Received: 8 December 2020 | Revised: 28 March 2021 | Accepted: 19 April 2021

DOI: 10.1111/mec.15947

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genomic and phenotypic divergence informs translocation strategies for an endangered freshwater fish

Liam U. Taylor[1] | Edgar Benavides[1] | Jeffrey W. Simmons[2] | Thomas J. Near[1,3]

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

[2]Tennessee Valley Authority, Chattanooga, TN, USA

[3]Yale Peabody Museum of Natural History, New Haven, CT, USA

Correspondence
Liam U. Taylor, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA.
Email: liam.taylor@yale.edu

Funding information
National Science Foundation Graduate Research Fellowship for Liam U. Taylor, Grant/Award Number: DGE1752134; Peabody Museum Bingham Oceanographic Fund; Tennessee Valley Authority

Abstract

Translocation, the movement of organisms for conservation purposes, can result in unintended introgression if genetic material flows between populations in new ways. The Bluemask Darter *Etheostoma akatulo* is a federally endangered species of freshwater fish inhabiting the Caney Fork River system and three of its tributaries (Collins River, Rocky River, and Cane Creek) in Tennessee. The current conservation strategy for Bluemask Darters involves translocating the progeny of broodstock from the Collins River (in the west) to the Calfkiller River (in the east) where the species had been extirpated. In this study, we use ddRAD sequence data from across the extant range to assess this translocation strategy in light of population structure, phylogeny, and demography. We also include museum specimen data to assess morphological variation among extant and extirpated populations. Our analyses reveal substantial genetic and phenotypic disparities between a western population in the Collins River and an eastern population encompassing the Rocky River, Cane Creek, and upper Caney Fork, the two of which shared common ancestry more than 100,000 years ago. Furthermore, morphological analyses classify 12 of 13 Calfkiller River specimens with phenotypes consistent with the eastern population. These results suggest that current translocations perturb the evolutionary boundaries between two delimited populations. Instead, we suggest that repopulating the Calfkiller River using juveniles from the Rocky River could balance conflicting signatures of demography, diversity, and divergence. Beyond conservation, the microgeographic structure of Bluemask Darter populations adds another puzzle to the phylogeography of the hyperdiverse freshwater fishes in eastern North America.

KEYWORDS
Caney Fork River system, darter, ddRAD, Endangered Species Act, *Etheostoma akatulo*, microendemism

1 | INTRODUCTION

During translocations, the movement of organisms across geographic boundaries disrupts ecological and evolutionary processes for the purposes of conservation (Conant, 1988; Moritz, 1999). Specific conservation priorities vary with different biological and logistical contexts. For example, some translocations are focused on the demographic problem of reintroducing organisms into open habitats (Seddon, 2010). Others aim to offset inbreeding depression or provide phenotypic capacities through the introduction of alleles to existing populations (i.e., "genetic rescue"; Tallmon et al., 2004; Weeks et al., 2011). Ultimately, the effectiveness of translocation can rely on both the flow

of organisms and the flow of alleles. Just as the success of genetic rescue depends on the survival and reproduction of introduced organisms (e.g., Bouzat et al., 2009), the success of organism reintroduction can depend on patterns of genetic diversity such as founder effects (e.g., Jamieson, 2011). Indeed, modern guidelines require explicit attention to not only the demographic risks and opportunities of translocation for imperiled populations, but also the genetic risks and opportunities of translocation between sites with different levels of allelic diversity (e.g., George et al., 2009; IUCN/SSC, 2013).

Further complications arise whenever translocations alter the flow of organisms and alleles across population or lineage boundaries (Laikre et al., 2010). Translocation of organisms across populations can break the genetic boundaries of allele variation that reflect unique processes of mutation, reproduction, and migration among populations (Allendorf et al., 2010; Johnson, 2000). In more extreme cases, translocating organisms between distinct lineages can collapse the phylogenetic structure that harbours diverged, and diverging, evolutionary histories (Faith, 1992; Moritz, 1995). The breakdowns of genetic and phylogenetic boundaries can result in corresponding phenotypic shifts, which might at least disturb novel trait diversity and at most threaten the adaptations that help organisms function in local habitats (Storfer, 1999; but see Weeks et al., 2011). The best translocation strategies must therefore balance patterns of divergence with the issues of demography and diversity.

One crucial difficulty is identifying the divergence boundaries relevant to translocation, especially when they are not clearly outlined by breaks in geography, habitat, or distribution. Such is often the case for North American freshwater fishes. Concentrated in the southeastern portion of the continent, North American fish fauna are the most species-rich of any non-tropical region on Earth (Lundberg et al., 2000). At the species level, diversity has been shaped by continent-scale allopatry initiated via Pliocene-Pleistocene glaciation, sea level change, and waterway discontinuity (Mayden, 1988; Near & Keck, 2005; Soltis et al., 2006). However, broad paleogeographic events do not fully capture the boundaries that have been important for the evolution of these fishes. For example, there is evidence that freshwater fish diversification and speciation has occurred within, as opposed to between, disjunct highland areas and river systems (Hollingsworth & Near, 2009; Keck & Near, 2010). Even within contiguous waterways, anthropogenic factors such as dams, reservoirs, and habitat fragmentation (e.g., Beneteau et al., 2009; Haponski et al., 2007; Yamamoto et al., 2004) can compound longer-term evolutionary processes such as isolation-by-distance (e.g., Argentina et al., 2018; Zieritz et al., 2010) and result in genetic structure that varies between populations.

Amidst these intricate patterns of diversity, at least 40% of North American freshwater fishes are imperiled (Jelks et al., 2008). Captive propagation, translocation, and reintroduction of species to areas of extirpation are being increasingly deployed to combat declines in diversity and abundance (George et al., 2009). In North America, the translocation of nongame fishes has mainly been focused at the organismal level, with the goals of reintroducing individuals to historical habitats or increasing the size and stability of existing populations (Andreasen & Springer, 2001; Doll et al., 2020; Shute et al., 2005). Although some of these translocations have also sought to bolster genetic variation of the target populations (e.g., Shute et al., 2005) there has been relatively little focus on the hidden histories of genetic material being moved with the fishes.

