A community genomics approach to natural hybridization

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

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

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

[*]Corresponding Author: zdzbinde@uark.edu

Abstract

Hybridization and introgression (admixture) play a complicated role in ecology and evolution. While once deemed unnatural and uncommon, admixture is now recognized as remarkably 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 admixture across 75 communities of freshwater fish within the Ozarks (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 because it does not require *a priori* assumptions of putative admixture and thus tests a wider breadth of parental species pairs. We found extensive admixture (*N*=70 individuals; 2.4%) across 18 parental species pairs involving 73% of the species analyzed (*N*=24). Most admixed individuals were minnows (Leuciscidae; *N*=15 species; *N*=66 individuals). Introgression was evident in 24 backcrossed individuals providing evidence for genetic exchange among fish lineages. Our community-level assessment suggested hybridization is even more common than now-contemporary considerations and suggests this approach is more robust for accurately surveying hybridization in nature. By contrast, literature-based per-individual hybridization rates predicted only 2-3 hybrids within our dataset.

Keywords:

admixture, ddRAD, freshwater fish, introgression, leuciscidae, multispecies hybridization

1. Introduction

Hybridization, once considered "exceedingly rare" by zoologists [1], is now recognized as relatively common, primarily because of the increased power to detect it quantifiably due to modern DNA sequencing technologies [2,3]. A more contemporary paradigm is that hybridization and introgression, i.e., admixture, play an important role in ecology and evolution [4–6], which diverges from a historical perspective rooted in the biological species concept [7].

Hybridization studies focus predominantly on single species pairs, emphasizing reproductive barriers and genomic consequences [8]. 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 [9], inflates divergence estimates in population genomic evaluations [10,11], and generally violates model assumptions underlying analytical frameworks [12]. In short, admixture's ecological, evolutionary, and analytical ramifications remain nebulous without understanding how frequently it occurs in communities.

Several issues have obscured 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 [13,14], where 'cryptic' hybridization beyond a study's purview often goes unrecognized [15,16]. This restricted focus is primarily due to the cost-prohibitive nature of multispecies evaluations. Mandates that stipulate funding for only 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 [17] to contend with reduced detection probabilities of introgressed alleles [18,19].

Broader comparative examinations often utilize meta-analyses [13] to quantify hybridization rates per species. However, studies collated within meta-analyses to detect hybridization are rarely equivalent in statistical power [20], with negative results often ignored [21]. An unbiased examination of

hybridization in nature is past due [22], and by determining its extent, we can better discern its ecological and evolutionary significance.

Freshwater fishes demonstrate a greater propensity for hybridization than other vertebrates [1,23,24,25]. Ostensibly this stems from weak reproductive isolating mechanisms, e.g., external fertilization and spatiotemporal overlap in spawning [24,26]. Furthermore, developmental incompatibilities are also reduced in fish [27].

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? And (ii) What are the rates of introgression among species?

2. Methods

(a) 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. 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 [28,29], and sequenced as pooled batches of 144 individuals per 1x100 lane on the Illumina HiSeq 4000 (Supplement S1).

(b) Data processing and assembly

Before hybrids could be identified, family-level SNP panel alignments had to be generated. Raw Illumina reads were demultiplexed and filtered in IPYRAD [30]. First, family-level phylogenies were generated to verify that individuals were correctly identified to species using PHYLORAD [31] (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 [30] 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 [32] (Supplement S1).

(c) 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; [33]). An initial DAPC (ADEGENET [34]) 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. Individual assignment probabilities were calculated via a second DAPC based on the optimum number of components.

Results were contrasted with a maximum-likelihood clustering approach that explicitly considers admixture categories (SNAPCLUST; [35]). 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 [35].

Admixture analysis employing sparse non-negative matrix factorization (sNMF) was used to refine ancestry coefficients for each family (LEA; [36]), 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 majority ancestry cluster.

(d) Hybrid classification

We contrasted models with and without introgression in SNAPCLUST to identify backcrossed individuals [2,35]. We confirmed assignments by assigning individuals as F_1, F_2, or backcrosses (NEWHYBRIDS; [37]). We also implemented HYBRIDDETECTIVE [38] 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).

