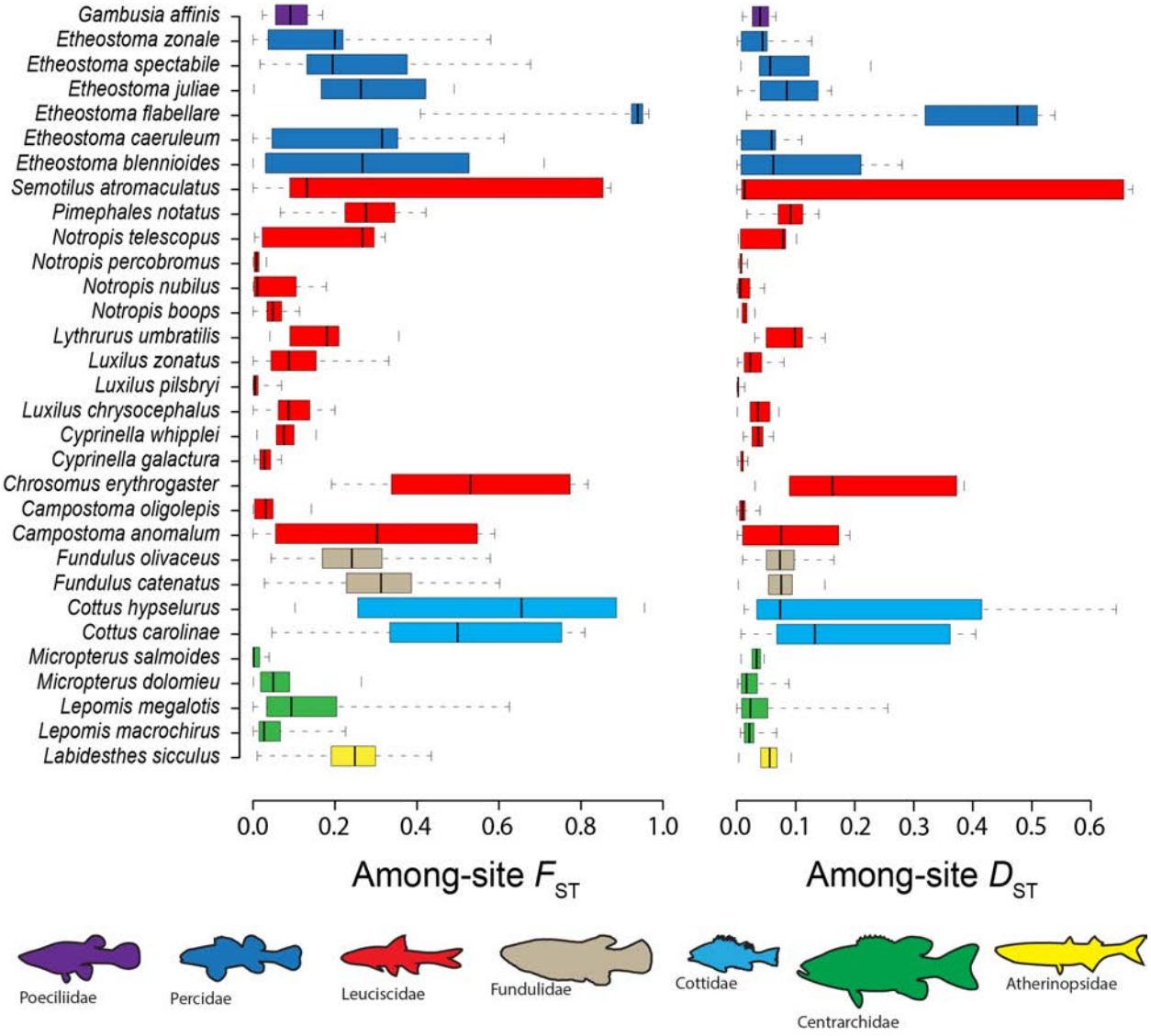


Figure 2: Genetic structure of *N*=31 fish species collected across the White River Basin (Ozark Mountains, U.S.A.) as summarized by among-site F_{ST} (Weir and Cockerham's θ) and Jost's *D*. Boxplots show the distributions of both pairwise estimates among sampling sites for each species. Inner quantiles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

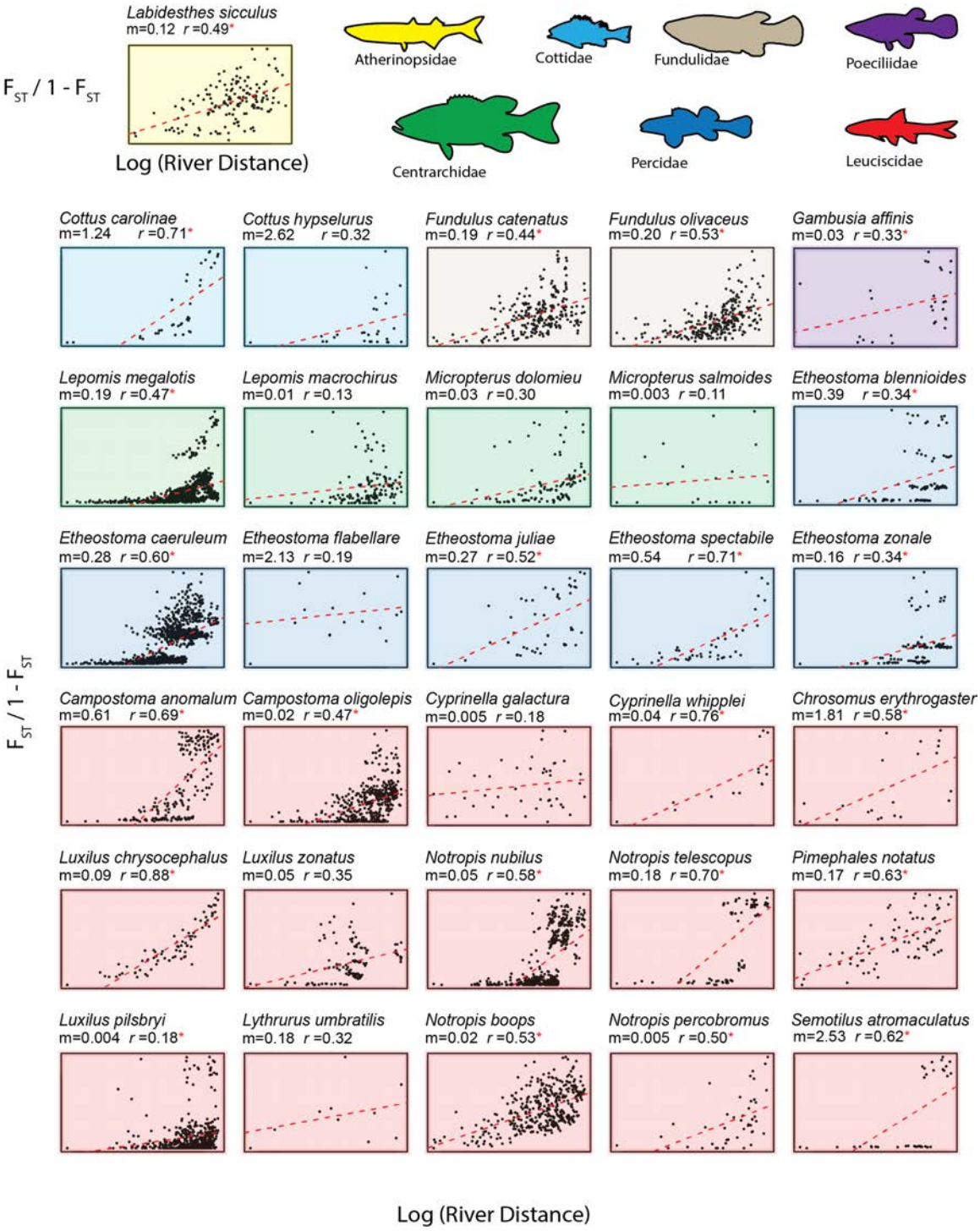


Figure 3: Isolation by distance plots for N=31 fish species collected across the White River Basin (Ozark Mountains, U.S.A.). Each depicts the relationship between among-site F_{ST} (linearized) and log river distance among sites. The following are represented below each species name: m=slope of the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

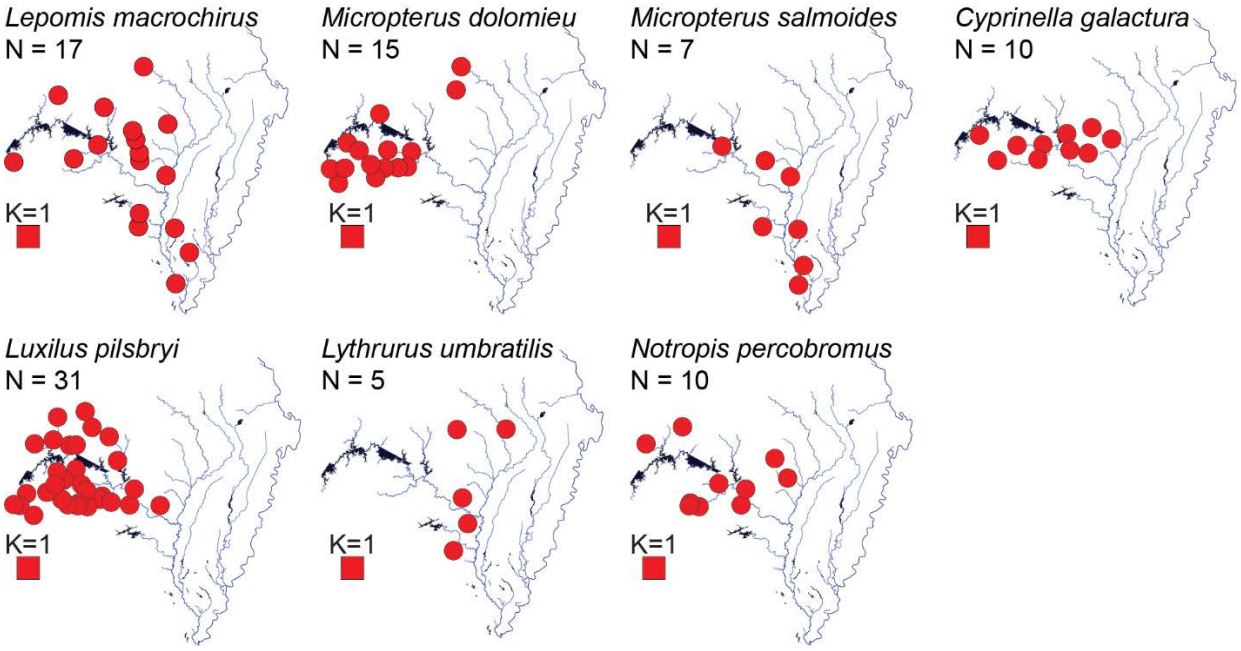


Figure 4: Sampling distribution maps of seven species which showed no evidence of discrete genetic population structure within the White River Basin (Ozark Mountains, U.S.A.). A total of *N*=31 species were sampled across 75 sites. The number of collection sites (red circles) for each species is denoted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

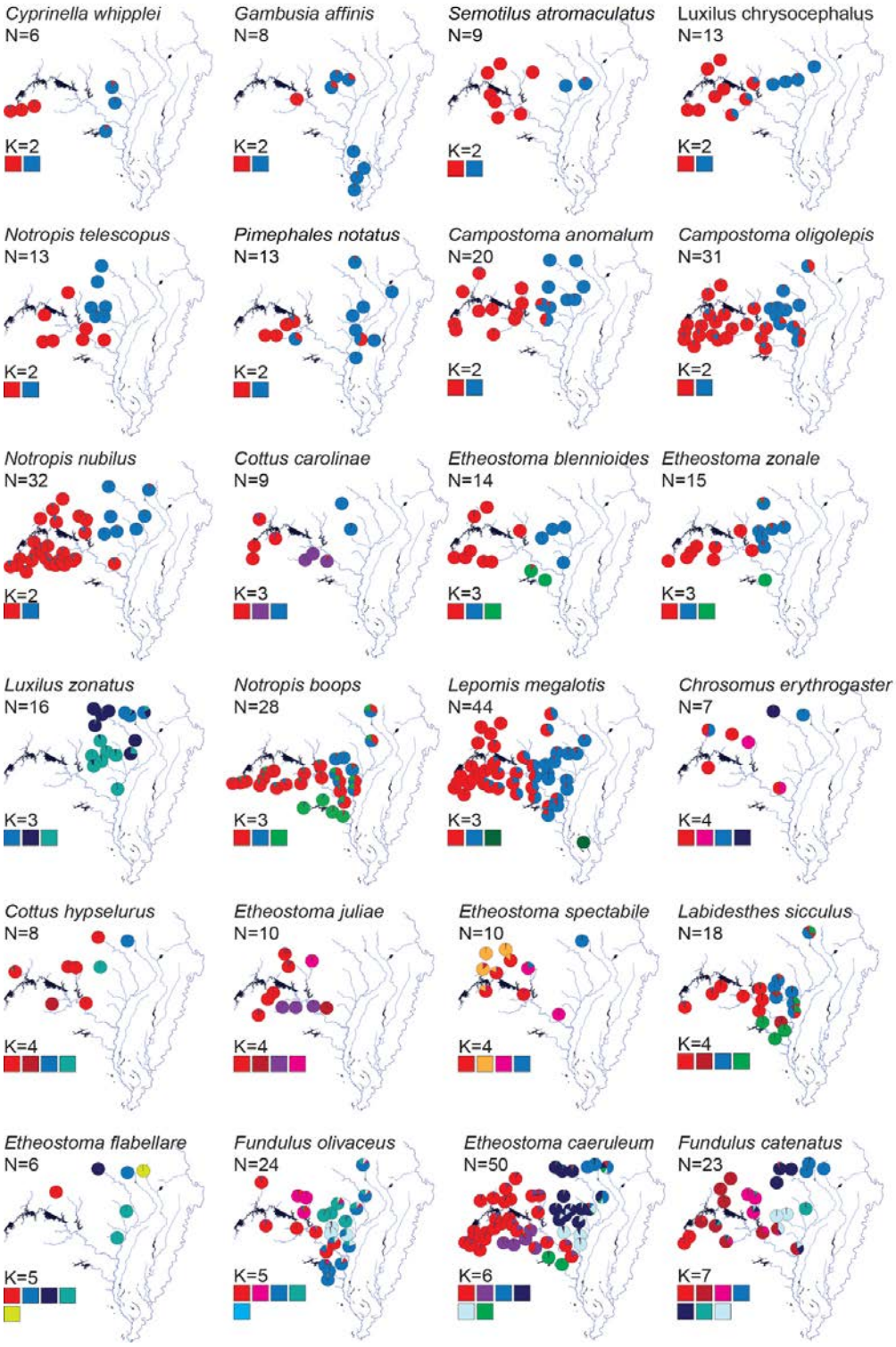


Figure 5: Sampling distribution maps of 24 species which showed evidence of genetic population structure within the White River Basin (Ozark Mountains, U.S.A.). *N*=31 species were sampled across 75 sites. K= the number of discrete genetic populations discerned from sparse non-negative matrix factorization.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