The Bluemask Darter (*Etheostoma akatulo* Layman & Mayden, 2009) is a federally endangered freshwater species currently the target of translocation efforts. This species is part of the diverse clade of North American darters that includes iconic imperiled species such as the Snail Darter (*Percina tanasi*). *Etheostoma akatulo* was listed as endangered 16 years before it was formally described as a species (Layman & Mayden, 2009; U. S. Fish & Wildlife Service, 1993). Endemic to the Caney Fork River system upstream of the Great Falls in central Tennessee, this species presently occurs in the Collins River, Rocky River, Cane Creek, and Caney Fork (Figure 1). *Etheostoma akatulo* also inhabited the Calfkiller River as recently as 1968 but has since been extirpated in that waterway (Layman et al., 1993; TVA unpublished data). The Collins River is in the west of the Caney Fork River system, while the Rocky River, Calfkiller River, Cane Creek, and upper Caney Fork comprise the eastern portion of the system.

Despite its restricted geographic range within a single river system, *Etheostoma akatulo* exhibits evidence of population structure among its extant populations. An analysis of 10 microsatellite loci identified three genetically distinct populations: one in the Collins River, one in the Rocky River, and one in the Cane Creek and Caney Fork (Robinson et al., 2013). The corresponding patterns of genetic variation support earlier reports that the Collins River population is more abundant and genetically diverse than other extant populations (Layman & Mayden, 2009; Layman et al., 1993; Jeffrey W. Simmons, unpublished data). In the context of these results, the current conservation strategy for the Bluemask Darter involves the capture of adult fish from the Collins River for captive propagation and reintroduction into open habitats in the Calfkiller River (Petty et al., 2020; Figure 1).

However, the microsatellite identification of three Bluemask Darter populations by Robinson et al. (2013) cannot fully inform the appropriate translocation source for Calfkiller River sites. There are three possibilities for a set of three populations: (i) they are equidistant from one another in genetic, phylogenetic, or phenotypic space; (ii) the central population is a mixture of the two more distant populations; or (iii) they are hierarchically nested, with two populations being more closely related to one another than to the third population. In consequence, the best translocation source depends on the structure of population structure, including an assessment of how the targets of reintroduction fit into that structure and how different populations might demographically sustain perturbations to structure.

Here, we combine genomic tools and museum specimen records to investigate the evolutionary history and translocation boundaries of *Etheostoma akatulo* via population structure assessment, phylogenetic inference, demographic modelling, and morphological analysis. First, we use genome-wide, next-generation sequencing (NGS) to parse fine-scale genetic population structure using hundreds of single-nucleotide polymorphisms (SNPs).

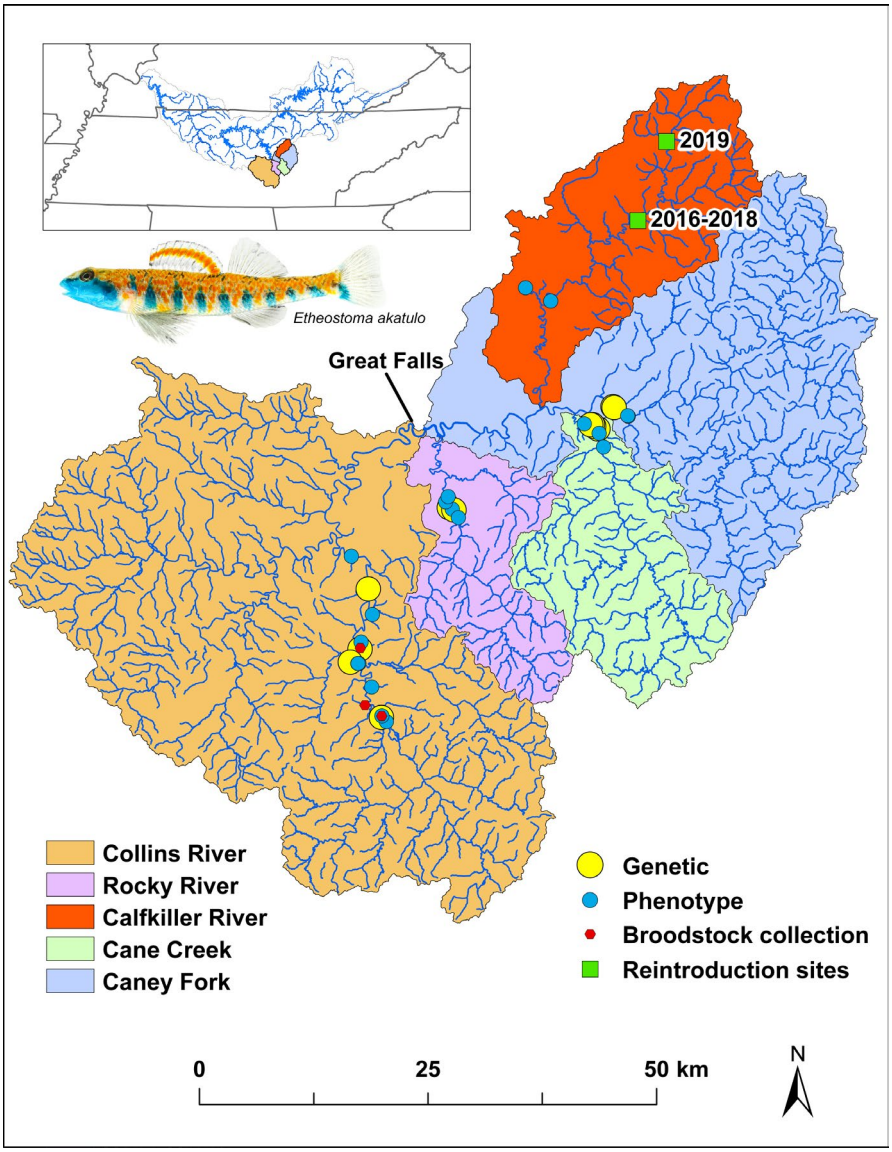

FIGURE 1 Collection sites across the range of the Bluemask Darter (*Etheostoma akatulo*). Broodstock from the Collins River are being propagated for reintroduction to sites in the Calfkiller River where the species has been extirpated (Photograph: Todd Amacker)

Second, we use tens of thousands of loci to generate an explicit phylogenetic hypothesis for lineages from different tributaries, including divergence time estimates. Third, we use demographic modelling to estimate effective population sizes and migration parameters from genetic data. Fourth, we couple our genetic analyses with an analysis of phenotypic variation across both extant and extirpated populations. We combine these methods to highlight risks and opportunities of different translocation strategies for this endangered species.