3. 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 admixed individuals across genotype frequency classes (Supplement S2–S19).

(a) Detecting admixed individuals

We found more evidence of admixture than expected, although the majority was within the minnows (Leuciscidae). Admixed individuals (*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 admixed individuals were minnows

(Leuciscidae), with exceptions being: (i) *Micropterus dolomieu* x *M. salmoides* (Black Basses/Centrarchidae); (ii) *Etheostoma juliae* x *E. zonale* (Darters/Percidae); (iii) *Etheostoma spectabile* x *E. caeruleum* (Darters/Percidae); and (iv) a putative multispecific hybrid: *Noturus maydeni* x *N. albater* x *N. exilis* (Madtom Catfishes/Ictaluridae). Thus, we recognized N=18 hybridizing species-pairs and N=8 hybridizing species-triplets, i.e., multispecies hybrids.

Within the minnow family Leuciscidae, we identified 66 admixed individuals and 15 hybridizing species pairs (Supplement S20). The most remarkable occurrence (N=29 individuals; 41%) involved *Campostoma anomalum* x *C. oligolepis*. All minnow species sampled herein (N=15) appear to be hybridizing with at least one other species. 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).

(b) Classifying admixed individuals

We also found evidence of gene exchange among species, i.e., introgression. 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 [35].

4. Discussion

Hybridization plays a pivotal role in ecology and evolution because it can negatively impact the fitness of individuals, drive species to extinction, or even generate new lineages through hybrid speciation [5,6,9,11,18]. While hybridization is now generally considered a "common" phenomenon, we still expect hybrid individuals to be rare in communities because of reproductive barriers [2]. 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 [9]. 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.

(a) 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; thus, the community-level approach provides a more robust way to survey the landscape of hybridization across taxa.

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%) [13,39,49]. Our results delineated substantially higher rates in leuciscids (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 [23]. Our

study using genomic signatures was more adept at uncovering cryptic admixture not easily identified via phenotype [16].

Hybridization may be higher in fishes [1,25]. Similar per-individual rates (0–4%) were identified in marine fishes [41], with even higher rates (22.5%) among invasive Mississippi River carp [42]. However, fish hybridization studies typically focus on single species pairs, often within hybrid zones or invasion fronts [43,44], which may impede a strict comparison by biasing rates. 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 [40,45]. 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.* [20] 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 than expected [20].

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 [1,26], potentially indicating hybrids with greater viability [46]. Not surprisingly, breeding behavior within Leuciscidae is also a significant predictor of hybridization rates [26]. Many minnow species distribute gametes widely within the water column or affix them onto substrate nests that are often a shared resource among species [26,47]. Some observations have noted that as many as six different minnow species simultaneously spawn over the same gravel substrate [1]. 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 [48]. 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 [24].

(b) Introgression between species

Not only did we find evidence of 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 [49]. This genetic exchange can provide a ready source of novelty for evolutionary forces to act upon [45], and the prevalence found here indicates how probable an extraneous genetic contribution to a species's ecological and 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 [8,18]. It is intriguing that one of the most diverse groups of freshwater fish — leuciscids — 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 a role in cyprinid evolutionary history [49]. Questions remain regarding how adaptive such hybridization is and how it may differ among species and environmental contexts [50]. Such data would facilitate our capacity to predict/mitigate species collapse and offer a potential avenue for future conservation and management [51].

(c) 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 leads to admixed individuals more often 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 [52].

Another methodological observation is the underperformance of SNAPCLUST when compared to NEWHYBRIDS (a conclusion in agreement with other studies [2]). 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.

5. 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 minnows 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 [2]. 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.

Acknowledgments

We thank M. Flurry, M. George, T. Goodhart, K. Hollar, and M. Reed, who assisted with DNA extractions. The Arkansas High-Performance Computing Center provided analytical resources.