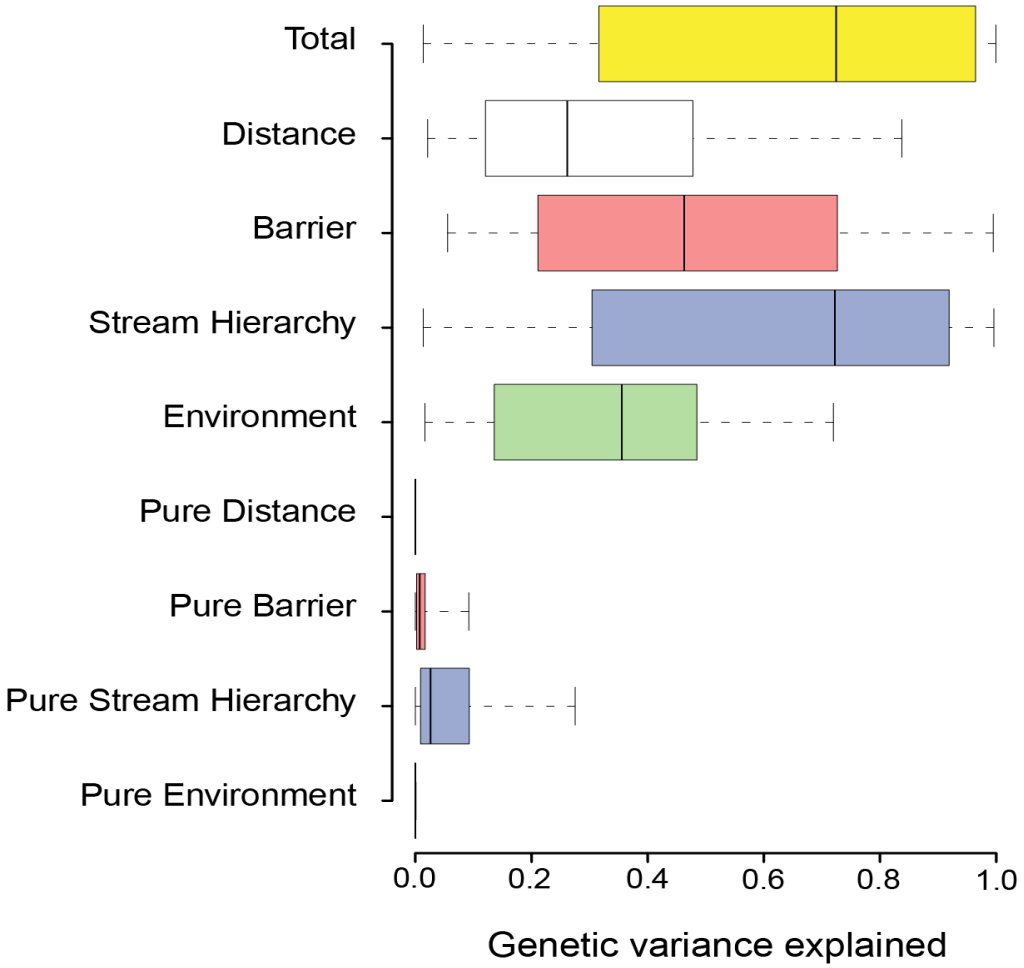


Figure 6: Neutral genetic variation was partitioned between four explanatory models for *N*=31 fish species sampled across the White River Basin (Ozark Mountains, U.S.A.). Partitioning was conducted separately for each species. The four models represent: (i) isolation by *distance*, the river network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

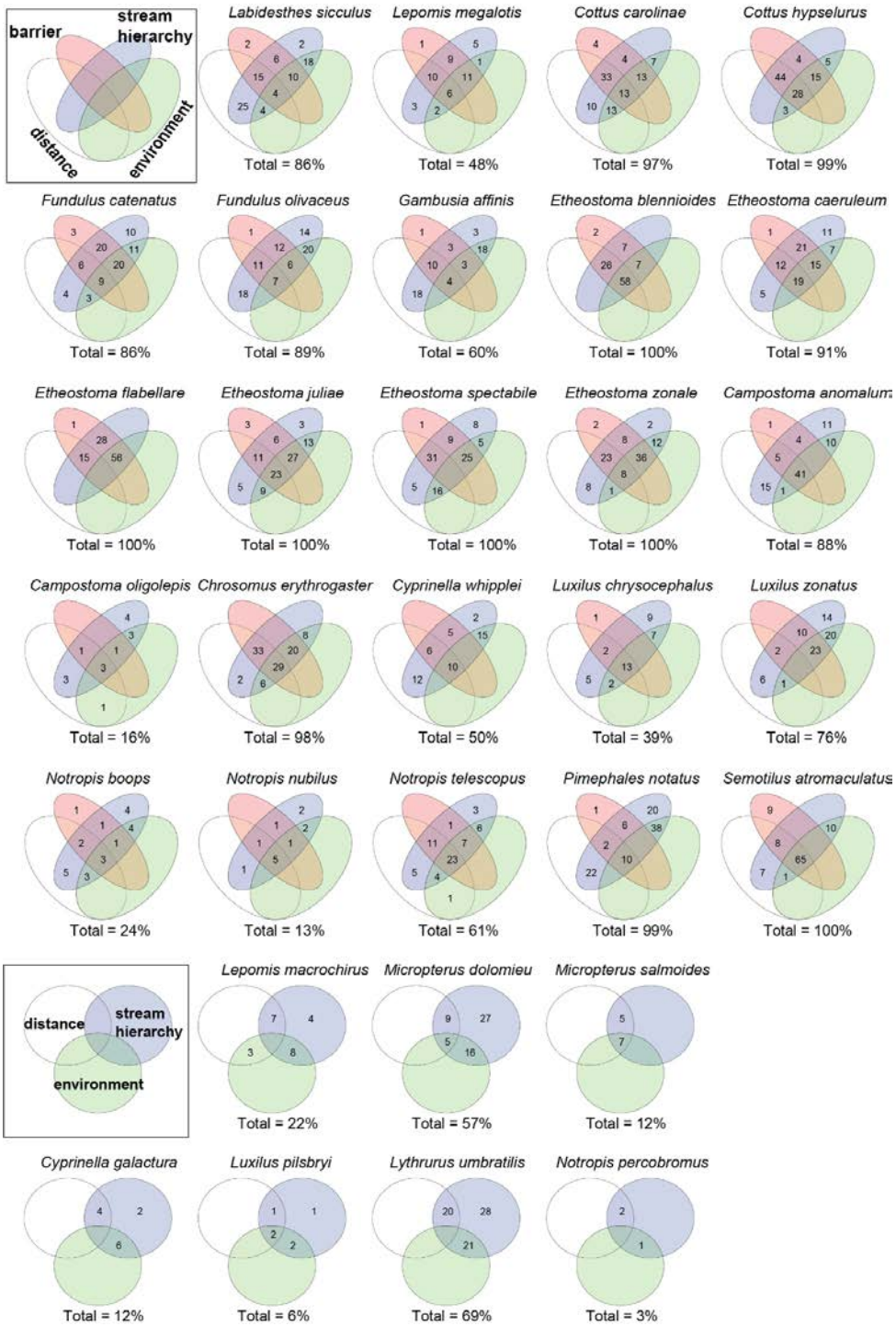


Figure 7: Venn diagrams depict neutral genetic variation resulting from four models as applied to *N*=31 fish species sampled from the White River Basin (Ozark Mountains, U.S.A.). Models were based on: (i) isolation by distance, isolation by barrier, isolation by stream hierarchy, and isolation by environment.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

CHAPTER III

Linking traits with evolutionary drivers within freshwater fish communities of the Ozark Highlands

Zachery D. Zbinden, Marlis R. Douglas, Tyler K. Chafin, Michael E. Douglas

ABSTRACT

Phenotypic traits modulate how species interact with one another and the environment and consequently shape population histories. These species-specific responses to a shifting environment are documented as distinct genetic patterns that emerge over time. While this linkage between genetic diversity and phenotypic traits is seemingly fundamental, its causality has been difficult to establish unequivocally. We quantify phenotypic traits across a freshwater fish community to test if they are significantly related to observed genetic patterns. Traits so identified were then parsed using a predictive model that generated genetic diversity indices. The test matrix for our application was a freshwater fish community (N=31 species) sampled across 75 sites within the White River Basin (Ozark Plateau, USA). For each species, we derived three genetic diversity indices (=H_S/H_T/G"$_{ST}$) using SNP analysis (N=2,000 loci) and assessed 28 phenotypic traits. After adjusting for phylogenetic autocorrelation and removing sample-size effects, we identified a series of traits (N=2-5, depending upon the index) strongly associated with diversity. These were subsequently applied in predictive models that explained 31-68% of the genetic variability in each species. Our approach effectively linked species-specific traits with diversity both within/among populations. Researchers can apply it to generate diversity metrics for species currently lacking such information, which can help prioritize freshwater fish management and streamline conservation plans.

INTRODUCTION

How species interact with their abiotic and biotic environment ultimately determines the evolutionary trajectory of populations (Hand et al., 2015; Hutchinson, 1965). Linking ecology and evolution remains a central challenge in biology (Avise et al., 2016). One route to link them is through the relationship between species traits and genetic diversity (Bohonak, 1999). Phenotypic characteristics — or traits — mediate how species deal with abiotic and biotic factors. Therefore, traits also play a role in the evolution of populations and should show strong relationships with different evolutionary metrics (i.e., genetic diversity; Meirmans et al., 2011). We use 'genetic diversity' to refer broadly to three different facets: within-site diversity (α); among-site diversity (β); and total diversity (γ) (Sherwin et al., 2017). Parsing the relationship between traits and genetic diversity to uncover generalities requires a comparative approach (López-Uribe et al., 2019; Selkoe et al., 2014).

Here, the most insight to be gained is at the community level, e.g., landscape community genetics (Hand et al., 2015). By community-level, we mean simultaneously assessing the genetic population structure of multiple, co-distributed species within a biogeographic region or metacommunity (Leibold et al., 2004; Rissler, 2016). At this level, the study of ecology and evolution is enhanced by quantifiable variation among species that can be leveraged to test hypotheses (Kelly & Palumbi, 2010). Why do some species have low genetic diversity? A partial answer, albeit not terribly useful, is undoubtedly stochastic and historical contingency, whereas a more direct conduit may be the organismal characteristics that drive such variability (Duminil et al., 2007). With a nod to Dobzhanksy (1973): Nothing in comparative community genetics makes sense except in the light of traits (Papadopoulou & Knowles, 2016). Hence, we can uncover why species differ regarding genetic diversity by employing a trait-based approach.

Such a predictive framework relies on a causal link between traits shaping ecology and evolutionary history, as demonstrated by statistical associations between genetic diversity and the traits in question, which emerge from demographic history and population connectivity. In this sense, life history influences demography and population size, while dispersal initiates connectivity. These processes, accumulated over time, impact genetic diversity through various effects on evolutionary processes. For example, smaller populations rapidly lose genetic variation to drift (Lynch & Lande, 1998), while those with limited connectivity diverge more rapidly in a genetic sense (Wright, 1931). Thus, the link between traits and genetic diversity is substantial and has been empirically demonstrated in animals (Bohonak, 1999), plants (Duminil et al., 2007; Loveless & Hamrick, 1984; Meirmans et al., 2011), marine organisms (Kelly & Palumbi, 2010; Riginos et al., 2014), and birds (Burney & Brumfield, 2009).

If life history and dispersal traits are associated in a causal sense with genetic diversity, then diversity could be predicted in other species by modeling their traits; this could be valuable because genetic diversity represents the variation available for evolution to act upon and serves as an indicator of population persistence (Jump et al., 2009). The spatial structure of genetic diversity further imposes constraints on a species' ability to adapt to a changing environment (López-Uribe et al., 2019). Many efforts have aimed to assess the genetic diversity of threatened species, although over 70% have yet to be assessed (Bachman et al., 2019; Hogg et al., 2022), underscoring the value of a predictive framework. Understanding levels of standing genetic variation and its structure is critical for managing declining species (Willoughby et al., 2015).

However, suppose we could forecast which species will become threatened in the future based on genetic diversity indicators – as a proxy for 'population persistence.' We could then take proactive measures to bolster such species against decline (Lunney et al., 2004; Martinez,

Willoughby, & Christie, 2018). These predictions can guide efforts to prioritize species of conservation concern and focus targeted efforts aimed at further data collection. Rather than conducting population genetic studies on hundreds or thousands of species, which would be costly and require technical skill and necessary infrastructure, we could make accurate predictions of critical conservation indicators by assessing a few dozen species representative of a regional community.

The primary focus of this study was to test the hypothesis that species-specific traits within a freshwater fish community are related to genetic diversity. We hypothesized that essential traits would be related to morphology, life history, and ecology, as these are most likely to impact population size, reproduction, and dispersal. Little work has been done to assess the relationship between traits and genetic diversity for freshwater fishes (but see Martinez et al., 2018; Mitton & Lewis, 1989; Pilger et al., 2017; Sousa-Santos et al., 2016). Secondly, we were interested in which traits are most strongly associated with diversity and what inferences, i.e., mechanisms promoting genetic diversity, could be established from these relationships. Finally, we attempted to build predictive models of different genetic diversity indices based on parsimonious sets of traits.

METHODS

The study region

The focal region for our study is the Ozark Plateau (Figure 1), specifically the White River Basin, which drains 71,911 km^2 as a tributary to the Mississippi River. Like the Appalachian Highlands, the Ozark Highlands served as a glacial refugium during the Pleistocene (Mayden, 1988). Its long temporal span of geologic stasis has acted to equilibrate gene flow and genetic drift, thus serving as a 'natural laboratory' for biodiversity diversification (Hutchison &

Templeton, 1999). It contrasts with more northern latitudes, where cyclical bottlenecks and subsequent expansions have effectively (and iteratively) recast population genetic signatures of resident taxa (Cammen et al., 2018). The extended stability of the Ozark Highlands system provides a platform upon which the evolutionary histories of resident biodiversity have been annealed by trait-mediated ecology rather than by substantial geomorphic disturbances that serve to 'reshuffle' the genetic deck, effectively masking phylogeographic patterns.

Sampling

Fishes were collected in wadable streams using seines during low flows between June 2017 and September 2018. Sampling was approved by the University of Arkansas Institutional Animal Care and Use Committee (IACUC: #17077), with collecting permits from Arkansas Game & Fish Commission (#020120191), Missouri Department of Wildlife Conservation (#18136), and U.S. National Parks Service (NPS: Buffalo River Permit; BUFF-2017-SCI-0013). Fish were euthanized by immersion in tricaine methanesulfonate (MS-222) at a concentration of 500 mg/L, buffered to pH=7, and then preserved in 95% ethanol. Species diagnosis occurred in the laboratory, as augmented by Eschmeyer's Catalog of Fishes (Fricke et al., 2022). The right pectoral fin was removed and stored in 95% ethanol at -20 °C prior to subsequent DNA extraction.

Genomic data collection and filtering

To estimate genetic diversity indices, we developed genotypic alignments for each species by using a double-digest restriction site-associated DNA (ddRAD) sequencing procedure (Peterson et al., 2012) that provided tens of thousands of loci for each individual. Our standard procedures

were appropriately modified (Chafin et al., 2019) and previously reported in detail (Zbinden et al., 2022a, Zbinden et al., 2022b).

Genomic DNA was first isolated and then digested with high-fidelity restriction enzymes *Msp*I and *Pst*I. Bead-purified samples were standardized to 100 ng DNA and then ligated with custom adapters containing in-line identifying barcodes. Samples were pooled in library sets of 48 and subsequently size-selected from 326–426 bp, including adapter length. Illumina adapters and i7 index were added via 12-cycle PCR with Phusion high-fidelity DNA polymerase. Three libraries were pooled per lane and then single-end sequenced on an Illumina HiSeq 4000 platform (1x100bp; Genomics & Cell Characterization Core Facility; University of Oregon, Eugene).

Raw Illumina reads were demultiplexed, clustered, filtered, and aligned with IPYRAD v.0.9.62 (Eaton & Overcast, 2020). Demultiplexing criteria allowed no more than a single barcode mismatch and individuals with extremely low reads ($< \bar{x} - 2s$) were removed. Next, individuals were screened for misidentifications and putative hybrids (Zbinden et al., 2022a), the former re-assigned and the latter removed. Raw sequence reads were partitioned by species and aligned *de novo.* Clusters of homologous loci were assembled using an 85% identity threshold. Putative homologs were removed according to the following criteria: <20x and >500x coverage per individual; >5% of consensus nucleotides ambiguous; >20% of nucleotides polymorphic; >8 indels present; or presence in <15% of individuals. Paralogs were identified (and removed) as those clusters with either >2 alleles per site in consensus sequence or excessive heterozygosity (>5% of consensus bases or >50% heterozygosity/site among individuals).

Biallelic SNP panels were then visualized for each species, with additional filtering implemented with the R package RADIATOR (Gosselin, 2020). To ensure high data quality, loci

were removed if: Monomorphic; minor allele frequency <3%; Mean coverage <20 or >200; Missing data >30%; SNP position on read >91; and significant HWE deviation in one or more sampling sites ($\alpha = 0.0001$). Linkage disequilibrium was reduced by retaining but one SNP per locus (i.e., that which maximized minor allele count). Finally, singleton individuals/species at a sampling site were removed, as were those with >75% missing data in the filtered panel. We standardized panels to eliminate bias caused by differences in the number of loci between species by randomly sampling 2000 SNPs per species panel. However, genetic diversity indices were almost perfectly correlated ($r > 0.99$) whether generated from full *versus* standardized panels.