2 | MATERIALS AND METHODS

2.1 | Sample collection and DNA extraction

We sampled specimens of *Etheostoma akatulo* from sites that encompassed the complete extant range of the species (Figure 1). Samples were small tissue biopsies from the upper lobe of the caudal fin. To minimize injury and mortality, tissue biopsies were only sampled from individuals with a total length >30 mm and all specimens were released at their collection site. All samples were collected in 2017, except for five Collins River samples from one site (Scott Creek) collected in 2006. We also included several previously collected specimens of species closely related to *Etheostoma akatulo* for outgroups in phylogenetic and molecular divergence time analyses: two specimens of *Etheostoma stigmaeum* (from Crooked Creek, Jefferson County, AL, USA), two *Etheostoma meadiae* (Copper Creek, Scott County, VA, USA), two *Etheostoma jessiae* (McNair Creek, Franklin County, AL, USA), and one *Etheostoma chlorosoma* (Wildcat Creek, Calloway County, KY, USA). We extracted genomic DNA from ethanol-preserved tissue samples using the Qiagen DNeasy Blood and Tissue Kit following the manufacturer's protocol (Qiagen). We verified the quality of DNA extractions with gel electrophoresis and quantified concentration with a Qubit 3.0 Fluorometer (Thermo Fisher Scientific).

2.2 | Sequencing and sequence assembly

We genotyped 76 *Etheostoma akatulo* specimens and seven outgroup specimens using genome-wide, double digest restriction site-associated DNA sequencing (ddRAD-seq; Table 1; Peterson et al., 2012; Poland et al., 2012). The sampling of *Etheostoma akatulo* included 32 specimens from the Collins River, 16 from the Rocky River, 20 from Cane Creek, and eight from the upper Caney Fork (Figure 1). Library preparation began with an ethanol precipitation to concentrate and purify DNA extractions, followed by DNA digestion with *PstI* and *MspI* restriction enzymes (New England BioLabs). Digested fragments were ligated with adapters that included both sample-specific barcodes and shared PCR amplification targets. After ligation, samples were pooled and purified with the Qiaquick PCR Purification Kit (Qiagen), amplified with PCR, and purified again. To maintain a balanced representation of samples in each library, we monitored pool concentrations using a Qubit 3.0 Fluorometer. Once the final libraries were prepared, we performed a size selection step for DNA fragments between 300 and 500 base pairs (bp) in length using a BluePippin 2% cassette (Sage Science). Libraries were sequenced for 100 bp single-end reads on an Illumina HiSeq2000 at the Genomics & Cell Characterization Core Facility, University of Oregon (Eugene, OR, USA). Sequencing resulted in a total of 236,583,914 reads across the 83 specimens (mean: 2,850,409 reads; min: 593,350 reads; max: 7,499,428 reads).

We assembled and aligned ddRAD reads using ipyrad v. 0.9.53 (Eaton & Overcast, 2020). Samples from different libraries were demultiplexed separately according to their barcode sequences. Assembly used an *Etheostoma spectabile* reference genome (GenBank: GCA_008692095.1; 854.807 Mb; 104,340 contigs; contig N50: 26,023). The more closely related of the two *Etheostoma* reference genomes currently available, *E. spectabile* shares a common ancestor with *E. akatulo* and outgroup species that dates to about 30 million years ago (Ma) (Near et al., 2011). Assembly parameters included strict adapter filtering, no barcode mismatches, and default values for minimum read length (>35 bp after adapter trimming), maximum SNPs per locus (0.2), maximum indels per locus (8), and maximum heterozygous sites per locus (0.5).

We created four separate workflow branches for ipyrad, allowing us to apply different assembly and filtering parameters while constructing genetic data sets for population structure, phylogeny, divergence time, and demographic analyses, respectively (Table 1). These branches differed in two ways: the specimens included for assembly and the thresholds for locus retention across specimens. Population structure analyses are sensitive to missing data (e.g., Wright et al., 2019), and do not require outgroup taxa. For these analyses, we assembled data across *Etheostoma akatulo* specimens only and retained only loci present across 100% of specimens. Because these loci could still contain missing sites, and therefore missing SNP data, we also established five subbranches to test the sensitivity of our population structure results to increasing amounts of missing data (Graham et al., 2020). These five subbranches retained loci present across at least 50%, 60%, 70%, 80%, and 90% of individuals, respectively, and underwent the same downstream filtering as our main population structure branch. Our demographic analyses did not allow for missing data and do not require outgroup taxa. Because we later apply additional filtering and projection steps to remove all missing sites before demographic analysis, we first required a more relaxed locus retention threshold of 80% for this branch. We included the relevant outgroup specimens in the assembly branches for both phylogeny and divergence time analyses (Table 1). Our phylogeny branch allowed for a greater proportion of

TABLE 1 Assembly branches for genetic data sets using ipyrad and additional filtering tools. The individuals-per-locus column shows the minimum number for individuals required to have data at a given locus in order to retain that locus in the data set. The number of loci per individual column indicates loci following assembly but before additional SNP filtering, including summary data for those branches that retained loci missing from some individuals. The number of SNPs column indicates the total number of SNPs across all individuals after loci filtering and subsampling. The divergence time branch included four *Etheostoma akatulo* specimens from the Collins River and two *E. akatulo* specimens each from the Cane Creek and upper Caney Fork. The demography branch shows the number of sites following projection down to 20 synthetic individuals in each of two populations (west and east) to eliminate missing data

Branch	Analyses	Individuals (*N*)	Individuals per locus	No. loci per individual	No. SNPs	Missing SNPs
Population structure	fastStructure, DAPC, fineRADstructure, F_{ST}, heterozygosity, isolation-by-distance	*E. akatulo* (76)	100%	1,593	497	15.1%
Phylogeny	IQTree	*E. akatulo* (76) *E. stigmaeum* (1) *E. chlorosoma* (1)	50%	Mean: 54,696 Min: 13,168 Max: 67,719	240,485	35.5%
Divergence time	SNAPP	*E. akatulo* (8) *E. stigmaeum* (2) *E. meadiae* (2) *E. jessiae* (2) *E. chlorosoma* (1)	100%	10,721	9,933	11.89%
Demography	fastsimcoal2	*E. akatulo* (76) (projected to 20 per population)	80%	Mean: 35,548 Min: 7,142 Max: 38,482	12,967 sites (projected)	0%

missing data (only 50% individuals required per locus), which does not impede phylogenetic inference given many loci (Eaton et al., 2017), whereas our divergence time branch required maximal loci coverage (100% individuals required per locus) to minimize missing SNP data.