Funding

Funding was provided by the University of Arkansas Distinguished Doctoral Fellowship and Harry and Jo Leggett Chancellor's Fellowship (ZDZ), the Bruker Professorship in Life Sciences (MRD), the Twenty-First Century Chair in Global Change Biology (MED), and an NSF Postdoctoral Research Fellowship in Biology (TKC) [DBI: 2010774]. The findings, conclusions, and opinions expressed in this article represent those of the authors and do not necessarily represent the views of the NSF nor other affiliated or contributing organizations.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

ZDZ conceived the research with input from all authors. Specimen collection was done by ZDZ & TKC. Laboratory work, bioinformatics, data analysis, and manuscript drafting were done by ZDZ. All authors contributed to analysis interpretation and critically revising the manuscript. MRD and MED administered funding through their endowments.

REFERENCES

1. Hubbs CL. 1955 Hybridization between fish species in nature. *Syst. Zool.* **4**, 1–20. (doi:10.2307/sysbio/4.1.1)
2. Cordonnier M, Gayet T, Escarguel G, Kaufmann B. 2019 From hybridization to introgression between two closely related sympatric ant species. *J. Zool. Syst. Evol. Res.* **57**, 778–788. (doi:10.1111/jzs.12297)
3. Ottenburghs J. 2019 Multispecies hybridization in birds. *Avian Res.* **10**, 1–11. (doi:10.1186/s40657-019-0159-4)
4. Pfennig KS. 2021 Biased Hybridization and Its Impact on Adaptive Introgression. *Trends Ecol. Evol.* **36**, 488–497. (doi:10.1016/j.tree.2021.02.010)
5. Lowe WH, Muhlfeld CC, Allendorf FW. 2015 Spatial sorting promotes the spread of maladaptive hybridization. *Trends Ecol. Evol.* **30**, 456–462. (doi:10.1016/j.tree.2015.05.008)
6. Rhymer JM, Simberloff D. 1996 Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83–109. (doi:10.1146/annurev.ecolsys.27.1.83)
7. Harrison RG, Larson EL. 2014 Hybridization, introgression, and the nature of species boundaries. *J. Hered.* **105**, 795–809. (doi:10.1093/jhered/esu033)
8. Taylor SA, Larson EL. 2019 Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* **3**, 170–177. (doi:10.1038/s41559-018-0777-y)
9. Bangs MR, Douglas MR, Brunner PC, Douglas ME. 2020 Reticulate evolution as a management challenge: Patterns of admixture with phylogenetic distance in endemic fishes of western North America. *Evol. Appl.* **13**, 1400–1419. (doi:10.1111/eva.13042)
10. Grant PR, Grant BR. 2019 Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23216–23224. (doi:10.1073/pnas.1913534116)
11. Arnold ML. 1992 Natural hybridization as an evolutionary process. *Annu. Rev. Ecol. Syst.* **23**, 237–261. (doi:10.1146/annurev.es.23.110192.001321)
12. McVay JD, Hipp AL, Manos PS. 2017 A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc. R. Soc. B Biol. Sci.* **284**, 20170300. (doi:10.1098/rspb.2017.0300)
13. Justyn NM, Callaghan CT, Hill GE. 2020 Birds rarely hybridize: A citizen science approach to estimating rates of hybridization in the wild. *Evolution.* **74**, 1216–1223. (doi:10.1111/evo.13943)
14. Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001 The problems with hybrids: Setting conservation guidelines. *Trends Ecol. Evol.* **16**, 613–622. (doi:10.1016/S0169-5347(01)02290-X)
15. Vallender R, Robertson RJ, Friesen VL, Lovette IJ. 2007 Complex hybridization dynamics between golden-winged and blue-winged warblers (Vermivora chrysoptera and Vermivora pinus) revealed by AFLP, microsatellite, intron and mtDNA markers. *Mol. Ecol.* **16**, 2017–2029. (doi:10.1111/j.1365-294X.2007.03282.x)
16. Justen H, Kimmitt AA, Delmore KE. 2021 Estimating hybridization rates in the wild: Easier said than done? *Evolution.* **75**, 2137–2144. (doi:10.1111/evo.14082)
17. Randi E. 2008 Detecting hybridization between wild species and their domesticated relatives. *Mol. Ecol.* **17**, 285–293. (doi:10.1111/j.1365-294X.2007.03417.x)
18. Meng C, Kubatko LS. 2009 Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theor. Popul. Biol.* **75**, 35–45. (doi:10.1016/j.tpb.2008.10.004)
19. Gerard D, Gibbs HL, Kubatko L. 2011 Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evol. Biol.* **11**, 1–12. (doi:10.1186/1471-2148-11-291)
20. Schwenk K, Brede N, Streit B. 2008 Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 2805–2811. (doi:10.1098/rstb.2008.0055)
21. Kotze D, Johnson C, O'Hara R. 2004 Editorial: The journal of negative results in ecology and