Genetic diversity

We calculated three facets of genetic diversity for each species: Within-site diversity (α), among-site diversity (β), and total diversity within and among sites (γ) (Sherwin et al., 2017). Three indices were employed to represent these facets: H_S, the average gene diversity (heterozygosity) within sites (Nei, 1973; Nei & Chesser, 1983); $G''_{ST,}$ the unbiased fixation index (Meirmans & Hedrick, 2011); and $H_{T,}$ total gene diversity across all sites (Nei, 1973; Nei & Chesser, 1983). We chose G''_{ST} because it is a more appropriate metric for comparison among species, and it does not underestimate differences when the number of sampled populations is small (Meirmans & Hedrick, 2011). These genetic diversity indices were calculated using R Statistical Software (R Core Team, 2022) via SNP panels formatted as 'genind' objects (Jombart, 2008), with calculations performed by the R package MMOD (Winter, 2012).

Explanatory factors

Covariates

Although our goal was to relate species traits with genetic diversity, we surmised that differences in sampling among species, e.g., number of sites, could also increase variability. Therefore, we included four covariates for each species in our exploratory analysis, to include: the number of individuals in the analysis (=Nindiv), the number of sites at which those individuals occurred (=Nsites), the mean number of individuals analyzed per occupied site (=Mean_Ind_Site), and the median value of pairwise river distances between all occupied sites (=Median_RivDist).

Phylogeny

Closely related species tend to be more similar, which can violate assumptions of independence in our models, i.e., phylogenetic autocorrelation (Felsenstein, 1985). However, removing autocorrelation entirely from analyses can eliminate other (ancillary) signals of potential interest (Legendre, 1993; Peres-Neto, 2006). Therefore, we explored phylogenetic autocorrelation using variation partitioning (Borcard et al., 1992), which allowed us to test the effects of traits with and without phylogenetic autocorrelation and determine the potential role of phylogenetically correlated traits.

A phylogeny of the study species was obtained using a previously constructed tree of ray-finned fishes based on a 27-gene multi-locus alignment (Rabosky et al., 2018). Our backbone tree was limited to the species collected in this study (via the R package FISHTREE; Chang et al., 2019). We decomposed the topology using phylogenetic eigenvector regression (PVR; Diniz-Filho et al., 1998) to create a matrix of *N*-1 eigenvectors for a given phylogeny (where *N*=number of species). Principal coordinate analysis decomposed phylogenetic distances among

tips into vectors representing the phylogenetic topology, with the first (e.g., EV1) representing deeper splits and more recent splits by the latter (e.g., EV20) (Diniz-Filho et al., 2012). We tested the relationships between these eigenvectors and genetic diversity indices using the approaches described below.

Traits

We assembled a data set of 28 species-level traits broadly related to morphology, life history, and ecology. Phenotypic characteristics within each category can leave an indelible imprint on genetic diversity by impacting dispersal and population demography and, in turn, gene flow and genetic drift. Traits were gathered from three primary databases: FishTraits, a public database for North American freshwater fishes (Frimpong & Angermeier, 2009); FISHMORPH, a global database on morphological traits of freshwater fishes; and an unpublished database (J.D. Olden, unpublished data, 2021) used in previous publications (Giam & Olden, 2016; Mims et al., 2010). Robison & Buchanan (2020) was consulted to corroborate/adjudicate disagreements [Supplementary Material (S1) and online data repository (https://osf.io/837vj/)].

Analyses

Data reduction

We developed three response variables (i.e., indices of genetic diversity) that we hypothesize as related to our predictor variables: (a) Covariates; (b) phylogenetic eigenvectors; and (c) species-level traits. We analyzed each response variable separately using an identical framework (Figure 2) as follows: We first applied vector fitting (R package VEGAN; Oksanen et al., 2020) to test each predictor set against a diversity index. Those variables exhibiting a significant relationship

($\alpha < 0.05$) were retained for further analysis. We did not adjust α for multiple comparisons in that we did not wish to increase the potential for Type-II error, given a subsequent data reduction step. Significant variables within each predictor set were then used in separate stepwise forward selection procedures to yield parsimonious models based on Akaike Information Criteria (AIC) (Oksanen et al., 2020). If covariates explained significant variation in a genetic index (only the case for H_T), those effects were removed using residuals extracted from a generalized linear model (GLM) between the genetic index and significant covariates.

Variance partitioning

We estimated the variation in genetic diversity explained by each predictor set after accounting for variation explained by the other sets. We partitioned the variation among species across reduced sets of predictor variables for each genetic diversity index to test for significant relationships. Partial multiple regressions (variation partitioning; Borcard et al., 1992; Legendre & Legendre, 2012; Peres-Neto et al., 2006) allowed us to elucidate the variation of each genetic index explained by traits, phylogeny, and phylogenetically correlated traits. Covariates were not included in that they were either not significantly associated with genetic diversity (H_S, $G"_{ST}$), or their significant effects had previously been removed per above (H_T). Results were presented as Venn diagrams depicting the adjusted R^2, i.e., the extent of variation explained by predictors. We tested our fractions of variation, e.g., "pure trait" variation after adjustment for phylogeny, using N=9,999 permutations (Anderson & Legendre, 1999).

Multi-model averaging

Reduced trait sets were used as 'global' models to generate all possible combinations, i.e., subsets, with the different models ranked according to second-order Akaike Information Criteria (AIC_C) (R package MuMIn; Barton, 2022). We derived coefficients for each trait and their standard errors based on model averaging via information criteria to produce predictions based on trait inputs. As a demonstration, we used the models to predict genetic diversity for Slender Madtom (*Noturus exilis* Nelson, 1876), a small-bodied benthic-dwelling catfish in the White River Basin (as an aside, it was not sampled extensively enough to be included in formal analyses; i.e., ≥5sites with ≥2 individuals). We used trait values gathered from the abovementioned sources and model coefficients to predict its genetic diversity values. Slender Madtom SNP genotypes were used to estimate H_S, H_T, and G''_{ST} (generated in the same manner as for the other species and standardized to N=2000 SNPs). These estimates allowed us to compare and contrast predicted *versus* observed genetic diversity indices.

RESULTS

We collected freshwater fish from N = 75 locations (Figure 1) and analyzed N = 2,861 individuals representing 31 fish species, as genotyped across standardized SNP panels generated with ddRAD. Each panel was based on 15–358 individuals collected across 5–50 sampling sites (Table 1). The mean number of individuals/species/site = 5.1. Mean within-site gene diversity (=H_S) ranged from 0.024–0.283 ($\bar{x}$=0.159; s=0.056); total gene diversity (=H_T) ranged from 0.138–0.350 ($\bar{x}$=0.221; s=0.053); and among-site diversity (=G''_{ST}) spanned from 0.022–0.965 ($\bar{x}$=0.303; s=0.253) [Supplementary Material (S2–S3)].

We gathered 28 traits related to morphology, life history, and ecology [Table 2; Supplementary Material (S1 & S4)]. We examined relationships between genetic diversity and four covariates (N individuals, N sites, individuals/site, and median river distance between sites). None of the covariates were significantly related to either within-site gene diversity (H_S) or among-site diversity (G''_{ST}). However, total gene diversity (H_T) was significantly related to N individuals, N sites, and Mean individuals/site, i.e., sample size (Table 3). These three covariates explained ~62% of the variation in H_T, and forward-selection revealed that N individuals alone could explain ~50%. Surprisingly, total diversity decreased with sample size. The remaining results related to total gene diversity, H_T, are therefore based on the residuals of a linear model (H_T ~ N individuals) to remove the effect of sampling variability.

Each facet of genetic diversity was significantly related to some component of the traits: H_S = 8 traits; H_T = 4; G''_{ST} = 6 (Table 3). These reduced sets explained 26–66% of the variation within diversity (Table 3). Following forward-selection, a parsimonious trait model was derived for each index, and involved fewer traits: H_S = 5 traits; H_T = 2; G''_{ST} = 2 (Table 3). These explained similar amounts of variation (31–68%), as did the vector-fitted sets (based on adjusted R^2) (Table 3).

Each facet of genetic diversity was also significantly related to phylogenetic distance (Table 3). Only one or two phylogenetic eigenvectors were significantly related to each genetic diversity index but explained between 15–36% of the variance. In each case, those phylogenetic eigenvectors that were significant in the vector fitting reduction step were also selected in forward-selection (Figure 2; Table 3). Relationships between each genetic diversity index and its corresponding vector-fitted variables are found in Supplementary Material (S5–S7).

After accounting for phylogenetic autocorrelation, significant variance in genetic diversity was explained by partial multiple regression among traits (Figure 3). Total variation (traits and phylogeny) ranged from 43–68%, whereas variance explained "purely" by traits ranged from 6–52%. Phylogeny explained from 0–18% of the variation. Finally, phylogenetically correlated traits explained 16–25%.

We confirmed our parsimonious models' suitability for each genetic diversity index by testing all possible model combinations. The full model constructed using forward-selection was either the second-best (H_S, Table 4) or best model (H_T, Table 5; & G''_{ST}, Table 6). For H_S, the difference in AIC_C between the best and full models was essentially meaningless (delta=0.20; Table 4).

The averaged trait coefficients were utilized to predict genetic diversity indices for Slender Madtom. Predicted H_S= 0.10 with 95% confidence interval=0.07–0.13 (observed H_S =0.16). Predicted G''_{ST}=0.56 with 95% confidence interval = 0.36–0.77 (observed G''_{ST}=0.64). Unfortunately, H_T predictions are not comparable with observed values because the *residuals* of the linear relationship between H_T and N individuals were modeled rather than raw values.

DISCUSSION

Quantifying the relationship between traits and genetic diversity is one way to bridge the gap between ecology and evolution (Bohonak, 1999). Theoretically, those traits that influence how genes are passed from one generation to the next, e.g., fecundity, or how genes are spread among populations, e.g., swimming ability, should impact genetic diversity (Papadopoulou & Knowles, 2016). Whether genetic diversity can be predicted from species traits alone is still an open question. A relationship between genetic diversity and variance among species traits must first be

established. Herein, we tested for such a relationship by gathering data on traits associated with individual movement and population size. We found a statistically significant signal after first accounting for phylogenetic relatedness. For inference purposes, we then reduced them to sets that statistically correlated with three genetic diversity indices. We then selected parsimonious sets of these traits to build predictive models of genetic diversity.

Previous studies tested the relationships between traits and genetic diversity and, in so doing, supported the link between ecology and evolution (freshwater fishes: Martinez et al., 2018; Mitton & Lewis, 1989; Sousa-Santos et al., 2016; other biodiversity elements: Bohonak, 1999; Duminil et al., 2007; Kelly & Palumbi, 2010). However, trait diagnosis and predictability remain inconsistent (Riginos et al., 2014), a conclusion not surprisingly derived from hundreds of different population-level studies across many different biogeographical regions (but see: Meirmans et al., 2011; Selkoe et al., 2014; Sousa-Santos et al., 2016). The search for targeted associations became arduous and muddled as numerous confounding factors were incorporated (Kelly & Palumbi, 2010). Additionally, meta-analyses are often forced to incorporate legacy markers with lower information content, e.g., isozymes, AFLPs, or microsatellites. A community-level approach that genotypes individuals and estimates genetic diversity using next-generation sequencing technology might allow for a potentially more robust link between traits and diversity and reduce inconsistencies among projected studies.

Key traits

We found a significant signal of association between at least one genetic diversity index and half of the traits so analyzed, thus supporting a hypothesized relationship between traits and genetic diversity. Key morphological traits represented aspects of body size and swimming/foraging

behavior: body elongation, lateral shape, pectoral fin position, caudal fin throttling, and position of the mouth. Important life-history traits included fecundity, egg size, length of spawning season, parental energy investment in offspring, and whether eggs were guarded post-spawn. The significant ecological traits were related to habitat use: Temperature preference, benthic feeding, and surface feeding.

Within-site genetic diversity (H_S)

Species with higher within-site genetic diversity tended to be larger, with greater fecundity, smaller eggs, longer spawning seasons, preferences for warmer water temperature, with mouths phenotypically inclined toward the surface rather than the benthos (*vice-versa* for species with low diversity). The above relationships between genetic diversity and body size, egg size, environmental tolerance, and benthic *versus* pelagic spawning have been previously noted (Husemann et al., 2012; Osborne et al., 2014; Sousa-Santos et al., 2016).

Larger-bodied fish that feed more toward the surface and prefer warmer temperatures may occupy much larger areas than fish lacking such characteristics, i.e., dispersal syndromes (Comte & Olden, 2018). This may consequently increase both population sizes and connectivities, positively influencing genetic diversity within sites. Genetic diversity erodes more slowly in larger populations (Mitton & Lewis, 1989).

Population size and genetic diversity may also benefit from higher fecundity, smaller egg sizes, and extended spawning periods. Egg dispersal may also promote lotic genetic diversity through population connectivity, whereas those that produce more and smaller eggs may benefit from stream-mediated dispersal (Platania & Altenbach, 1998). The genetic diversity of fishes in lotic environments as it relates to egg dispersal is not well studied. However, it is consistent for

marine fishes in that smaller eggs released into the water column disperse widely due to ocean currents (Riginos et al., 2014). Moreover, a more extended spawning season may provide greater environmental variability and, by chance or cue, provide optimal conditions for downstream dispersal and/or survival of eggs and larvae. For example, more turbulent flows during broadcast spawning may enhance viability by preventing egg clumping (Jager et al., 2001). Faster flows may also reduce competition by broadcasting embryos more widely and providing a more extensive range of ecological opportunities (McCabe Jr. & Tracy, 1994).

Total genetic diversity (H_T)

Species with higher total genetic diversity tend to be deeper bodied, invest more energy into parental care, and construct/ guard nests. This finding underscores the role of parental investment in determining overall genetic diversity. However, this finding stands in opposition to the suggestion that '*r*-strategists' (high fecundity/low parental investment) tend to have larger population sizes and concomitantly higher genetic diversity than '*K*-strategists' (low fecundity/ high parental investment) (Mitton & Lewis, 1989; Romiguier et al., 2014). Although, those findings align more with our observation of within-site diversity above. Interestingly, a large meta-analysis of marine and freshwater fishes (N=463) also provided mixed support for the polarity of *r* vs. *K*-driven genetic diversity (Martinez et al., 2018). Romiguier et al. (2014) assessed 76 metazoan species across a diversity of evolutionary histories and observed an inverse relationship between propagule size (quality) and fecundity (quantity). Yet, we found fecundity positively correlated with higher parental investment and nest construction/guarding behavior. In our study species, those investing more in offspring also deposited more eggs – although smaller in size. Presumably, species exhibiting these characteristics also had greater numbers of

offspring surviving to reproduce, thus promoting larger, more stable populations and higher genetic diversity. Ultimately, categorizing species as *r*- or *K*-strategists based on only a few traits may lead to spurious conclusions.

Among-site genetic diversity (G''_{ST})

Generally, species with higher among-site diversity tend to have mouths more ventrally positioned, trophically more benthic in orientation, and pectoral fins more dorsally positioned (for active swimming). Caudal fins were smaller (less influential in propulsion), whereas eggs were larger-sized (morphological interpretations from Brosse et al., 2021). Species with higher among-site diversity also seemingly displayed reduced within-site diversity. Indeed, these characteristics seem opposed to those promoting higher within-site diversity (as above). Species in the latter category are more benthic-oriented (per mouth and pectoral fin positions), with trophic resources gleaned from the bottoms of streams/rivers. Hence, they display less movement than heterospecifics trophically oriented within the water column or the surface. Moreover, their larger eggs are perhaps less likely to be dispersed downstream. These results are consistent with the expectation that benthic habitat specialists demonstrate greater divergences among populations than habitat generalists (Pilger et al., 2017).

Can genetic diversity be predicted?

The processes that have shaped global biodiversity must be clearly understood before attempting to mitigate its loss (Manel et al., 2020). In this regard, threatened taxa lose heterozygosity more rapidly due to genetic drift acting on declining populations. Concomitantly, their genetic diversities should be depressed (Spielman et al., 2004), which imparts a substantial, adverse

effect on fitness (DeWoody et al., 2021). If managers could identify which species tend to have lower levels of genetic diversity, they could prioritize them for targeted surveys and subsequent management. The capacity to forecast which species have lower genetic diversity based on easily estimated traits would provide much-needed focus, as there are far too many species for each to be evaluated for population genetic metrics.