After ipyrad assembly, the population structure branch (and missing data subbranches) were filtered with VCFtools v. 0.1.16 (Danecek et al., 2011) to exclude singleton and private doubleton alleles (Linck & Battey, 2019) and to retain only biallelic SNPs >10,000 bp apart on the reference genome. The phylogeny branch required no additional filtering following assembly. The divergence time branch was filtered to retain one biallelic SNP per locus and formatted for analysis using the phrynomics package (https://github.com/bbanbury/phrynomics). Demographic analyses were performed using site frequency spectrum (SFS) information that cannot include missing data. To reduce linkage effects, we first filtered output in this branch with VCFtools to retain only SNPs >10,000 bp apart. We then used SFS-scripts (https://github.com/marqueda/SFS-scripts) to project filtered SNP data from the demography branch down to a folded SFS with a smaller number of synthetic individuals (20 per population) and fewer segregating sites but no missing data. We used PLINK v.1.9 (Chang et al., 2015) as well as the vcfR and tidyverse packages in R v. 3.6.3 (Knaus & Grünwald, 2017; R Core Team, 2020; Wickham et al., 2019) to format filtered files for analysis and visualization.

2.3 | Population structure

We used three methods to assess *Etheostoma akatulo* population structure. First, the Bayesian method fastStructure was used to assess the number of populations (*k*, from $k = 1$ to $k = 20$ populations) that maximized marginal likelihood and best explained variation of SNPs in our genetic data (Raj et al., 2014). Second, we used a *k*-means clustering approach, discriminant analysis of principal components (DAPC), to identify genetic clusters with the R package adegenet (Jombart, 2008). We used the *find.clusters* function to determine the number of genetic clusters (*k*, from $k = 1$ to $k = 20$ clusters) that minimized the Bayesian information criterion (BIC) followed by the cross-validated *xvalDapc* function to describe those assignments along discriminant axes. For DAPC, missing SNPs were imputed as the mean value of nonmissing data at that locus. Third, we assessed fine-scale, hierarchical population structure in fineRADstructure v. 0.3.2 (Malinsky et al., 2018). The fineRADstructure Bayesian method maximizes the posterior probability of a partitioned genetic similarity ("coancestry") matrix and then derives a tree by sequentially merging inferred populations together (Lawson et al., 2012). We ran five independent fineRADstructure replicates for 3,000,000 iterations each with burnin of 2,000,000 iterations and sampling every 1,000th iteration. Convergence for each replicate was assessed by viewing parameter traces and confirming posterior effective sample sizes of at least 300 in Tracer v. 1.7 (Rambaut et al., 2018; Figure S1). We retained the tree corresponding to the highest posterior probability across all replicates.

We calculated pairwise F_{ST} values as an average across all loci for individuals grouped by population assignment and tributary using the Weir and Cockerham (1984) method as executed in the R package hierfstat (Goudet, 2005). We also calculated pairwise F_{ST} between collection sites within the Collins River. We estimated heterozygosity rate as the proportion of heterozygous biallelic sites with VCFtools.

To test for patterns of isolation-by-distance, we used a Mantel test for correlations between Prevosti's genetic distance (Prevosti et al., 1975) and along-waterway distances between collection sites. Genetic distances were calculated in R with the poppr package (Kamvar et al., 2014) and a Mantel test with 10,000 permutations was performed with the ade4 package (Dray & Dufour, 2007).

To assess the sensitivity of population structure results to data assembly parameters, we replicated our fastStructure, DAPC, and F_{ST} analyses for the five population structure subbranches that allowed for increasing amounts of missing data.

2.4 | Phylogeny and divergence time estimates

We inferred a maximum likelihood phylogeny of *Etheostoma akatulo* and outgroup specimens using IQTree v. 1.6.12 (Nguyen et al., 2015) with a GTR +Gamma model and 1,000 ultrafast bootstrap replicates (Minh et al., 2013). To improve likelihood searching given short sequence data, we lowered nearest neighbour interchange perturbation strength (-pers) from 0.5 to 0.2 and increased the number of unsuccessful iterations required to stop search (-nstop) from 100 to 500. We used the R package ggtree to visualize phylogenies (Yu et al., 2017).

Divergence times among the lineages of *Etheostoma akatulo* were estimated in the context of outgroup divergence times using the multispecies coalescent module SNAPP v. 1.4.2 for BEAST v. 2.5.2 (Bouckaert et al., 2019; Bryant et al., 2012). Following the methods and scripts of Stange et al. (2018), we fixed mutation rate parameters at 1.0 and assumed a symmetric substitution model with equal frequencies, a strict molecular clock across all branches, and a pure-birth Yule branching process. Based on results of the maximum likelihood phylogenetic analysis, we assigned *Etheostoma akatulo* specimens from the Collins River to a "west" taxon and specimens from the Cane Creek and upper Caney Fork to an "east" taxon, allowing us to estimate the age of the node in the phylogeny that represents the most recent common ancestor of all extant *E. akatulo* populations. Inferences from multiple nuclear genes by Near and Keck (2013) were used to constrain and initialize the phylogeny and node ages of the darter clade *Doration* that includes *Etheostoma akatulo*, *E. stigmaeum*, *E. jessiae*, and *E. meadiae* as well as the outgroup *E. chlorosoma*. The 95% highest posterior density intervals inferred in that study were used as the upper and lower boundaries of uniform priors for our node ages: 10.63 to 17.66 Ma for the common ancestor of *E. chlorosoma* and *Doration*; 4.28 to 8.65 Ma for the common ancestor of *Doration*; and 2.02 to 4.87 Ma for the common ancestor of *E. stigmaeum*, *E. jessiae*, and *E. meadiae*. We combined the output of five independent SNAPP replicates that each ran for

1,000,000 iterations with a burnin of 500,000 iterations and sampling every 1,000th iteration. Convergence for each replicate was assessed by viewing parameter traces and confirming posterior effective sample sizes of at least 400 in Tracer (Figure S2).