evolutionary biology. *J. Negat. Results - EEB* **1**, 1–5.
22. Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018 New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am. J. Bot.* **105**, 364–375. (doi:10.1002/ajb2.1018)
23. Keck BP, Near TJ. 2009 Patterns of natural hybridization in darters (Percidae: Etheostomatinae). *Copeia* **2009**, 758–773. (doi:10.1643/CI-09-008)
24. Scribner KT, Page KS, Bartron ML. 2000 Hybridization in freshwater fishes: A review of case studies and cytonuclear methods of biological inference. *Rev. Fish Biol. Fish.* **10**, 293–323. (doi:10.1023/A:1016642723238)
25. Wallis GP, Cameron-Christie SR, Kennedy HL, Palmer G, Sanders TR, Winter DJ. 2017 Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes. *Mol. Ecol.* **26**, 3116–3127. (doi:10.1111/mec.14096)
26. Corush JB, Fitzpatrick BM, Wolfe EL, Keck BP. 2021 Breeding behaviour predicts patterns of natural hybridization in North American minnows (Cyprinidae). *J. Evol. Biol.* **34**, 486–500. (doi:10.1111/jeb.13751)
27. Thorgaard GH, Allendorf FW. 1988 Developmental genetics in fishes. In *Developmental Genetics of Animals and Plants* (ed GM Malacinski), pp. 369–391. New York, NY: MacMillan.
28. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135. (doi:10.1371/journal.pone.0037135)
29. Chafin TK, Douglas MR, Martin BT, Douglas ME. 2019 Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity.* **123**, 759–773. (doi:10.1038/s41437-019-0259-2)
30. Eaton DAR, Overcast I. 2020 Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594. (doi:10.1093/bioinformatics/btz966)
31. Fan H, Ives AR, Surget-Groba Y. 2018 Reconstructing phylogeny from reduced-representation genome sequencing data without assembly or alignment. *Mol. Ecol. Resour.* **18**, 1482–1491. (doi:10.1111/1755-0998.12921)
32. Gosselin T. 2020 R package radiator: RADseq Data Exploration, Manipulation and Visualization using R. See https://thierrygosselin.github.io/radiator/.
33. Jombart T, Devillard S, Balloux F. 2010 Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 1–15. (doi:10.1186/1471-2156-11-94)
34. Jombart T. 2008 Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)
35. Beugin MP, Gayet T, Pontier D, Devillard S, Jombart T. 2018 A fast likelihood solution to the genetic clustering problem. *Methods Ecol. Evol.* **9**, 1006–1016. (doi:10.1111/2041-210X.12968)
36. Frichot E, François O. 2015 LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929. (doi:10.1111/2041-210X.12382)
37. Anderson EC, Thompson EA. 2002 A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229. (doi:10.1093/genetics/160.3.1217)
38. Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. 2017 hybriddetective: A workflow and package to facilitate the detection of hybridization using genomic data in r. *Mol. Ecol. Resour.* **17**, e275–e284. (doi:10.1111/1755-0998.12704)
39. Mayr E. 1963 *Animal Species and Evolution*. Cambridge: Harvard University Press.
40. Mallet J. 2005 Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237. (doi:10.1016/j.tree.2005.02.010)
41. Burford MO, Bernardi G, Carr MH. 2011 Analysis of individual year-classes of a marine fish reveals little evidence of first-generation hybrids between cryptic species in sympatric regions. *Mar. Biol.* **158**, 1815–1827. (doi:10.1007/s00227-011-1694-7)
42. Lamer JT, Dolan CR, Petersen JL, Chick JH, Epifanio JM. 2010 Introgressive Hybridization between Bighead Carp and Silver Carp in the Mississippi and Illinois Rivers. *North Am. J. Fish.*