While a significant relationship between traits and genetic diversity is of interest, it may be of scant applicability if traits lack predictive power. For example, deterministic trait-based models are of little help if genetic diversity variation among species is indeed stochastic. However, previous studies suggest that a trait-based framework may be valuable. Meirmans et al. (2011) found that ecological and life-history traits explained 30% of the variation in the genetic structure of alpine plant species, a remarkably high value given it was based on but six characters, with other (non-assayed) processes also influential. Other studies across a diverse group of animals identified even higher correlations (e.g., R^2=0.79) between genetic diversity and either life-history traits (Romiguier et al., 2014) or dispersal abilities (Bohonak, 1999). By comparison, our models of genetic diversity performed well with adj. R^2 ranging from 0.31 to 0.68. We acknowledge potential drawbacks in evaluating the predictive capacity of models solely based on coefficients of determination (Onyutha, 2020). In particular, our within-population genetic diversity model (H_S) provided the best predictive capacity (Std Err=0.016), whereas our among-population genetic diversity model (G''_{ST}) had a greater standard error (Std Err=0.11). However, both reasonably predicted H_S and G''_{ST} for the Slender Madtom.

Phylogenetic autocorrelation

Genetic diversity is not a heritable trait of a species *per se* but rather an emergent property (Duminil et al., 2007). We thus hypothesize that similarity in genetic diversity among related species is driven by species-specific traits that mediate vital components of genetic diversity, such as population sizes and connectivities (Abrams, 2019; Fobert et al., 2019; Naish et al., 2013). Since closely related species tend to be more similar with respect to traits (Felsenstein, 1985), the same might be expected for genetic diversity in that it emerges via trait combinations (Duminil et al., 2007). We thus suggest that species with similar traits should reflect similar genetic diversities.

To avoid overestimating significance due to autocorrelation, we considered the effects of shared evolutionary history in testing the overall relationship between genetic diversity and traits (Hawkins, 2012). However, we did not incorporate phylogeny in predictive modeling (Meirmans et al., 2011). Strictly speaking, our study species are not independent of one another, and our statistical application of linear models would violate the assumption of independence. However, we would also remove favorable aspects of statistical signal by forcing independence by factoring out variance due to shared ancestry (Duminil et al., 2007). We agree with others (Legendre, 1993; Peres-Neto, 2006) that a balanced approach to autocorrelation is necessary. In this sense, phylogenetic autocorrelation is interpreted not as bias or *artifact* but instead as what we are interested in (Hawkins, 2012; Legendre, 1993).

Importance of covariates

Given the differences encountered in sampling among species, associated covariates could also be related to variability in genetic diversity. For example, one might predict that an increase in

sample size might also concomitantly increase observed genetic diversity. While sample size differences or spatial variance in a sampling regime can be controlled, other more nebulous covariates such as biogeographic and/or climactic histories can be challenging to incorporate. They conspire to make the meta-analyses of genetic diversity less tractable across numerous independent studies. In contrast, single studies of multiple co-distributed species within the same biogeographic region provide more opportunities for sound interpretation by eliminating metadata variance.

Only our estimates of total genetic diversity, H_T, had related covariates. We observed a decrease in total genetic diversity for a species as we sampled more individuals, sites, or individuals/sites. This result conflicts with potential predictions because more variation should be inherent with an increasing sample size (McCusker & Bentzen, 2010). However, two possible explanations emerge: First, heterozygosity is most influenced by common alleles and less by rare ones (Schmidt et al., 2021). The former requires relatively few individuals and sites to be manifested, whereas additional sampling adds only rare variants without substantially increasing heterozygosity. Second, lower heterozygosity should emerge if our sample of a species' genetic diversity contains observed population structure due to the Wahlund effect (Wahlund, 1928). Effective population size is reduced by subpopulation structure, resulting in a greater loss of heterozygosity due to genetic drift and thereby lower than expected variability for a single, panmictic population. However, other alternatives, such as age structure, may produce similar effects (Waples & Allendorf, 2015).

Most study species demonstrated discrete population structures resulting from a hierarchical stream configuration (Zbinden et al., 2022b). However, the number of sub-

populations was not significantly related to H_T. We also noted that H_T and G''_{ST} were positively related (results not shown), which conflicts with the presence of a hypothesized Wahlund effect.

Another alternative explanation for greater total diversity within smaller sample sizes may be a spurious correlation reflecting sampling bias due to the fishing gear employed. Seine nets are not equally effective across all species and are less adequate for sampling larger-bodied fishes with more rapid propulsion and greater visual acuity. Indeed, many species with elevated levels of genetic diversity were larger species (i.e., bass, sunfish, and large cyprinids) that were sampled with less efficiency and therefore totaled fewer individuals, sites, and individuals/site. The relationship could be coincidental; therefore, it is unclear whether the connection between genetic diversity and these covariates has any biological interpretation.

Conclusion

Life history traits evolve over deep and shallow timeframes as populations interact, environments differentiate, and species-level genetic diversities coalesce. Although viewed as conduits between ecology and evolutionary history, the causality of these traits has been difficult to establish unequivocally. However, if so verified, the link between ecology and evolution can be estimated by modeling appropriate traits. Genetic diversities could then be quantified for species deemed 'sibling' (i.e., near-identical morphologically), 'cryptic' (as previous but non-hybridizing), or with narrow niches/restricted distributions (e.g., short-range endemics; Davis et al., 2015). Thus, forecasting genetic diversity would be valuable for deriving conservation policy and subsequently applying management decisions (Hoban et al., 2022). A mechanism is needed to focus and prioritize management and conservation, particularly given the multiplicity of species spanning numerous distinct geographic regions. The procedure described herein can

facilitate management and streamline conservation plans by deriving diversity metrics for freshwater species currently lacking such information. Expanding the trait-based approach by incorporating more species, regions, greater phylogenetic breadth, and numerous traits will undoubtedly provide more focused insights and possibly greater predictive power. Furthermore, this community genomics approach should incorporate biotic interactions among species (sensu Hand et al., 2015) to integrate an overlooked essential biotic component. The framework herein also provides a platform for bridging the gap between micro- and macro-evolution, in that traits impinging upon genetic diversity (microevolution) could also play a significant role in speciation and extinction (macroevolution).

REFERENCES

Abrams, P.A. (2019). How does the evolution of universal ecological traits affect population size? Lessons from simple models. *American Naturalist*, *193*(6), 814–829.

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

Avise, J.C., Bowen, B.W., & Ayala, F. J. (2016). In the light of evolution X: Comparative phylogeography. *Proceedings of the National Academy of Sciences*, *113*(29), 7957–7961.

Bachman, S.P., Field, R., Reader, T., Raimondo, D., Donaldson, J., Schatz, G.E., & Lughadha, E. N. (2019). Progress, challenges and opportunities for Red Listing. *Biological Conservation*. *234*, 45–55.

Barton, K. (2022). MuMIn: Multi-Model Inference. R package version 1.46.0. https://cran.r-project.org/package=MuMIn

Bohonak, A.J. (1999). Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology*, *74*(1), 21–45.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045–1055.

Brosse, S., Charpin, N., Su, G., Toussaint, A., Herrera-R, G.A., Tedesco, P.A., & Villéger, S. (2021). FISHMORPH: A global database on morphological traits of freshwater fishes. *Global Ecology and Biogeography*, *30*(12), 2330–2336.

Burney, C.W., & Brumfield, R.T. (2009). Ecology Predicts Levels of Genetic Differentiation in Neotropical Birds. *American Naturalist*, *174*(3), 358–368.

Cammen, K.M., Schultz, T.F., Don Bowen, W., Hammill, M.O., Puryear, W.B., Runstadler, J., … Kinnison, M. (2018). Genomic signatures of population bottleneck and recovery in Northwest Atlantic pinnipeds. *Ecology and Evolution*, *8*(13), 6599–6614.

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

Chang, J., Rabosky, D.L., Smith, S.A., & Alfaro, M.E. (2019). An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in Ecology and Evolution*, *10*(7), 1118–1124.

Comte, L., & Olden, J.D. (2018). Evidence for dispersal syndromes in freshwater fishes. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172214.

Davis, M.A., Douglas, M.R., Webb, C.T., Collyer, M.L., Holycross, A.T., Painter, C.W., ... & Douglas, M.E. (2015). Nowhere to go but up: impacts of climate change on demographics of a short-range endemic (Crotalus willardi obscurus) in the sky-islands of southwestern North America. *PLoS One*, *10*(6), e0131067.

DeWoody, J.A., Harder, A.M., Mathur, S., & Willoughby, J.R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular ecology*, *30*(17), 4147-4154.

Diniz-Filho, Jose Alexandre F., Bini, L.M., Rangel, T.F., Morales-Castilla, I., Olalla-Tárraga, M.Á., Rodríguez, M.Á., & Hawkins, B.A. (2012). On the selection of phylogenetic eigenvectors for ecological analyses. *Ecography*, *35*(3), 239–249.

Diniz-Filho, José Alexandre Felizola, de Sant'Ana, C.E.R., & Bini, L.M. (1998). An eigenvector method for estimating phylogenetic inertia. *Evolution*, *52*(5), 1247–1262.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *American Biology Teacher*, *35*(3), 125–129.

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G.G., & Petit, R.J. (2007). Can Population Genetic Structure Be Predicted from Life-History Traits? *The American Naturalist*, *169*(5), 662–672.

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

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, *125*(1), 1–15.

Fobert, E.K., Treml, E.A., & Swearer, S.E. (2019). Dispersal and population connectivity are phenotype dependent in a marine metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1909), 20191104.

Frimpong, E.A., & Angermeier, P.L. (2009). Fish Traits: A Database of Ecological and Life-history Traits of Freshwater Fishes of the United States. *Fisheries*, *34*(10), 487–495.

Fricke, R., Eschmeyer, W.N., & Van der Laan, R. (2022) Eshmeyer's catalog of fishes: genera, species, references. Retrieved from https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

Giam, X., & Olden, J.D. (2016). Environment and predation govern fish community assembly in temperate streams. *Global Ecology and Biogeography*, *25*(10), 1194–1205.

Gosselin, T. (2020). R package radiator: RADseq Data Exploration, Manipulation and Visualization using R. https://thierrygosselin.github.io/radiator/

Hand, B.K., Lowe, W.H., Kovach, R.P., Muhlfeld, C.C., & Luikart, G. (2015). Landscape community genomics: Understanding eco-evolutionary processes in complex environments. *Trends in Ecology and Evolution*, *30*(3), 161–168.

Hawkins, B.A. (2012). Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography*, *39*(1), 1–9.

Hoban, S., Archer, F.I., Bertola, L.D., Bragg, J.G., Breed, M.F., Bruford, M.W., … Vernesi, C. (2022). Global genetic diversity status and trends : towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biological Reviews*. Advance online publication.

Hogg, C.J., Ottewell, K., Latch, P., Rossetto, M., Biggs, J., Gilbert, A., … Belov, K. (2022).

Threatened Species Initiative: Empowering conservation action using genomic resources. *Proceedings of the National Academy of Sciences*, *119*(4), e2115643118.

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

Hutchison, D. W., & Templeton, A. R. (1999). Correlation of Pairwise Genetic and Geographic Distance Measures: Inferring the Relative Influences of Gene Flow and Drift on the Distribution of Genetic Variability. *Evolution*, *53*(6), 1898–1914.

Hutchinson, G. E. (1965) *The Ecological Theater and the Evolutionary Play,* Yale University Press.

Jager, H. I., Chandler, J. A., Lepla, K. B., & Van Winkle, W. (2001). A theoretical study of river fragmentation by dams and its effects on white sturgeon populations. *Environmental Biology of Fishes*, *60*(4), 347–361.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Jump, A. S., Marchant, R., & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science, 14*(1), 51–58.

Kelly, R. P., & Palumbi, S. R. (2010). Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLOS ONE*, *5*(1), e8594.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology*, *74*(6), 1659–1673.

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (3rd ed.). Elsevier.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., … Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613.

López-Uribe, M. M., Jha, S., & Soro, A. (2019). A trait-based approach to predict population genetic structure in bees. *Molecular Ecology*, *28*(8), 1919–1929.

Loveless, M. D., & Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, 15, 65–95.

Lunney, D., Matthews, A., Cogger, H., & Dickman, C. (2004). The neglected 74% - the non-threatened vertebrates - and a reflection on the limitations of the process that fashioned the current schedules of threatened species in New South Wales. In P. Hutchings, D. Lunney, & C. Dickman (Eds.), *Threatened species legislation: is it just an Act?* (pp. 145–157). Royal Zoological Society of New South Wales.

Lynch, M., & Lande, R. (1998). The critical effective size for a genetically secure population. *Animal Conservation*, *1*, 70–72.

Manel, S., Guerin, P. E., Mouillot, D., Blanchet, S., Velez, L., Albouy, C., & Pellissier, L. (2020). Global determinants of freshwater and marine fish genetic diversity. *Nature Communications*, *11*(1), 1–9.

Martinez, A. S., Willoughby, J. R., & Christie, M. R. (2018). Genetic diversity in fishes is influenced by habitat type and life-history variation. *Ecology and Evolution*, *8*(23), 12022–12031.

Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Biology*, *37*(4), 329–355.

McCabe Jr., G. T., & Tracy, C. A. (1994). Spawning and early life history of white sturgeon, *Acipenser transmontanus*, in the Lower Columbia River. *Fishery Bulletin*, *92*(4), 760–772.

McCusker, M. R., & Bentzen, P. (2010). Positive relationships between genetic diversity and abundance in fishes. *Molecular Ecology*, *19*(22), 4852–4862.

Meirmans, P. G., Goudet, J., & Gaggiotti, O. E. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, *20*(15), 3144–3155.

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

Mims, M. C., Olden, J. D., Shattuck, Z. R., & Poff, N. L. (2010). Life history trait diversity of native freshwater fishes in North America. *Ecology of Freshwater Fish*, *19*(3), 390–400.

Mitton, J. B., & Lewis, W. M. (1989). Relationships between genetic variability and life-history features of bony fishes. *Evolution*, *43*(8), 1712–1723.

Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., & Quinn, T. P. (2013). Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (Oncorhynchus mykiss) population released in the wild. *Molecular Ecology*, *22*(5), 1295–1309.

Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*, *70*(12), 3321–3323.

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, *47*(3), 253–259.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2020). vegan: Community ecology package. https://cran.r-project.org/package=vegan

Olden, J.D. (2021) North American freshwater fish trait database. Unpublished data. University of Washington.

Onyutha, C. (2020). From R-squared to coefficient of model accuracy for assessing "goodness-of-fits." *Geoscientific Model Development Discussions*, *2020*, 1–25.

Osborne, M.J., Perkin, J.S., Gido, K.B., & Turner, T.F. (2014). Comparative riverscape genetics reveals reservoirs of genetic diversity for conservation and restoration of Great Plains fishes. *Molecular Ecology*, *23*(23), 5663–5679.

Papadopoulou, A., & Knowles, L.L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences*, *113*(29), 8018–8024.

Peres-Neto, P., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, *87*(10), 2614–2625.

Peres-Neto, P.R. (2006). A unified strategy for estimating and controlling spatial, temporal and phylogenetic autocorrelation in ecological models. *Oecologia Brasiliensis*, *10*(1), 105–119.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., & Hoekstra, H.E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model

and non-model species. *PLoS ONE*, *7*(5), e37135.

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

Platania, S.P., & Altenbach, C.S. (1998). Reproductive strategies and egg types of seven Rio Grande Basin cyprinids. *Copeia*, 559–569.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., … Alfaro, M.E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*(7714), 392–395.

Riginos, C., Buckley, Y.M., Blomberg, S.P., & Treml, E.A. (2014). Dispersal Capacity Predicts Both Population Genetic Structure and Species Richness in Reef Fishes. *The American Naturalist*, *184*(1), 52–64.

Rissler, L.J. (2016). Union of Phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences, 113*(29), 8079–8086.

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

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., … Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, *515*(7526), 261–263.

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.

Selkoe, K.A., Gaggiotti, O.E., Bowen, B.W., & Toonen, R.J. (2014). Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, *23*(12), 3064–3079.

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.