2.5 | Demography

To estimate *Etheostoma akatulo* effective population sizes and migration rates, we assessed four demographic models with fastsimcoal2 (Excoffier et al., 2013). Following the results of our population structure and phylogenetic analyses, all models specified one "west" population (Collins River) and one "east" population (Rocky River, Cane Creek, and upper Caney Fork), the two of which diverged from a putative ancestral parent population *T* generations in the past. Given a generation time of two years (Robinson et al., 2013; Simmons et al., 2008), we fixed *T* as one half the mean parent node age estimated in our divergence time analyses. With a fixed divergence time, we ignored monomorphic sites in fastsimcoal2 (--removeZeroSFS), and therefore did not need to specify or estimate mutation rate. Effective population size parameters were initialized with minimum-bounded log-uniform distributions spanning 5 to 5,000 diploid individuals, and unnormalized migration rate priors were initialized with minimum-bounded log-uniform distributions spanning .001 to 10. Following the demographic conclusions for this species based on microsatellite analyses (Robinson et al., 2013), all our models assumed zero population growth. Each of the four models differed in terms of migration rate parameters: (i) no migration rate; (ii) two continuous migration rate parameters, one from west to east and the other from east to west; (iii) one continuous migration rate parameter, from west to east; and (iv) one continuous migration rate parameter, from east to west (see Figure S3 for visual model summaries).

We ran 1,000 replicates of each model (1,000,000 iterations and 40 ECM cycles per replicate) and chose the initial best replicate with the maximum estimated likelihood for each model. We compared models using AIC values calculated from the maximum of 1,000 likelihood estimates after fixing the initial best replicate parameters (Excoffier et al., 2013). To assess model fit given underlying uncertainty in our data set, we generated 1,000 bootstrap SFS replicates by randomly drawing SNP sites with replacement using SFS-scripts (https://github.com/marqueda/SFS-scripts). Sites were drawn from the demography branch data set that was already filtered to reduce SNP linkage and projected to remove missing data. We ran one replicate of the AIC-selected models (1,000,000 iterations and 40 ECM cycles) for each of the 1,000 bootstrap SFS replicates. Parameter starting values were set to the initial best replicate parameters for the given model. The result was a distribution of 1,000 estimates for each parameter from each AIC-selected model.

2.6 | Morphological divergence

We investigated if morphological traits commonly used to identify and delimit species of North American freshwater fishes exhibited variation consistent with population structure and phylogenetic relationships in *Etheostoma akatulo*. The meristic data and qualitative scoring of nape, cheek, opercle and breast squamation from 207 specimens used in the *Etheostoma akatulo* species description were supplied by Dr. Steven Layman (Layman & Mayden, 2009; Figure 1). This data set included 109 specimens from the Collins River, 49 specimens from the Rocky River, and 36 specimens from mixed Cane Creek and Caney Fork sites. Critically, this data set also included 13 specimens from the Calfkiller River, the translocation target site from which this species was extirpated.

We used 10 characters in our analyses: (i) number of lateral line scales (LL), (ii) number of transverse scale rows (Trans), (iii) number of scale rows around the caudal peduncle (CD), (iv) number of first dorsal fin spines (D1), (v) number of second dorsal fin rays (D2), (vi) number of anal fin rays (A2), (vii) number of left pectoral fin rays (P1), (viii) percentage of the nape scaled (Nape), (ix) percentage of the cheek scaled (Cheek), and (x) percentage of the opercle scaled (Opercle). In the data set provided, Cane Creek and Caney Fork records were combined and all except two of the squamation percentages were rounded to the nearest multiple of 10.

Phenotypic clustering among extant and extirpated populations was explored through a scaled principal component analysis of the 10 morphological characters using the *prcomp* function in R (R Core Team, 2020). We calculated pairwise group mean Mahalanobis distances in full principal component space using the *pairwise.mahalanobis* function from the HDMD package (McFerrin, 2013). We tested the robustness of phenotypic assignment into either a "west" (Collins River) or "east" (Rocky River, Cane Creek, or upper Caney Fork) population with cross-validated, leave-one-out, linear discriminant analysis (LDA) executed with the R package MASS (Venables & Ripley, 2013). We then used a west-east LDA model fit to all data from extant populations to predict the phenotypic assignment of historical specimens from the Calfkiller River.

3 | RESULTS

3.1 | Population structure

The fastStructure analysis indicated that $k = 2$ populations was the model complexity that best explained genetic structure and maximized the approximate marginal likelihood of genetic variation in the ddRAD SNPs (Figure S4A). DAPC also indicated a best-fit value of $k = 2$ clusters described by a single linear discriminant axis (Figure S4B). Given $k = 2$, both fastStructure and DAPC analyses assigned all specimens from the Collins River, in the western Caney Fork, to one population and assigned all specimens from the Rocky River, Cane Creek, and upper Caney Fork to a second, eastern population (Figure 2, and see Figure S5 for admixture proportions at other *k* values). Results from fineRADstructure indicated 11 coancestry groups (Figure S6). This analysis suggested that fine-scale population structure within each tributary was nested within a broader division between eastern and western coancestry groups, and there

was no single coancestry group containing all and only Rocky River individuals at any level of the hierarchy. Coancestry estimates from fineRADstructure revealed that one Collins River individual exhibited elevated coancestry with three Cane Creek individuals and one Rocky River individual (Figure S6).

Population structure results were robust to different levels of missing data. For fastStructure, the population structure assembly subbranches with differing locus retention thresholds all selected k = 2 populations as the number of model components that best explained genetic structure, and all except one data set indicated that k = 2 maximized marginal likelihoods despite a large decrease in comparative resolution as missing data increased (the 70% individuals-per-locus subbranch narrowly selected k = 3 in this criterion; Figure S7A). For DAPC, all subbranches supported k = 2 genetic clusters (Figure S7B). For both analyses, all subbranches sorted Collins River individuals into one population and all Rocky River, Cane Creek, and upper Caney Fork individuals into the other at k = 2.

Overall F_{ST} between the western and eastern population was 0.271. When calculated among tributaries, pairwise F_{ST} values highlighted moderate genetic differentiation between Collins River individuals and both Cane Creek and Caney Fork individuals (Table 2). Individuals from the Rocky River were more genetically differentiated from individuals in the other eastern tributaries, and less differentiated from Collins River individuals. Pairwise F_{ST} estimates among the four Collins River collection sites ranged from 0.02 to negative values, suggesting little to no genetic differentiation. Replicate F_{ST} estimates from subbranches with different amounts of missing data showed only minor quantitative differences (Table S1).

The mean estimated heterozygosity rate was 56% higher in Collins River specimens than across specimens from the eastern tributaries (west: 0.165; east: 0.106). Mean heterozygosity rate was highest in the Collins River and lowest in the upper Caney Fork (Collins River: 0.165; Rocky River: 0.112; Cane Creek: 0.105; Caney Fork: 0.096).