Manag. **30**, 1452–1461. (doi:10.1577/m10-053.1)

43. Matthews WJ, Turner TF, Osborne MJ. 2016 Breakdown of a hybrid swarm between two darters (Percidae), Etheostoma radiosum and Etheostoma spectabile, with loss of one parental species. *Copeia* **104**, 873–878. (doi:10.1643/CE-16-409)
44. Dowling TE, Hoeh WR. 1991 The extent of introgression outside the contact zone between Notropis cornutus and Notropis chrysocephalus (Teleostei: Cyprinidae). *Evolution* **45**, 944–956. (doi:10.1111/j.1558-5646.1991.tb04362.x)
45. Twyford AD, Ennos RA. 2012 Next-generation hybridization and introgression. *Heredity* **108**, 179–189. (doi:10.1038/hdy.2011.68)
46. Avise JC, Gold JR. 1977 Chromosomal Divergence and Speciation in Two Families of North American Fishes. *Evolution* **31**, 1–13. (doi:10.2307/2407539)
47. Peoples BK, Floyd SP, Frimpong EA. 2016 Nesting microhabitat comparison of Central stoneroller and Bluehead chub: potential inference for host-switching by nest associates. *J. Freshw. Ecol.* **31**, 251–259. (doi:10.1080/02705060.2015.1091390)
48. Matthews WJ, Marsh-Matthews E. 2017 *Stream fish community dynamics*. Baltimore, MD: Johns Hopkins University Press.
49. Dowling TE, Demarais BD. 1993 Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature* **362**, 444–446. (doi:10.1038/362444a0)
50. De Santis V, Quadroni S, Britton RJ, Carosi A, Gutmann Roberts C, Lorenzoni M, Crosa G, Zaccara S. 2021 Biological and trophic consequences of genetic introgression between endemic and invasive Barbus fishes. *Biol. Invasions* **23**, 3351–3368. (doi:10.1007/s10530-021-02577-6)
51. Gibson I, Welsh AB, Welsh SA, Cincotta DA. 2019 Genetic swamping and possible species collapse: tracking introgression between the native Candy Darter and introduced Variegate Darter. *Conserv. Genet.* **20**, 287–298. (doi:10.1007/s10592-018-1131-2)
52. Miller JM, Cullingham CI, Peery RM. 2020 The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity* **125**, 269–280. (doi:10.1038/s41437-020-0348-2)

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 are only considered for ancestry between two species (i.e., species triplets were not counted).

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, that 8 multispecific hybrid individuals are 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

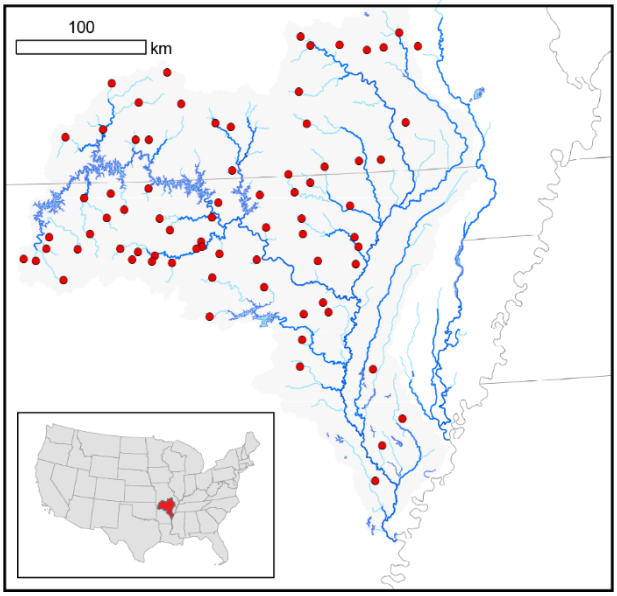

FIGURE 1 Collection sites (*N*=75; red circles) in the White River Basin (red area of insert; Missouri and Arkansas, USA). Streams/rivers are colored/sized by Strahler Stream Order (larger rivers darker/bolder).