Sousa-Santos, C., Robalo, J.I., Pereira, A.M., Branco, P., Santos, J.M., Ferreira, M.T., … Doadrio, I. (2016). Broad-scale sampling of primary freshwater fish populations reveals the role of intrinsic traits, inter-basin connectivity, drainage area and latitude on shaping contemporary patterns of genetic diversity. *PeerJ*, *2016*(2), e1694.

Spielman, D., Brook, B.W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*, *101*(42), 15261–15264.

Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, *11*, 65-106.

Waples, R.S., & Allendorf, F. (2015). Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity*, *106*(1), 1–19.

Willoughby, J.R., Sundaram, M., Wijayawardena, B.K., Kimble, S.J.A., Ji, Y., Fernandez, N.B., … DeWoody, J.A. (2015). The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, *191*, 495–503.

Winter, D.J. (2012). MMOD: An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, *12*(6), 1158–1160.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*, 97–159.

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

Zbinden, Z.D., Douglas, M.R., Chafin, T.K., & Douglas, M. E. (2022b) Riverscape community genomics: A comparative analytical framework to identify common drivers of genetic structure. Manuscript submitted for publication.

TABLES AND FIGURES

Table 1: Freshwater fish species (N=31; listed by Family and Species) were collected across the White River Basin (Ozark Plateau, USA). SNP panels resulted from genotyping individuals using ddRAD seq. Panels of randomly selected SNPs (N=2000) were used to calculate average site genetic diversity (H_S), total genetic diversity (H_T), and global genetic fixation/differentiation index (G''_{ST}). Four covariates are shown: Indiv = Total number of individuals/species analyzed; Sites = Total number of sampling locations/species analyzed; MedianRivDist = Median of pairwise river distance between analyzed set of sites; Mean Indiv/Site = Mean number of individuals analyzed across sites where a species occurred.

Family	Species Name	Indiv	Sites	Median RivDist	Mean Indiv/Site	H_S	H_T	G''_{ST}
Atherinopsidae	*Labidesthes sicculus*	99	18	320	5.5	0.139	0.182	0.289
Centrarchidae	*Lepomis macrochirus*	63	17	423	3.7	0.221	0.239	0.104
	Lepomis megalotis	242	44	434	5.5	0.151	0.182	0.204
	Micropterus dolomieu	56	15	400	3.7	0.216	0.233	0.099
	Micropterus salmoides	15	7	309	2.1	0.283	0.303	0.102
Cottidae	*Cottus carolinae*	24	9	424	2.7	0.109	0.263	0.689
	Cottus hypselurus	40	8	476	5.0	0.063	0.241	0.813
Fundulidae	*Fundulus catenatus*	112	23	478	4.9	0.137	0.196	0.362
	Fundulus olivaceus	131	24	317	5.5	0.173	0.232	0.316
Leuciscidae	*Campostoma anomalum*	93	20	427	4.7	0.126	0.200	0.439
	Campostoma oligolepis	119	31	416	3.8	0.160	0.169	0.062
	Chrosomus erythrogaster	53	7	459	7.6	0.109	0.272	0.713
	Cyprinella galactura	72	10	325	7.2	0.176	0.183	0.050
	Cyprinella whipplei	29	6	728	4.8	0.235	0.258	0.137
	Luxilus chrysocephalus	57	13	412	4.4	0.230	0.258	0.148
	Luxilus pilsbryi	244	31	275	7.9	0.136	0.139	0.022
	Luxilus zonatus	98	16	323	6.1	0.175	0.196	0.137
	Lythrurus umbratilis	23	5	293	4.6	0.250	0.302	0.273
	Notropis boops	233	28	407	8.3	0.164	0.176	0.078
	Notropis nubilus	191	32	454	6.0	0.128	0.138	0.083
	Notropis percobromus	62	10	332	6.2	0.179	0.185	0.046
	Notropis telescopus	81	13	342	6.2	0.162	0.199	0.237
	Pimephales notatus	47	13	487	3.6	0.179	0.250	0.363
	Semotilus atromaculatus	30	9	407	3.3	0.092	0.302	0.793
Percidae	*Etheostoma blennioides*	52	14	489	3.7	0.134	0.208	0.430
	Etheostoma caeruleum	358	50	446	7.2	0.101	0.139	0.305
	Etheostoma flabellare	22	6	578	3.7	0.024	0.350	0.965
	Etheostoma juliae	57	10	300	5.7	0.161	0.227	0.371
	Etheostoma spectabile	49	10	309	4.9	0.139	0.198	0.373
	Etheostoma zonale	74	15	450	4.9	0.130	0.163	0.248
Poeciliidae	*Gambusia affinis*	35	8	415	4.4	0.234	0.260	0.148
	MEAN	92.3	16.8	408.3	5.1	0.159	0.221	0.303
	ST.DEV.	80.8	11.2	94.1	1.5	0.056	0.053	0.253

Table 2: Summary statistics for trait data compiled for freshwater fish species (N=31; Table 1) collected across the White River Basin (Ozark Plateau, USA). Traits are grouped as Morphological, Life history, and Ecological, and are hypothesized as relating to variation in facets of genetic diversity among species. Trait = Trait definition; Units = As measured; Code = Trait acronym; Categories = 0 (not present)/ 1 (present); Stdev = Standard deviation.

	Trait	Units	Code	Minimum	Median	Maximum	Mean	Stdev
Morphological	Max body length	millimeters	MaxTL	40.0	130.0	970.0	186.5	191.7
	Mean length at maturity	millimeters	LengMature	22.0	57.5	265.0	71.7	53.0
	Caudal fin aspect ratio	height^2/surface area	AspectRatio	0.5	1.4	3.1	1.5	0.6
	Body elongation	length/depth	BEl	1.9	4.1	7.0	4.1	1.0
	Vertical eye position	eye position/body depth	VEp	0.4	0.5	0.7	0.5	0.1
	Relative eye size	eye diameter/head depth	REs	0.2	0.4	0.7	0.4	0.1
	Oral gape position	mouth height/body depth	OGp	0.1	0.4	0.7	0.4	0.1
	Relative maxillary length	maxillary length/head depth	RMl	0.2	0.5	0.8	0.5	0.1
	Body lateral shape	head depth/body depth	BLs	0.4	0.6	0.7	0.6	0.1
	Pectoral fin vertical position	pectoral fin vert position/body depth	PFv	0.1	0.3	0.6	0.3	0.1
	Pectoral fin size	pectoral fin length/body length	PFs	0.1	0.2	0.3	0.2	0.1
	Caudal peduncle throttling	caudal fin depth/caudal peduncle depth	CPt	1.8	2.5	3.5	2.6	0.4
	Trait	**Units**	**Code**	**Minimum**	**Median**	**Maximum**	**Mean**	**Stdev**
Life history	Age at maturation	years	AgeMature	0.1	1.5	3.0	1.6	0.7
	Longevity	years	Longevity	2.0	4.0	16.0	5.2	3.4
	Fecundity	number eggs/season/female	Fecundity	136.5	1045.0	49000.0	5191.1	11232.7
	Mean egg diameter	mm	EggSize	0.7	1.4	2.7	1.6	0.4
	Length of spawning season	months	SpawnLength	1.0	2.5	8.0	3.0	1.6
				Categories				
	Non-guarding brood hiders	binary	NonGuardBroodHide	0.No=16	1.Yes=15	-	-	-
	Non-guarding substrate spawners	binary	NonGuardOpenSubstrate	0.No=24	1.Yes=7	-	-	-
	Egg Guarding nest spawners	binary	GuarderNestSpawner	0.No=23	1.Yes=8	-	-	-
	Spawning frequency	binary	SpawnFreq	1.Single=16	2.Multi=15	-	-	-
	Parental Care Index	discrete index	ParentEnergy	0.None=3	1.Low=16	2.Med=5	3.High=7	-
	Trait	**Units**	**Code**	**Minimum**	**Median**	**Maximum**	**Mean**	**Stdev**
Ecological	Diet Generality Index	discrete index	DietIndex	1	2	6	2.8	1.4
	Habitat Generality Index	discrete index	HabitatIndex	7	10	18	11.3	2.7
				Categories				
	Trophic guild	categorical	TrophGuild	0.Herb/detrit=4	1.Omniv=4	2.Invertiv=19	3.Inv/piscivore=4	-
	Water temperature preference	categorical	TempPref	0.Cold=1	1.Cold/Cool=3	2.Cool=9	3.Cool/Warm=12	4.Warm=6
	Benthic dwelling	binary	Benthic	0.Non-benthic=19	1.Benthic=12	-	-	-
	Surface dwelling	binary	Surface	0.Water column=12	1.Surface=19	-	-	-

Table 3: Categorization of SNPs (N=200) genotyped for freshwater fish species (N=31; Table 1) collected across the White River Basin (Ozark Plateau, USA), and scored for each of three facets of genetic diversity quantified for all species: Within-site diversity (H_S); total diversity (H_T); and among-site diversity (G''_{ST}). We tested relationships between genetic diversity variation among species and: covariates, traits (Table 2), and phylogenetic relatedness. Results for those tests are provided here.

		H_S	H_T	G''_{ST}
Covariates	Fitted Vars	none	Nindiv; Nsites; Mean_Ind_Site	none
	Multi regression	-	adj. R^2 = 0.62; p = 0.0001	-
	Selected Vars	-	Nindiv	-
	Selected adj. R^2	-	adj. R^2 = 0.5	-
Traits	Fitted Traits	MaxTL; OGp; Fecundity; EggSize; SpawnLength; TempPref; Benthic; Surface	BEl; BLs; ParentEnergy; GuarderNestSpawner *	OGp; PFv; CPt; EggSize; Benthic; Surface
	Multi regression	adj. R^2 = 0.66; p = 0.0001	adj. R^2 = 0.26; p=0.02	adj. R^2 = 0.27; p = 0.03
	Selected Traits	Surface; TempPref; EggSize; MaxTL; SpawnLength	BLs; GuarderNestSpawner	Surface; EggSize
	Selected adj. R^2	adj. R^2 = 0.68	adj. R^2 = 0.31	adj. R^2 = 0.31
Phylogeny	Fitted Eigen Vectors	EV2	EV18 *	EV4; EV6
	Multi regression	adj. R2 = 0.15; p = 0.02	adj. R^2 = 0.36; p = 0.0003	adj. R2 = 0.36; p = 0.003
	Selected Eigen Vectors	EV2	EV18	EV4; EV6
	Selected adj. R^2	adj. R^2 = 0.15	adj. R^2 = 0.36	adj. R^2 = 0.36

* For Ht, Traits and Phylogeny variables were fitted to the residuals of Ht after accounting for the relationship with the covariates

Table 4: Model selection results for average site genetic diversity (H_S) based on traits and diversity values measured for freshwater fish species (N=31; Table 1) collected across the White River Basin (Ozark Plateau, USA). Coefficients include intercept and a reduced set of traits (Table 2). R^2 = correlation coefficient; DF = degrees of freedom; LogLik = log-likelihood; AIC_c = second-order Akaike Information Criterion; Delta = change in AIC_C from best model; Weight = model weight used for averaging.

H_S Coefficients											
Intercept	**EggSize**	**MaxTL**	**SpawnLength**	**Surface**	**TempPref**	**R^2**	**DF**	**LogLik**	**AICc**	**Delta**	**Weight**
0.1484	-0.0597	0.0001	0.0146	-	0.0128	0.70	6	64.72	-113.94	0.00	0.32
0.1340	-0.0509	0.0001	0.0113	0.0243	0.0127	0.73	7	66.30	-113.74	0.20	0.29
0.1760	-0.0623	0.0002	0.0163	-	-	0.66	5	62.52	-112.65	1.29	0.17
0.1612	-0.0534	0.0001	0.0130	0.0245	-	0.69	6	63.92	-112.33	1.61	0.14
0.1415	-0.0435	0.0001	-	0.0436	0.0161	0.65	6	62.30	-109.10	4.84	0.03
0.1259	-0.0345	-	-	0.0522	0.0208	0.59	5	59.54	-106.68	7.26	0.01

Table 5: Model selection results for total genetic diversity (H_T) based on traits and diversity values measured for freshwater fish species (N=31; Table 1) collected across the White River Basin (Ozark Plateau, USA). Coefficients include intercept and a reduced set of traits (Table 2). R^2 = correlation coefficient; DF = degrees of freedom; LogLik = log-likelihood; AIC_c = second-order Akaike Information Criterion; Delta = change in AIC_C from best model; Weight = model weight used for averaging.

H_T Coefficients								
Intercept	**BLs**	**GuarderNestSpawner**	**R^2**	**DF**	**LogLik**	**AICc**	**Delta**	**Weight**
0.1286	-0.2444	0.0354	0.36	4	65.74	-121.95	0.00	0.88
0.1244	-0.2209	-	0.17	3	61.90	-116.91	5.04	0.07
-0.0081	-	0.0314	0.14	3	61.35	-115.81	6.15	0.04
0.0000	-	-	0.00	2	58.93	-113.43	8.53	0.01

Table 6: Model selection results for global genetic fixation/differentiation index (G''_{ST}) based on traits and diversity values measured for freshwater fish species (N=31; Table 1) collected across the White River Basin (Ozark Plateau, USA). Coefficients include intercept and a reduced set of traits (Table 2). R^2 = correlation coefficient; DF = degrees of freedom; LogLik = log-likelihood; AIC_c = second-order Akaike Information Criterion; Delta = change in AIC_C from best model; Weight = model weight used for averaging.

G''_{ST} Coefficients								
Intercept	**EggSize**	**Surface**	**R^2**	**DF**	**LogLik**	**AICc**	**Delta**	**Weight**
0.1183	0.1988	-0.2017	0.36	4	5.96	-2.39	0.00	0.67
0.4575	-	-0.2519	0.24	3	3.49	-0.08	2.31	0.21
-0.1040	0.2624	-	0.21	3	2.84	1.20	3.59	0.11
0.3031	-	-	0.00	2	-0.84	6.11	8.50	0.01

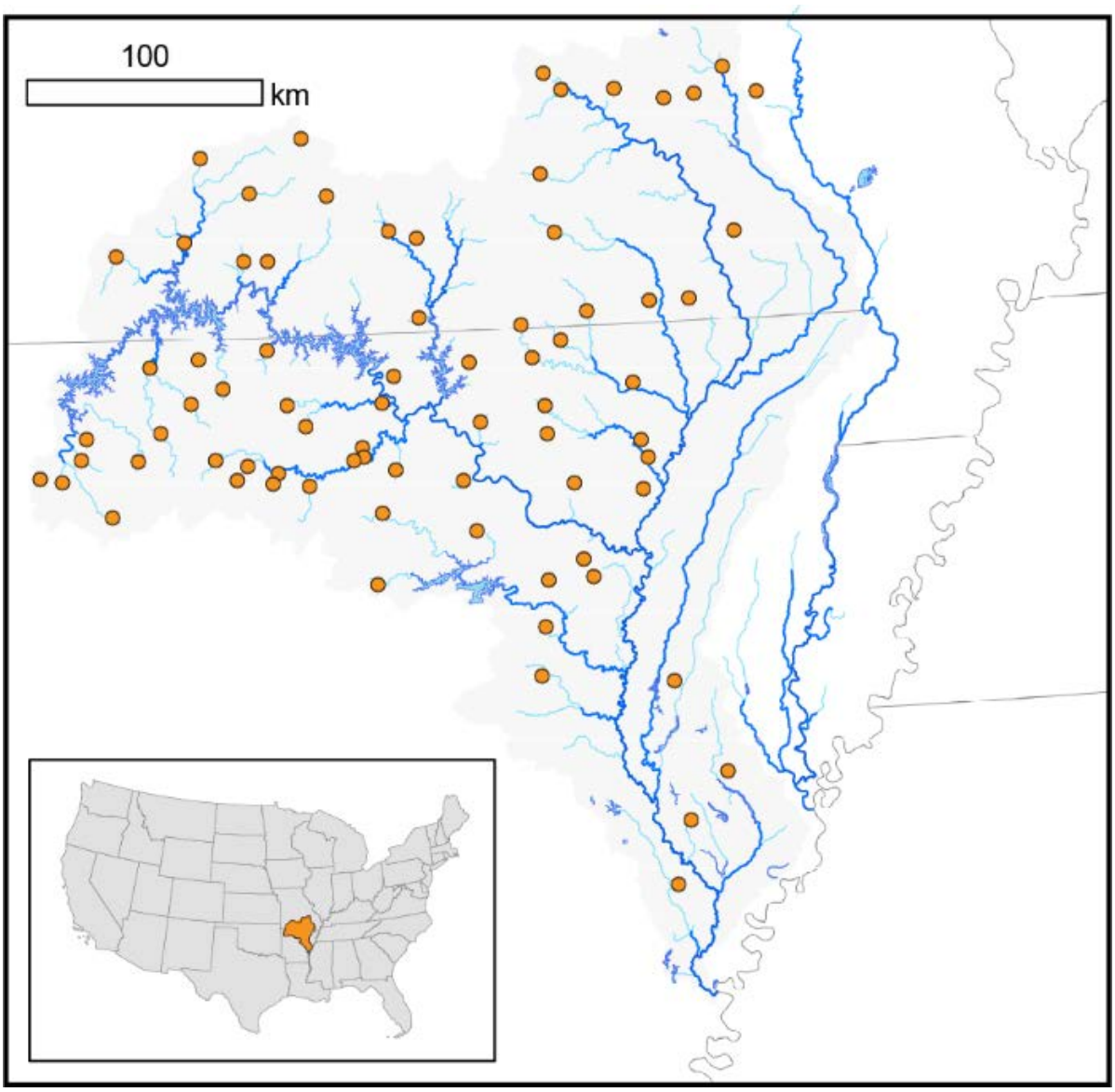


Figure 1: Freshwater fish species (N=31; Table 1) collected at *N*=75 sampling locations primarily across the White River Basin (Ozark Plateau, USA). Individuals were genotyped at N = 2000 SNPs to quantify facets of genetic diversity.