There was a statistically significant correlation between genetic distance and collection site distance when all collection sites from both western and eastern tributaries were included (Mantel test, simulated p <.001; max site distance: 109 km). There was also a significant isolation-by-distance correlation across eastern sites (p <.001; max site distance: 38 km). An isolation-by-distance test among Cane Creek and Caney Fork sites had a low but not significant p-value (p =.07; max site distance: 13 km) and a test across Collins River sites was also not significant (p =.46; max site distance: 32 km).

3.2 | Phylogeny and divergence time

The maximum-likelihood phylogeny resolved three *Etheostoma akatulo* clades: the Collins River, the Rocky River, and Cane Creek-Caney Fork (Figure 2, and see Figure S8 for outgroups and additional tip information). The Collins River and Cane Creek-Caney Fork clades were well-supported with bootstrap values of 1.00, whereas

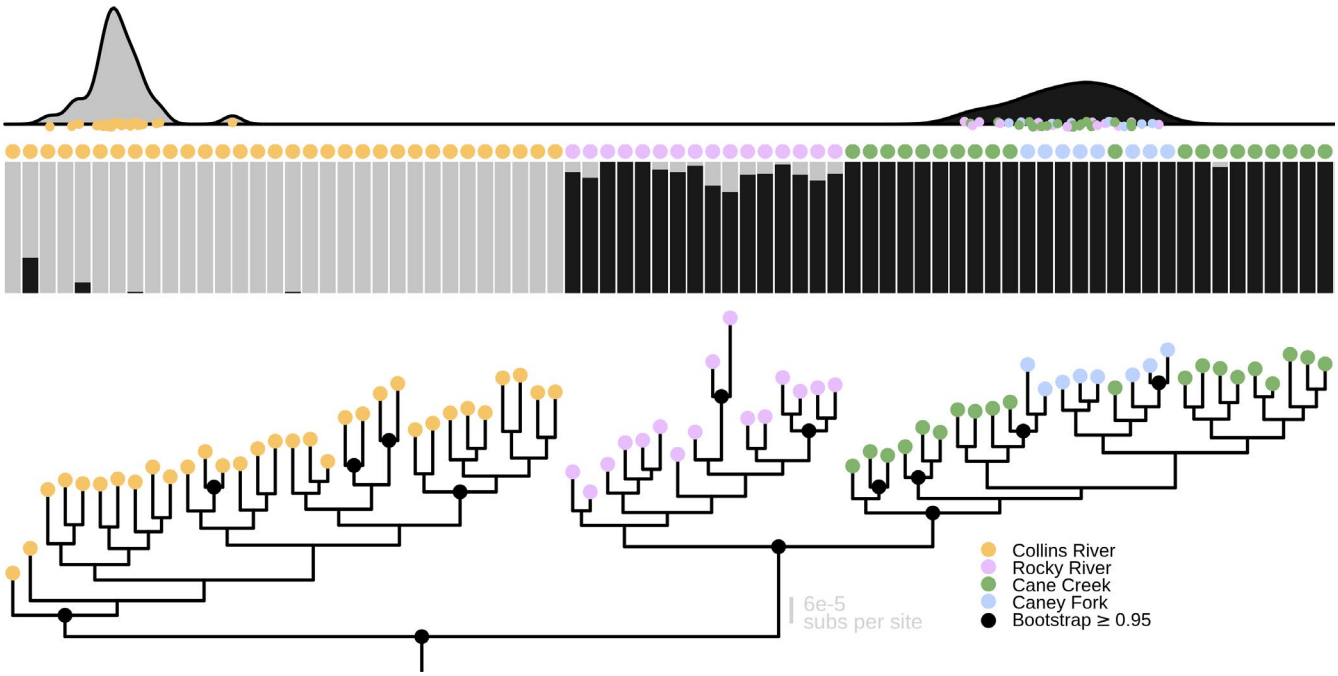

FIGURE 2 Range-wide genetic analysis of *Etheostoma akatulo* reveals population structure and phylogenetic divergence between western (Collins River) and eastern (Rocky River, Cane Creek, and upper Caney Fork) populations. Top: Discriminant analysis of principal components (DAPC) assigned individuals to k = 2 genetic clusters along a single linear discriminant axis with 100% cross-validation accuracy. Middle: Posterior mean admixture proportions from fastStructure at the favoured model complexity of k = 2 populations. Bottom: Maximum-likelihood phylogeny from IQTree resolved a well-supported split between western and eastern lineages

monophyly of the Rocky River lineage had bootstrap support of 0.81. Divergence time analysis in SNAPP estimated a posterior mean age of 146,500 years and a 95% highest posterior density interval ranging from 106,600 years to 188,100 years for the common ancestor of *E. akatulo*.

3.3 | Demography

Given a divergence time of 73,250 generations ago (146,500 years/two years per generation) between the western and eastern populations of *Etheostoma akatulo*, the model with the lowest AIC score was the most complicated model that included migration in both directions, followed by the model that allowed for migration from east to west only (ΔAIC =341; and see Figure S3 for a visual summary of demographic results). AIC scores also strongly disfavoured migration from west to east only (ΔAIC =601) and the model with no migration showed a remarkably lower estimated likelihood (ΔAIC =48,216). Bootstrapped parameter estimates from the top two preferred models suggested that both populations had effective population sizes of dozens to hundreds of individuals, and both models indicated that the western population was approximately twice as large as the eastern population (Table 3, and see Figures S9–S10 for full parameter distributions). Per-capita, per-generation immigration rate estimates for the western population were more than twice as high when only east to west migration was allowed (Table 3). However, the model that allowed for migration in both directions still estimated near-zero immigration rates for the eastern population (Table 3).

TABLE 2 Pairwise F_{ST} values for populations of *Etheostoma akatulo*. The Collins River population comprises the western portion of the Caney Fork River system and all other tributaries comprise the eastern Caney Fork River system

	Rocky River	Cane Creek	Caney Fork
Collins River	0.20	0.29	0.27
Rocky River		0.11	0.09
Cane Creek			0.05

3.4 | Morphological divergence

The summary of *Etheostoma akatulo* variation in meristic and squamation traits across all tributaries is shown in Table 4 (see Table S2 for frequency tables). Specimens from the extirpated Calfkiller River had the highest mean number of lateral line scales (LL), the lowest percentages of nape and opercle covered with scales, and the fewest average number of left pectoral fin rays (P1).