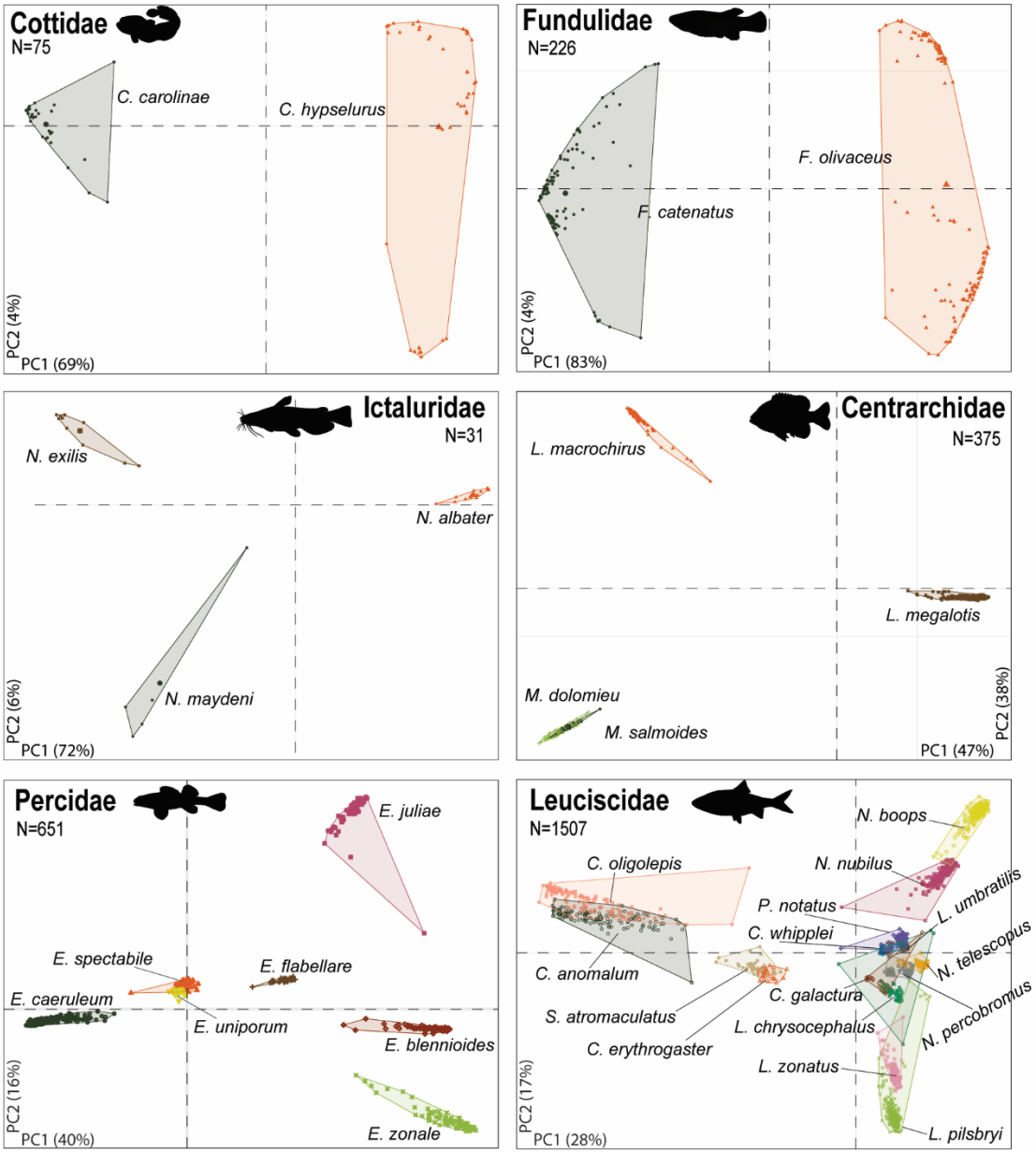

FIGURE 2 First two principal components derived from SNP genotypes in six stream fish families. Variance explained by each component is in the bottom left or right corner of each plot. N=number of individuals/family.