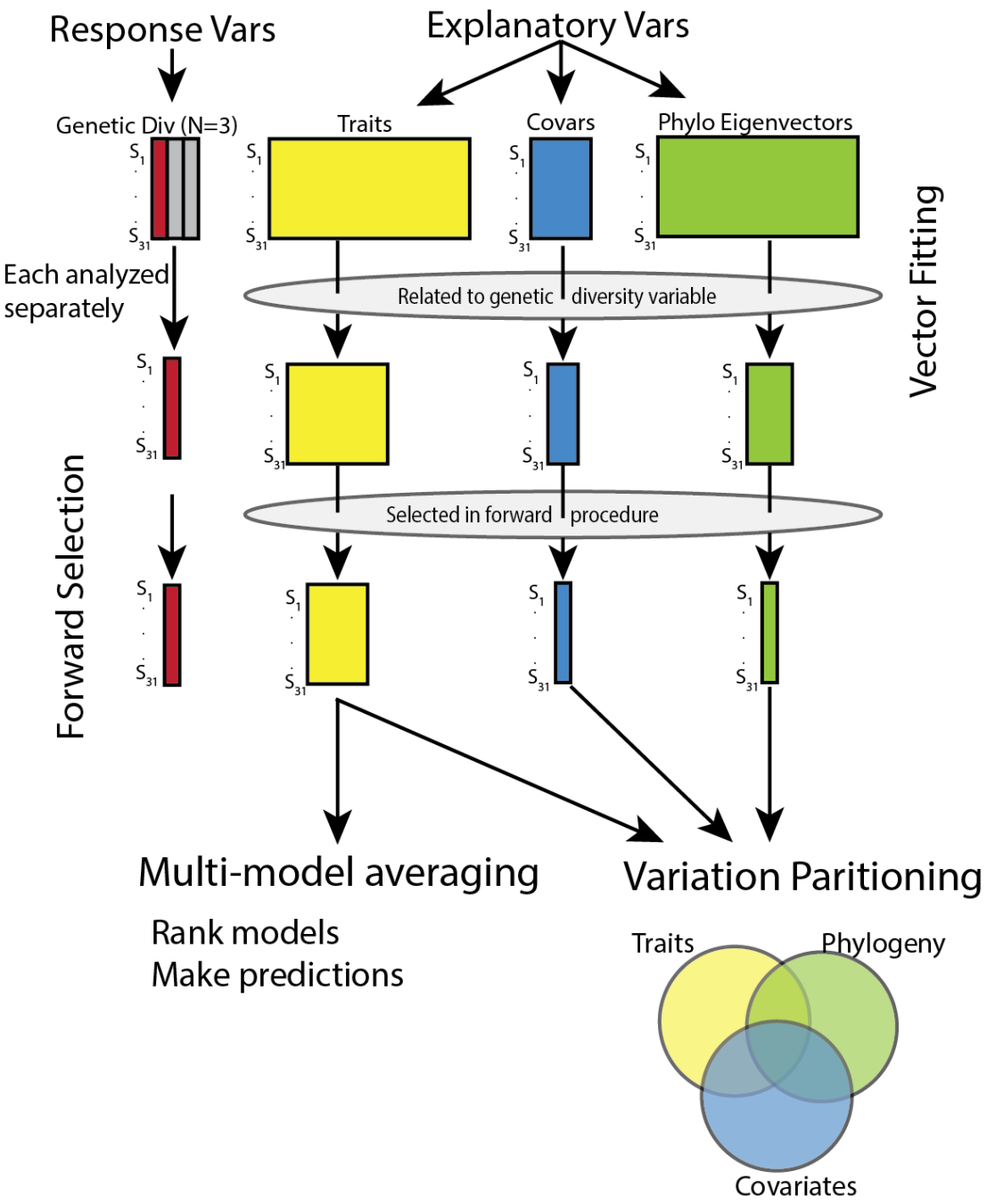


Figure 2: Overview of the framework for hypothesis testing and model building used in the evaluation of freshwater fish species (N=31; Table 1) collected at *N*=75 sampling locations primarily across the White River Basin (Ozark Plateau, USA).

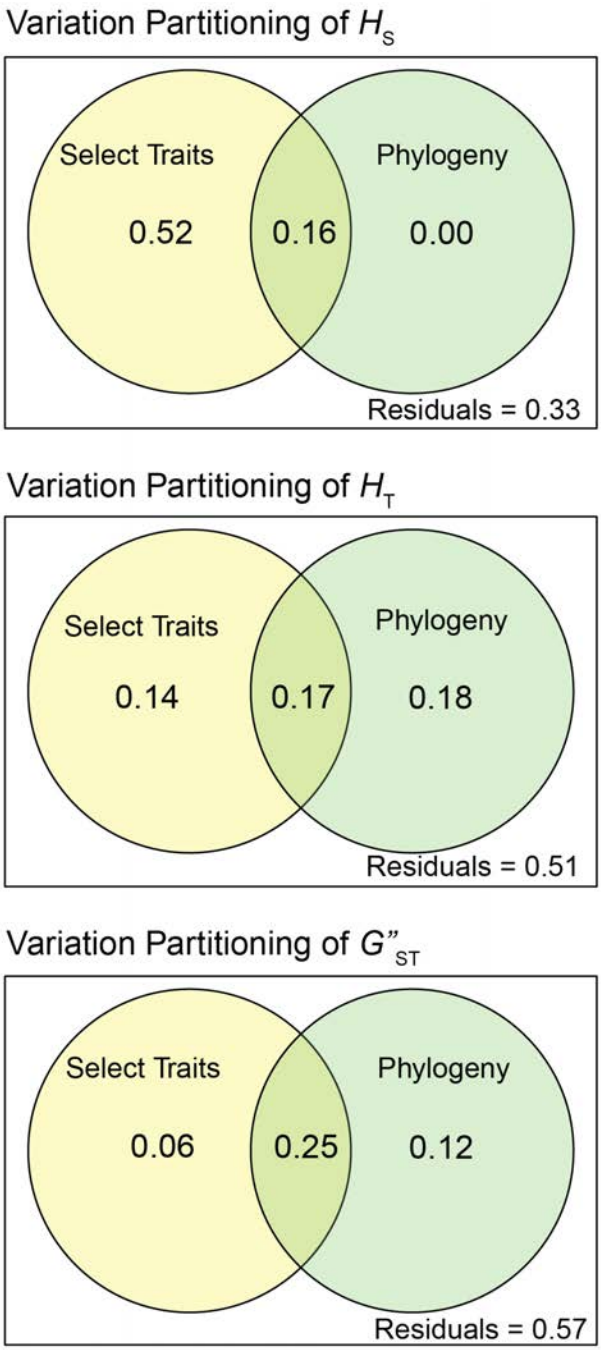
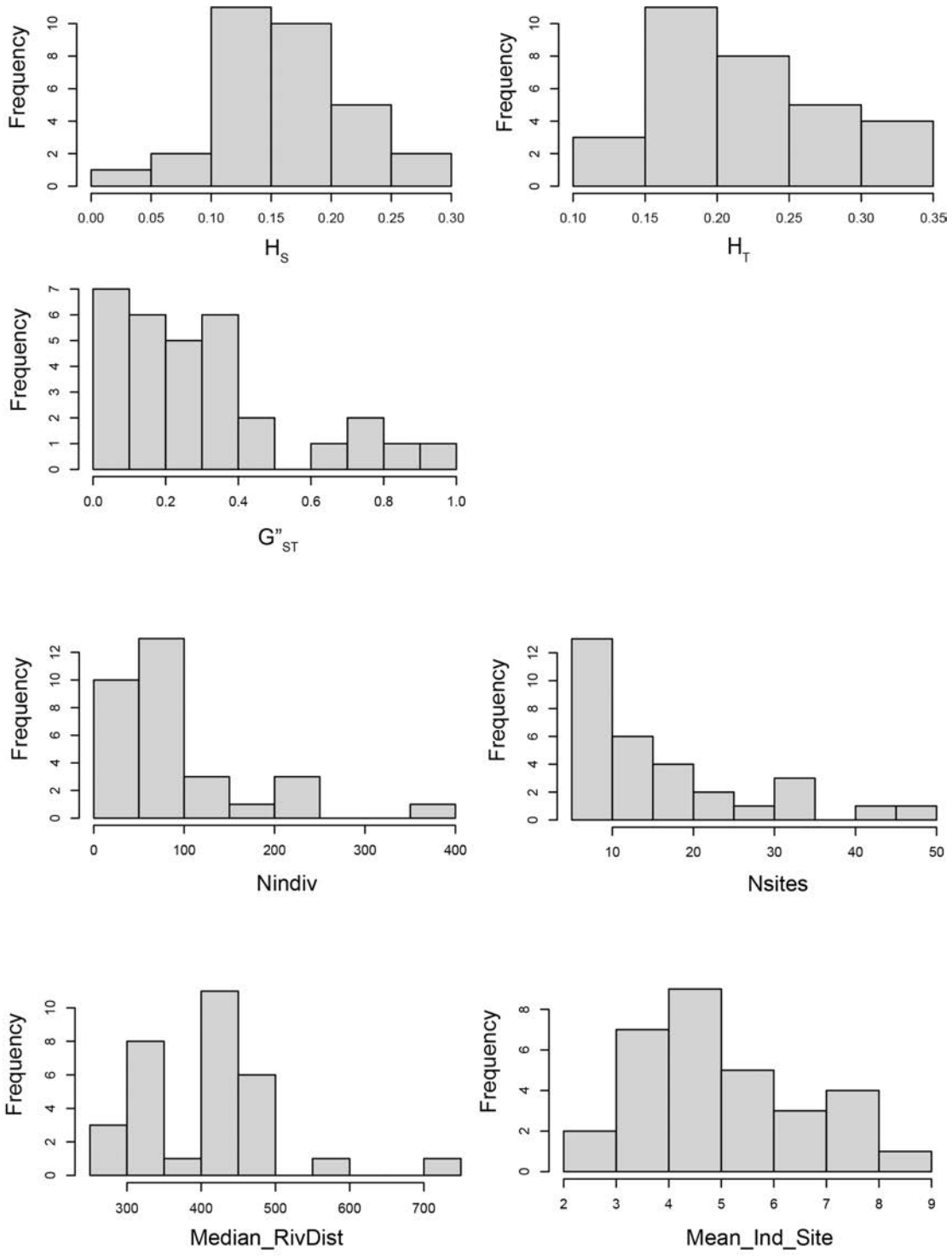


Figure 3: Variation across freshwater fish species (N=31; Table 1) collected at *N*=75 sampling locations primarily across the White River Basin (Ozark Plateau, USA) for each of three genetic diversity index: Within-site diversity (H_S); total diversity (H_T); and among-site diversity (G''_{ST}). Each index was calculated by species based on single nucleodtide polymorphisms (N=2000). Interspecific variability within each index was partitioned among selected sets of traits (Table 2) and phylogenetic eigenvectors. Values represent adjusted R^2 or the percentage of variation explained by either [A] select traits, [C] phylogeny, [B] both/phylogenetically correlated traits, or [D] unexplained variation/residuals. All values [A] & [C] > 0 were significant ($\alpha = 0.05$).

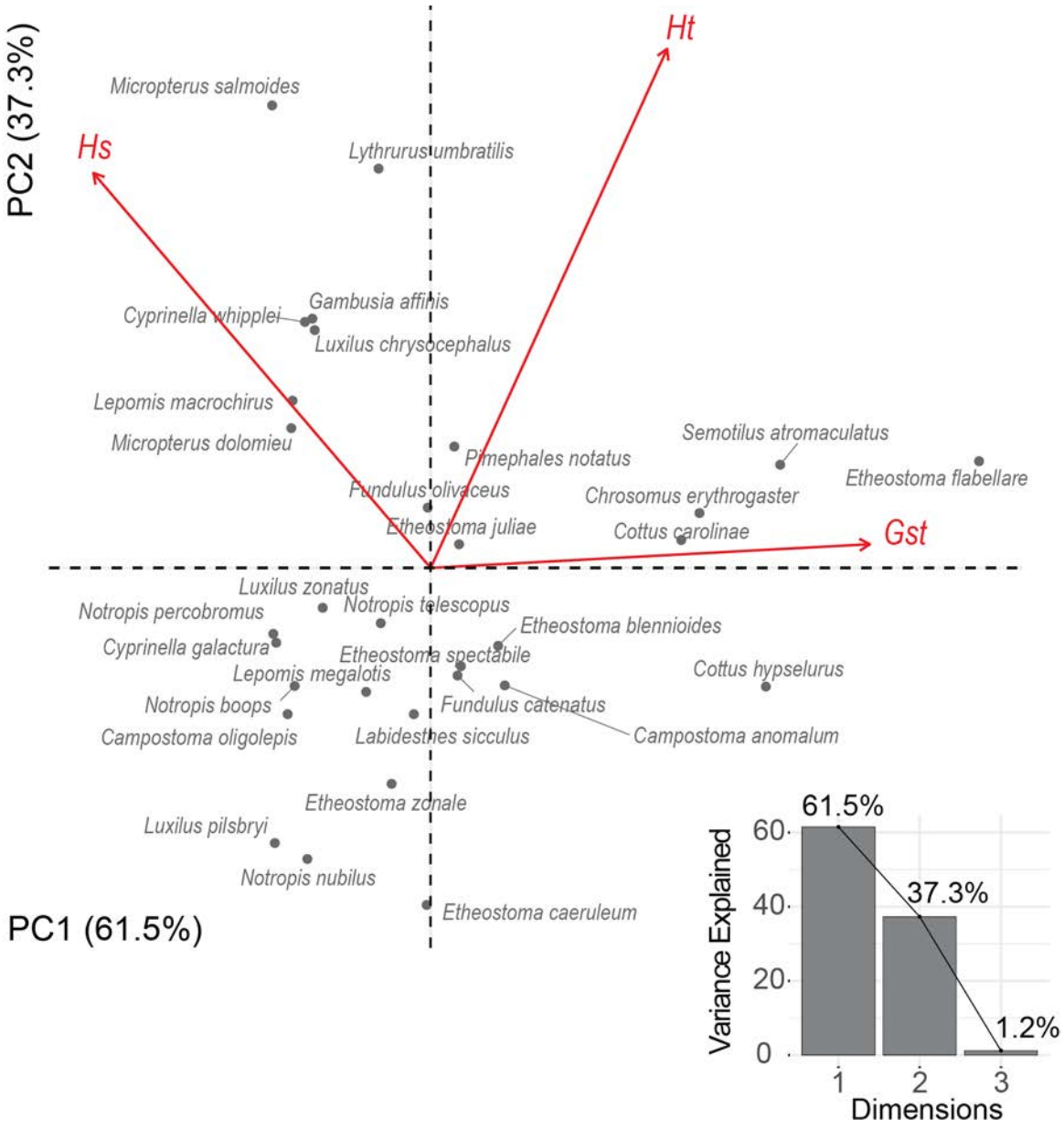
SUPPLEMENTARY MATERIALS

Supplement S1: Fish species collected across the White River Basin of the Ozarks, USA (N=31) were defined by characteristics (N=28) broadly related to morphology, life history, and ecology. These traits were gathered from a vareity of sources shown here.