The first principal component of morphology data highlighted divergence between the western specimens from the Collins River and the eastern specimens from the Rocky River, Cane Creek, Caney Fork, and Calfkiller River (Figure 3a). The traits loading most heavily on this axis were the number of scale rows around the caudal peduncle (CD), which was higher in the eastern tributaries, and percentages of nape and cheek covered in scales, which were higher in the Collins River to the west (Table 4). Mahalanobis distances revealed the Collins River and extirpated Calfkiller River populations had the greatest phenotypic disparity among contrasts within the species (Table 5). Overall contrasts within the extant eastern population exhibited less morphological disparity than contrasts involving the Collins River (Table 5). In cross-validated LDA, only 5 of 109 (4.6%) of specimens from the Collins River were assigned to the eastern phenotype and seven out of 85 (8.2%) of specimens from the Rocky River, Cane Creek, and Caney Fork were assigned to the western phenotype. Given the full model fit to all specimens from extant tributaries, 12 of 13 (92%) Calfkiller River specimens were assigned to the eastern phenotype (Figure 3b). The one Calfkiller River specimen assigned to the western phenotype maintained a 34% posterior probability of classification with the eastern phenotype.

4 | DISCUSSION

The microsatellite analyses of Robinson et al. (2013) suggested three genetically distinct populations of *Etheostoma akatulo*: one in the Collins River, one in the Rocky River, and one comprising the Cane Creek and Caney Fork. Our results suggest that genetic and phenotypic variation reflects a deeper divergence between western (Collins River) and eastern (Rocky River, Cane Creek, and upper

TABLE 3 Demographic estimates for western and eastern *Etheostoma akatulo* populations under the two models with the lowest AIC scores. Parameters values are the median, with 95% highest density intervals in parentheses, across 1,000 model replicates from bootstrapped data sets initiated with maximum likelihood parameters from the full empirical data set. Effective population sizes (N_e) indicate the number of diploid individuals rounded to the nearest whole number and migration parameters indicate per-capita, per-generation immigration rates

Model	No. Params	Log_{10} likelihood	AIC	N_e West	N_e East	Proportion N_e West/East	Migration East->West	Migration West->East
Migration Both	5	−15,272	70,340	256 (138, 405)	101 (59, 161)	2.58 (2.18, 3.05)	4.1e−3 (2.1e−3, 6.4e−3)	5.8e−4 (3.3e−4, 9.8e−4)
Migration East->West	4	−15,346	70,681	108 (40, 199)	53 (21, 102)	2.04 (1.69, 2.41)	0.01 (3.5e−3, 0.02)	0

Caney Fork) populations. Population structure analyses (Figure 2) and F_{ST} values (Table 2) demonstrated that genetic structure among Rocky River, Cane Creek, and Caney Fork individuals is at a lower magnitude than observed in comparisons between western and eastern populations. The large number of loci from ddRAD sequencing should increase the ability to parse high-resolution population

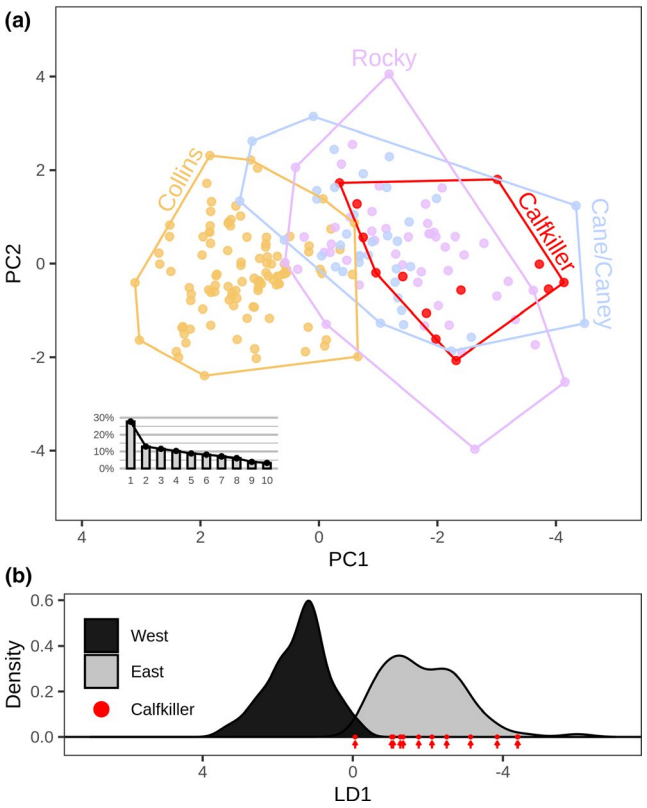

FIGURE 3 Morphological traits from *Etheostoma akatulo* museum specimens reveal phenotypic divergence between western (Collins River) and eastern (Rocky River, Cane Creek, and upper Caney Fork) tributaries, with records from the extirpated Calfkiller River clustering with records from eastern tributaries. (a) The first principal component (PC1) of meristic and squamation traits highlights variation between western and eastern phenotypes. Inset bars show the percentage of variation explained by each PC. (b) All except one Calfkiller River specimens were classified with eastern phenotypes in a linear discriminant model of morphological data from the extant tributaries. Both x-axes are reversed for aesthetic purposes

TABLE 4 Summary of mean meristic and squamation trait values among populations of *Etheostoma akatulo*. Standard deviations are given in parentheses. The bottom row shows the scaled loadings of each trait on the first principal component (PC1) axis across all traits, which highlights divergence between western (Collins River) and eastern (Rocky River, Cane Creek, upper Caney Fork, and Calfkiller River) populations. See main text for trait definitions

Tributary	LL	Trans	CD	D1	D2	A2	P1	Nape	Cheek	Opercle
Collins River	45.4 (1.9)	11.7 (0.7)	14.6 (0.8)	11.3 (0.6)	11.2 (0.5)	8.0 (0.4)	14.6 (0.5)	96.9 (6.9)	96.8 (5.2)	97.5 (4.7)
Rocky River	44.6 (2.2)	12.7 (0.8)	16.1 (0.8)	11.9 (0.5)	10.8 (0.5)	7.9 (0.5)	14.6 (0.6)	75.8 (10.6)	83.9 (16.3)	96.5 (7.5)
Cane Creek- Caney Fork	45.6 (1.7)	12.5 (0.8)	15.8 (0.7)	11.5 (0.7)	11.1 (0.6)	8.1 (0.5)	14.6 (0.5)	72.2 (13.5)	84.2 (12.5)	96.9 (6.7)
Calfkiller River	47.3 (2.5)	12.6 (0.8)	15.9 (0.8)	11.7 (0.6)	10.7 (0.6)	8.0 (0.7)	13.8 (0.6)	66.9 (14.4)	86.2 (10.4)	84.6 (9.7)
PC1 loading	−0.01	−0.39	−0.44	−0.29	0.29	0.16	0.16	0.48	0.41	0.21