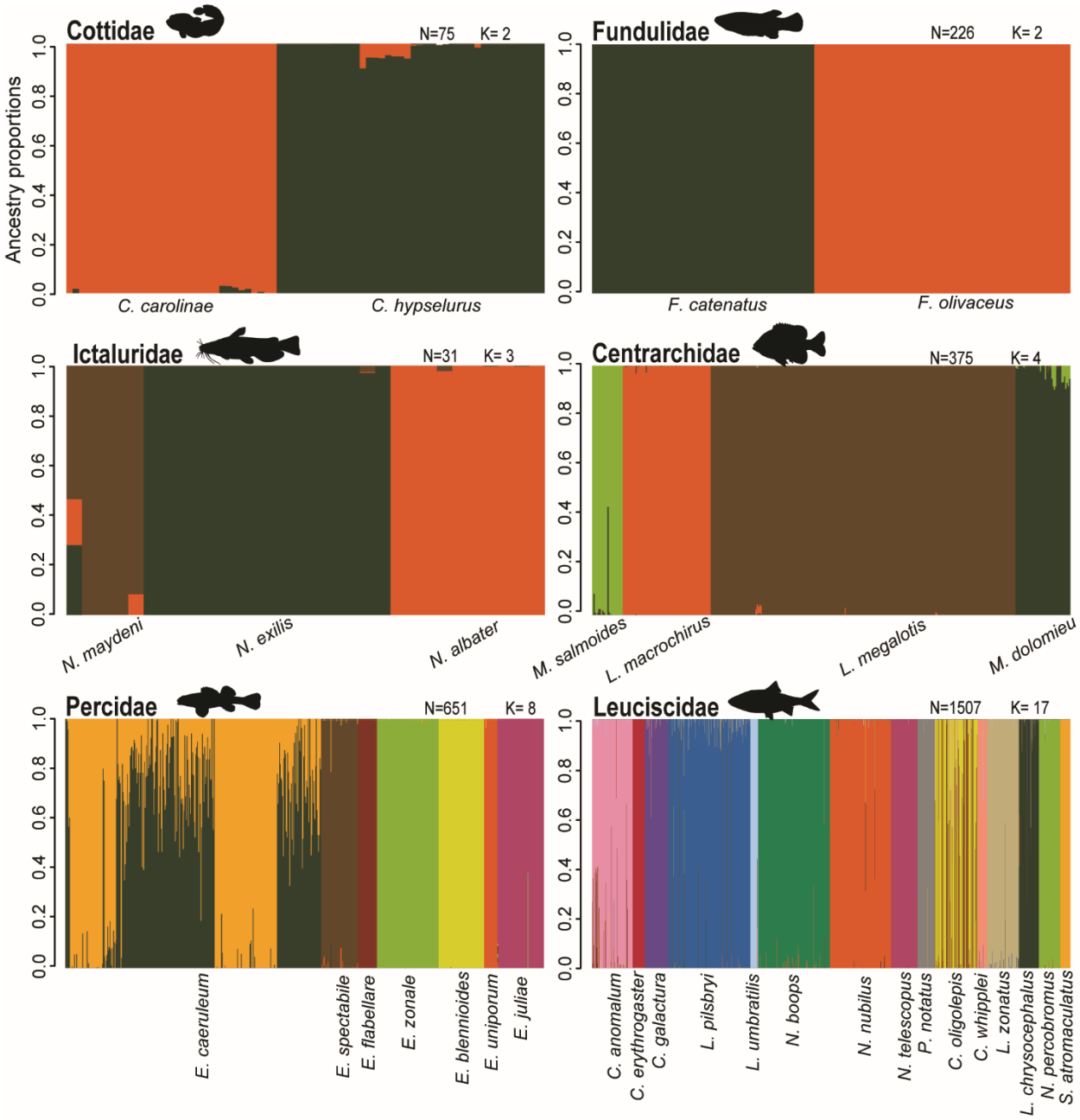

FIGURE 3 Calculated ancestral proportions for N=2,865 individuals from N=33 fish species grouped by family, where *N*=number of individuals, *K*=optimum number of clusters. When *K*>number of species (Percidae and Leuciscidae), population structure is present. Individual assignment to parental species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