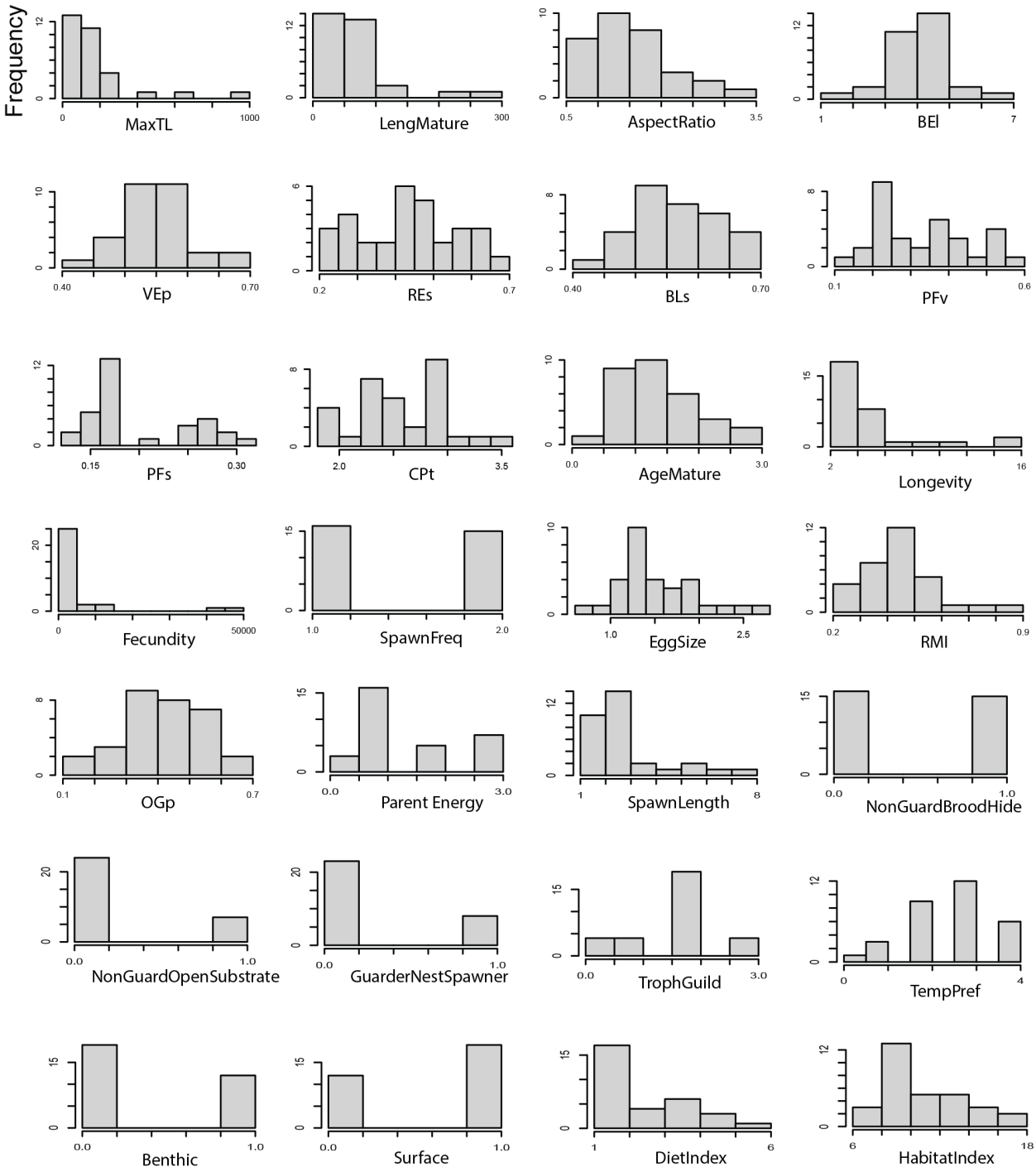
Code	Trait	Units	Source	Notes
MaxTL	Max body length	millimeters	Olden; FishMorph; FishTraits; FoA	Deferred to Olden
LengMature	Mean length at maturity	millimeters	Olden	
AspectRatio	Caudal fin aspect ratio	height^2/surface area	Olden	
BEl	Body elongation	length/depth	FishMorph	
VEp	Vertical eye position	eye position/body depth	FishMorph	
REs	Relative eye size	eye diameter/head depth	FishMorph	
OGp	Oral gape position	mouth height/body depth	FishMorph	
RMl	Relative maxillary length	maxillary length/head depth	FishMorph	
BLs	Body lateral shape	head depth/body depth	FishMorph	
PFv	Pectoral fin vertical position	pectoral fin vert position/body depth	FishMorph	
PFs	Pectoral fin size	pectoral fin length/body length	FishMorph	
CPt	Caudal peduncle throttling	caudal fin depth/caudal peduncle depth	FishMorph	
AgeMature	Age at maturation	years	Olden; FishTraits; FoA	Deferred to Olden
Longevity	Longevity	years	Olden; FishTraits; FoA	Deferred to Olden
Fecundity	Fecundity	number eggs/season/female	Olden; FishTraits; FoA	Deferred to Olden
EggSize	Mean egg diameter	mm	Olden	
SpawnLength	Length of spawning season	months	FishTraits	
NonGuardBroodHide	Non-guarding brood hiders	binary	Olden; FishTraits; FoA	
NonGuardOpenSubstrate	Non-guarding substrate spawners	binary	Olden; FishTraits; FoA	
GuarderNestSpawner	Egg Guarding nest spawners	binary	Olden; FishTraits; FoA	
SpawnFreq	Spawning frequency	binary	Olden; FishTraits	
ParentEnergy	Parental Care Index	discrete index	Olden	
DietIndex	Diet Generality Index	discrete index	Derived from FishTraits	Sum of binary diet variables
HabitatIndex	Habitat Generality Index	discrete index	Derived from FishTraits	Sum of binary habitat vars
TrophGuild	Trophic guild	categorical	Olden	
TempPref	Water temperature preference	categorical	Olden	
Benthic	Benthic dwelling	binary	Olden; FishTraits; FoA	Deferred to Olden
Surface	Surface dwelling	binary	FishTraits	



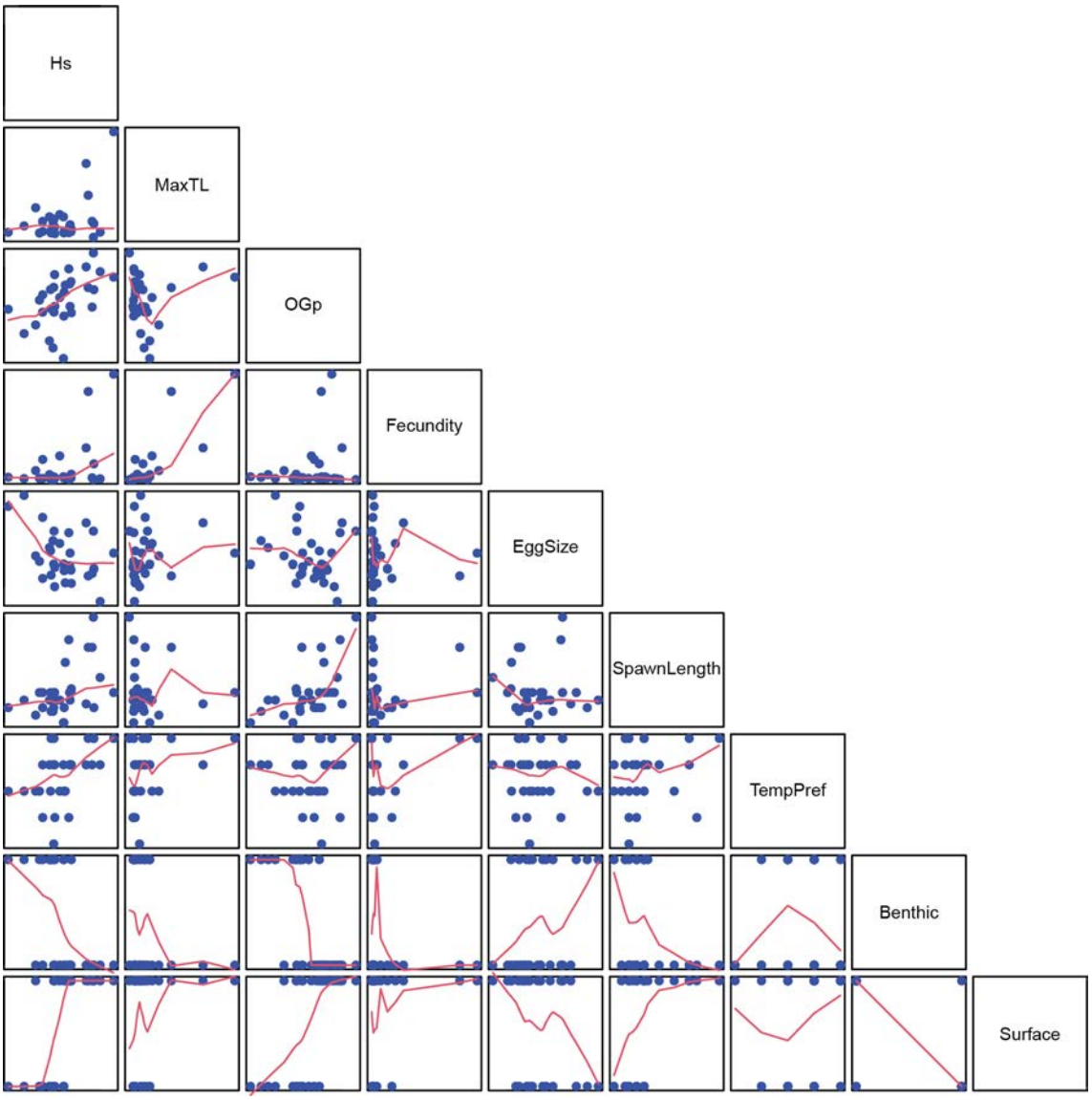
Supplement S2: Three statistics summarizing genetic diversity (HS, HT, and G”ST) and four covariates were calculated for fish species (N=31) collected across the White River Basin of the Ozarks, USA. Nindiv = the total number of individuals genotyped/analyzed; Nsites = total number of sites where genotyped/analyzed individuals were collected; Median_RivDist = the median value of pairwise river network distance between sites at which each species occurred; Mean_Ind_Site = mean number of individuals genotyped/analyzed per site.



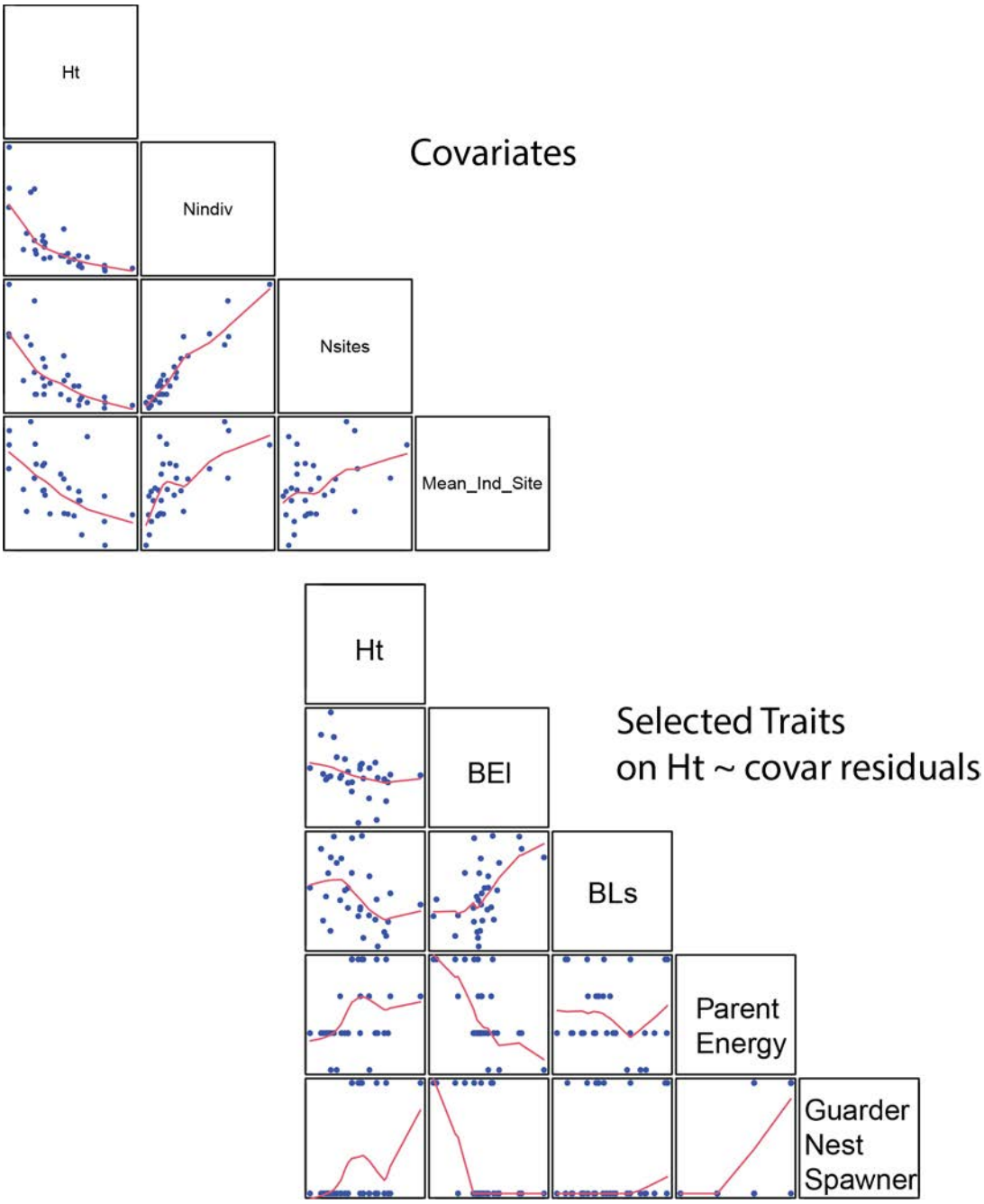
Supplement S3: Principal components analysis of three genetic summary variables quantified for thirty-one fish species collected across the White River Basin of the Ozarks, USA. H_S = heterozygosity within sites; H_T= total heterozygosity; G''_{ST} = genetic differentiation/fixation coefficient.



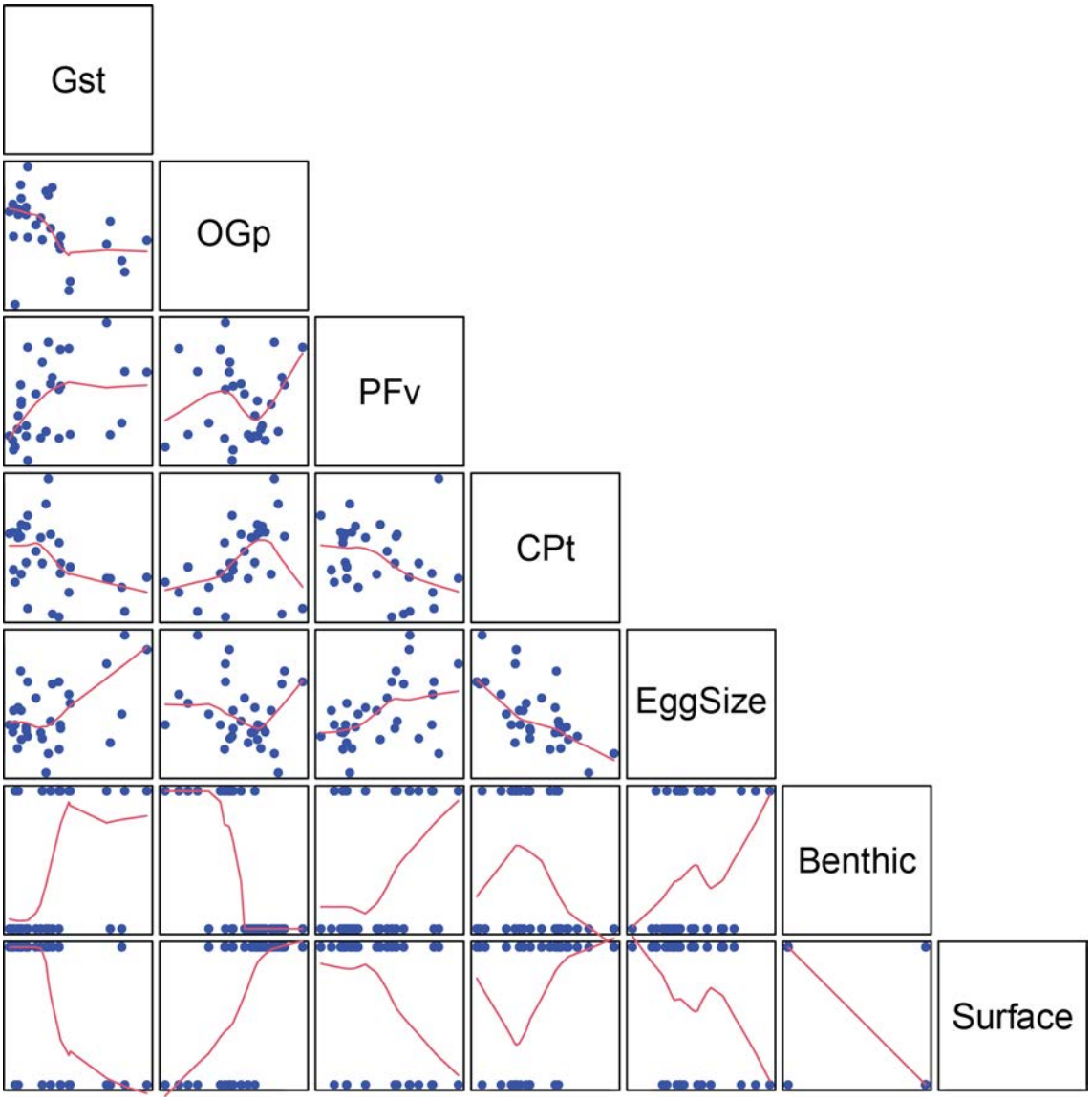
Supplement S4: These histograms show distributions of 28 trait measurments for 31 fish species collected across the White River Basin of the Ozarks, USA.



Supplement S5: Traits showing significant correlation with average site genetic diversity (H_S) quantified for N=31 freshwater fish species collected across the White River Basin of the Ozarks, USA.



Supplement S6: Covariates and traits showing significant correlation with total genetic diversity (H_T) quantified for N=31 freshwater fish species collected across the White River Basin of the Ozarks, USA. Note: traits were related to the residuals of the relationship between H_T the covariates.



Supplement S7: Traits showing significant correlation with global genetic fixation/differentiation (G''_{ST}) quantified for N=31 freshwater fish species collected across the White River Basin of the Ozarks, USA.

CONCLUSIONS

A multispecies comparative approach is necessary to disentangle the "freshwater paradox," that is the similar proportions of diversity in freshwater and marine habitats despite their tremendous difference in size (Tedesco et al., 2017). Presumably, something unique to freshwater habitats, e.g., fragmentation, underlies this paradox, and uncovering it could be essential to more effectively managing and conserving freshwater species. To probe the paradox, so to speak, a macroevolutionary approach using comparative phylogenetics could be used to infer how differences among species impact diversification rates (Bloom et al., 2013; Miller, 2021). However, this approach merely allows us to parse several broad—and non-mutually exclusive — alternatives such as differences in speciation, extinction, habitat transitions, or clade ages (Briggs, 1994). At the same time, this approach does little to reveal mechanistic underpinnings at the ecological or contemporary scale, which is most important for conservation (Rice & Emery, 2003). Therefore, my inquiry herein focused on the microevolutionary and ecological levels by assessing genetic diversity as the key metric to quantify extrinsic and intrinsic differences between species.

Understanding the FP by unraveling the processes responsible for the impressive diversity of freshwater fishes will lead to foundational knowledge and predictive schemes that can be applied to conservation and management. Similar processes have shaped the uneven richness between continents and oceans observed for all biodiversity (Briggs, 1994). Therefore, this pursuit benefits all conservation aims and is not limited to fish or aquatic organisms (Musher et al., 2022).

River networks are isolating

Appreciating the effects of the spatial structure and complexity of river networks on biodiversity is critical for freshwater fish conservation (Campbell Grant et al., 2007). River networks are hierarchically structured into repeating subunits, i.e., smaller tributaries lead to larger channels (Tarboton et al., 1988). This hierarchy also correlates with environmental gradients (Vannote et al., 1980). This structure is remarkably different from landscapes or oceanscapes and is reflected in the ecology and evolution of riverine organisms (Tedesco et al., 2012). While the isolating nature of river networks promotes divergence among lineages and may lead to higher speciation rates, the same isolation processes can lead to unstable metapopulations and extinction (Briggs, 1994). Therefore, species that can thrive in these environments must have phenotypic traits and adaptations that allow them to traverse the treacherous path between existence and extinction in isolated networks.

Species differ by nature

By their nature, species differ in some respect; otherwise, we would not differentiate between them (Simpson, 1951). Each species has a unique evolutionary history that culminates in differences we can observe today. In one sense, ecology boils down to how the spatial distribution of organisms reflects their interactions with the biotic and abiotic environment. These interactions are mediated by traits that vary within and among species. For example, larger organisms may travel further over a day or a lifetime, so they may come into contact with more types of environments and establish larger populations. Thus, differences in traits between species lead to differences in dispersal, population size, fluctuations, and connectivity (Lester et al., 2007). In turn, the traits indirectly impact evolutionary change — an impression that can be

left as an imprint on patterns of genetic diversity. A comparative approach can explore these relationships (Bohonak, 1999).

Comparative approach

A multispecies comparative approach unites ecology and evolution (Hand et al., 2015; Lowe et al., 2017). Because ecological and evolutionary processes forge differences among species, we can integrate ecology and evolution through a comparative approach to analyze how differences among species are related to each other (Bohonak, 1999). These differences include a measurable quantity of interest which can be compared to intrinsic or extrinsic properties hypothesized to govern it. For example, if one hypothesized that dispersal is an important driver of population size, we would expect a significant relationship between dispersal-related traits and quantitative measures related to population size, i.e., number of individuals or genetic diversity. Such a comparative approach used at the level of microevolutionary indicators, such as genetic diversity, can reveal impactful differences among species regarding intrinsic (i.e., traits) and extrinsic (e.g., abiotic) factors, which can be used in a comparative phylogenetic framework to test for differences in transition rates versus differences in speciation and extinction. Demonstrating that the same traits dictate micro- and macroevolutionary processes is an interesting pursuit that could further illuminate the intersection of ecology and evolution (Harvey et al., 2019; Singhal et al., 2018, 2022).

Hybridization

Hybridization plays a more critical role in ecology and evolution than previously thought (Chafin et al., 2019). In Chapter I, I infer the importance of hybridization through evidence of its widespread occurrence documented when investigated at the community level, as well as

evidence for viable hybrids and interspecies genetic exchange evidenced by admixed individuals. A multispecies approach to hybridization provides an unbiased way to investigate the occurrence of hybrids. If hybridization is only assessed between a single pair of species, we are likely to miss hybrids not involving both species but rather just one of them and a third, unstudied species. Therefore, we need to "cast a wider net," especially when exploring hybridization in fish.

Fish are likely to hybridize more than other vertebrates, and certain groups — like minnows —may be more prone to admixture than others due to their reproductive biology and genetic compatibility (Hubbs, 1955). Admixture can allow for novel genetic exchange among species leading to shifts in evolutionary trajectories and even hybrid speciation (Seehausen, 2004).

While some evidence indicates that admixture has likely played an important role in the evolution of fishes (Dowling & Demarais, 1993), is there any reason to believe that it has led to the freshwater fish paradox? In other words, is hybridization a phenomenon that promotes speciation and is somehow higher in freshwaters? Hubbs (1955) reviewed the literature then and noted that hybridization appeared rarer in marine fishes than in freshwater. However, this idea, echoed in following publications (Lagler et al., 1982), was based on hybrid references from only a handful of scientists (Schwartz, 2001). It is noteworthy that most evidence for marine fish hybridization comes from coral reef fishes, which happen to be very speciose (Walker & Ryen, 2007). Moreover, more contemporary evidence has emerged supporting higher hybridization rates in freshwater fishes (Wallis et al., 2017). This pattern may arise due to the isolating nature of river networks and their tendency to morph over long periods, which may promote secondary contact between divergent forms (Musher et al., 2022). Therefore, hybridization may play a part in the freshwater fish paradox.

Stream hierarchy

A multispecies comparative approach allows testing for general extrinsic processes governing ecology and evolution. If extrinsic, spatio-environmental factors constrain ecology and evolution, then we expect general patterns to emerge among species regarding the factors that control population structure. The Stream Hierarchy Model (SHM) suggests that fish dispersal is constrained by river network architecture and that dispersal occurs more within than between hierarchical units and across scales (Meffe & Vrijenhoek, 1988). Chapter II supported the SHM as a general model for explaining spatial patterns of genetic diversity across all 31 fish species investigated. Notably, the spatial scale at which the constraints of the river network become most apparent is the eight-digit HUC spatial scale. Furthermore, although the same general model best approximated genetic diversity, population divergence varied considerably among species, which led nicely to the final section on the trait-based approach to modeling genetic diversity.

Phenotypic traits

How species interact with the abiotic and biotic environment (ecological processes) ultimately determines their population histories (evolutionary processes). A species' characteristics mediate those interactions and thus play a role in determining the fate of evolution. Support for this hypothesis was demonstrated in Chapter III by a significant association between different measures of genetic diversity (indicators of the evolutionary process) and phenotypic traits related to dispersal, ecology, and life history. Thus, species traits could be used as an indicator of species' persistence or risk assessment. These insights offer avenues for new conservation strategies that leverage generalized ecological traits to forecast diversity metrics relevant to evolutionary trajectories in thousands of freshwater species currently lacking genetic information. These predictions can guide management efforts and prioritize species conservation

efforts, particularly in developing countries where funding, infrastructure, and training constraints limit pro-active management programs.

Final thoughts on application

Modern molecular approaches are uniquely positioned to help unite ecology, evolution, and conservation biology. With recent advances, thousands of informative genetic markers spanning entire genomes can be assessed for non-model organisms. By elevating population genomics to the community level, we add new dimensions to our potential for inquiry, namely the variation among species. Here, I demonstrated three major areas that benefit from the multispecies approach and yield valuable insights for applied efforts: (i) hybridization is widespread yet not evenly distributed among species. A multispecies approach is less biased and yields a more accurate picture of hybridization in natural communities. (ii) Spatio-environmental constraints on biodiversity should be general among species; therefore, a comparative approach allows these overarching processes to be more reliably inferred. The complexity of stream and riverine networks must be integral to the conservation and management of freshwater biodiversity because the structure of biodiversity—from genes to ecosystems—mirrors that of the network. (iii) measures of genetic diversity represent not only the history of populations and species but also the future. We can start forecasting the future by understanding the relationship between phenotypic traits and genetic diversity. The insights gathered from riverscape community genomics will only grow as the framework is applied across more diverse regions and taxa.

REFERENCES

Bloom, D.D., Weir, J.T., Piller, K.R., & Lovejoy, N.R. (2013). Do Freshwater Fishes Diversify Faster Than Marine Fishes? A Test Using State-Dependent Diversification Analyses And Molecular Phylogenetics Of New World Silversides (Atherinopsidae). *Evolution, 67*(7),

2040–2057.

Bohonak, A.J. (1999). Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology*, *74*(1), 21–45.

Briggs, J.C. (1994). Species Diversity: Land and Sea Compared. *Systematic Biology*, *43*(1), 130–135.

Campbell Grant, E.H., Lowe, W.H., & Fagan, W.F. (2007). Living in the branches: Population dynamics and ecological processes in dendritic networks. *Ecology Letters*, *10*(2), 165–175.

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

Dowling, T.E., & Demarais, B.D. (1993). Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature*, *362*, 444–446.

Hand, B.K., Lowe, W.H., Kovach, R.P., Muhlfeld, C.C., & Luikart, G. (2015). Landscape community genomics: Understanding eco-evolutionary processes in complex environments. *Trends in Ecology and Evolution*, *30*(3), 161–168.

Harvey, M.G., Singhal, S., & Rabosky, D.L. (2019). Beyond reproductive isolation: Demographic controls on the speciation process. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 75-95.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, *4*(1), 1–20.

Lagler, K.F., Bardach, J.E., & Miller, R.R. (1982). *Ichthyology*. John Wiley & Sons.

Lester, S.E., Ruttenberg, B.I., Gaines, S.D., & Kinlan, B.P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, *10*(8), 745–758.

Lowe, W.H., Kovach, R.P., & Allendorf, F.W. (2017). Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology and Evolution*, *32*(2), 141–152.

Meffe, G.K., & Vrijenhoek, R.C. (1988). Conservation Genetics in the Management of Desert

Fishes. *Conservation Biology*, *2*(2), 157–169.

Miller, E.C. (2021). Comparing diversification rates in lakes, rivers, and the sea. *Evolution*, *75*(8), 2055-2073.

Musher, L.J., Giakoumis, M., Albert, J., Del-Rio, G., Rego M., Thom, G., . . . Cracraft, J. (2022). River network rearrangements promote speciation in lowland Amazonian birds. *Science Advances*, *8*(14), eabn1088.

Rice, K.J., & Emery, N.C. (2003). Managing microevolution: restoration in the face of global change. *Frontiers in Ecology and Evolution*, *1*(9), 469–478.

Sschwartz, F.J. (2001). Freshwater and marine fish family hybrids: a worldwide changing scene revealed by the scientific literature. *Journal of the Elisha Mitchell Scientific Society*, *117*(1), 62-65.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, *19*(4), 198–207.

Simpson, G.G. (1951). The species concept. *Evolution*, *5*(4), 285–298.

Singhal, S., Colli, G.R., Grundler, M.R., Costa, G.C., Prates, I., & Rabosky, D.L. (2022). No link between population isolation and speciation rate in squamate reptiles. *Proceedings of the National Academy of Sciences*, *119*(4), e2113388119.

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

Tarboton, D.G., Bras, R.L., & Rodriguez-Iturbe, I. (1988). The fractal nature of river networks. *Water Resources Research*, *24*(8), 1317–1322.

Tedesco, P.A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H.H., Beauchard, O., … Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography*, *21*(10), 977–987.

Tedesco, P.A., Paradis, E., Lévêque, C., & Hugueny, B. (2017). Explaining global-scale diversification patterns in actinopterygian fishes. *Journal of Biogeography*, *44*(4), 773–783.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., & Cushing, C.E. (1980). The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(1), 130–137.

Walker, S.P.W., & Ryen, C. (2007). Opportunistic hybridization between two congeneric tropical reef fish. *Coral Reefs*, *26*(3), 539.

Wallis, G.P., Cameron-Christie, S.R., Kennedy, H.L., Palmer, G., Sanders, T.R., & Winter, D.J. (2017). Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes. *Molecular Ecology*, *26*(12), 3116–3127.

APPENDIX – Research Compliance

UNIVERSITY OF ARKANSAS

Office of Research Compliance

To:	Marlis Douglas
Fr:	Craig Coon
Date:	May 4th, 2017
Subject:	IACUC Approval
Expiration Date:	May 3rd, 2020

The Institutional Animal Care and Use Committee (IACUC) has APPROVED your protocol # **17077:** *Effective gene-flow between stream fish populations.*

In granting its approval, the IACUC has approved only the information provided. Should there be any further changes to the protocol during the research, please notify the IACUC in writing (via the Modification form) prior to initiating the changes. If the study period is expected to extend beyond May 3rd, 2020 you must submit a newly drafted protocol prior to that date to avoid any interruption. By policy the IACUC cannot approve a study for more than 3 years at a time.

The following individuals are approved to work on this study: Marlis Douglas and Zachery Zbinden. Please submit personnel additions to this protocol via the modification form prior to their start of work.

The IACUC appreciates your cooperation in complying with University and Federal guidelines involving animal subjects.

CNC/aem