TABLE 5 Mahalanobis distances between *Etheostoma akatulo* phenotypes. Values are calculated between group means across all principal components

	Rocky River	Cane Creek-Caney Fork	Calfkiller River
Collins River	2.86	2.44	3.87
Rocky River		1.13	2.67
Cane Creek-Caney Fork			2.56

structure (Sunde et al., 2020), consistent with the results of our fineRADstructure analysis which identified subpopulation structure within each tributary (Figure S6). However, this analysis did not identify a well-supported Rocky River subpopulation and further showed that population structure is hierarchically nested within a broader genetic divergence between overall western and eastern populations. Our phylogenetic analysis paralleled the population structure results, showing a western lineage of all Collins River individuals along with a well-supported eastern clade containing a monophyletic lineage from the Cane Creek and upper Caney Fork, as well as a potential clade from the Rocky River (Figure 2).

Molecular divergence time analysis estimated that the western and eastern *Etheostoma akatulo* lineages shared a common ancestor 146,500 years ago. Morphological analyses with LDA revealed a marked disparity in meristic and squamation traits that exceeded 90% cross-validated classification accuracy for eastern and western phenotypes. Although a possible monophyletic Rocky River lineage helps explain the three-population result derived from 10 microsatellite loci in Robinson et al. (2013), our analyses of ddRAD data reveal a deeper divergence between western and eastern populations of *Etheostoma akatulo*. Our results highlight the advantages of utilizing ddRAD data, which yielded hundreds of SNPs to render high-resolution population structure and demographic estimates (Lemopoulos et al., 2019) as well as tens of thousands of loci to construct a well-resolved phylogeny (Eaton et al., 2017).

The current conservation strategy for Bluemask Darters involves propagating broodstock from the Collins River, in the western portion of the Caney Fork River system, to repopulate empty Calfkiller River habitats in the eastern portion of the system (Petty et al., 2020; Figure 1). Our results suggest that this strategy will disrupt the genetic, phylogenetic, and phenotypic divergence that signals western and eastern *Etheostoma akatulo* populations as two independent, evolutionarily significant units (ESUs) for the purposes of conservation (Crandall et al., 2000; Moritz, 1999). Our critique of translocation from the Collins River to Calfkiller River is also substantiated by facts about the morphology of Calfkiller River fish themselves. First, historical Calfkiller River phenotypes are on average of 48% more disparate from those of the Collins River than from populations in other eastern tributaries of the Caney Fork River system (Table 5). Second, all but one of the specimens from the Calfkiller River were classified as having eastern phenotypes using LDA (Figure 3b). Although limited by the small number of available phenotypic records, these results constitute the only direct information about extirpated *E. akatulo* from the Calfkiller River. Beyond simply repopulating empty habitats, reintroduction of *E. akatulo* from the Collins River to the eastern portion of the Caney Fork River system not only establishes a possible route of anthropogenically-induced introgression via translocated individuals into the broader eastern population, but also introduces members of a phenotypically divergent population into the Calfkiller River itself.

An alternative translocation strategy of moving individuals from other eastern sites into the Calfkiller River would not threaten these evolutionary boundaries. Although the Rocky River lineage is differentiated from the lineage occupying the Cane Creek and upper Caney Fork, none of our results showed relevant differences in these lineages with respect to the Calfkiller River, which joins the river system between the mouths of Rocky River and Cane Creek (Figure 1). However, our analyses still pointed to demographic and genetic risks of translocating adults from the eastern tributaries to Calfkiller River habitats. Our best demographic models suggest that the eastern *Etheostoma akatulo* population is small in absolute size and approximately half the size, or less, than the western population (Table 3). Further, the genetic data showed that all eastern tributaries harboured lower heterozygosity than the western population. While the precision of our population size estimates is limited by compound uncertainty in age estimates for the common ancestor of *E. akatulo*, our consistent estimate of a larger western population parallels the genetic results of Robinson et al. (2013) as well as field reports (Layman & Mayden, 2009; Layman et al., 1993; Jeffrey W. Simmons, unpublished data), all of which indicate that populations are more abundant in the Collins River and increasingly limited in both abundance and habitat across the eastern tributaries.

In this case, the competing aims of maintaining evolutionary divergence, managing demographic resources, and encouraging allelic diversity creates a translocation conundrum. On the one hand, the translocation of adults from the Collins River is suboptimal because such movement could collapse the population structure, phylogenetic, and phenotypic boundaries between western and eastern populations of *Etheostoma akatulo*. Indeed, because the eastern population is so restricted in size and genetic diversity—and because the normal flow of migration appears to be from east to west—translocation from the more diverse and distinct Collins River population carries the additional risk of "swamping" local alleles in the east (Lenormand, 2002; Yeaman, 2015). On the other hand, translocation from the extant eastern populations is suboptimal precisely because the available population size and genetic variation is limited.

One solution to this conundrum could use juvenile fish from the Rocky River as the translocation source for Calfkiller River reintroduction. Rocky River individuals exhibited the highest heterozygosity among the eastern tributaries. Important to this strategy is the fact that many larvae spawned in the Rocky River make their way downstream to unsuitable, impounded habitats in the fluctuation zone of the Great Falls Reservoir each year (Simmons et al., 2008). The recruitment patterns of these juvenile fish are presently unknown and further study is necessary. However, preliminary surveys across multiple years suggests that adult fish do not persist in this reservoir zone and that >90% of fish are young-of-year (Jeffrey W. Simmons, unpublished data). If juveniles in the impounded portion of the Rocky River have a low probability of survival and recruitment, translocating these individuals would allow for eastern recruitment into Calfkiller River sites and sampling from the highest heterozygosity eastern tributary, all without directly drawing down the limited eastern adult breeding population. Note that the possible advantages of this translocation strategy only become clear after our full suite of analyses for not only the population structure, phylogenetic, and phenotypic patterns that identify eastern sites as a

better translocation source, but also the demographic patterns that suggest population limits in that source.

Although the parallel genetic and phenotypic divergence revealed here is consistent with a hypothesis of local adaptation, none of our analyses investigated the functional associations between phenotypic and environmental variation sufficient to actually demonstrate local adaptations important for conservation (Storfer, 1999). Nor did we determine whether western and eastern lineages have established reproductive boundaries that might contribute to the full sedimentation of species differences (Coyne & Orr, 2004). A fruitful avenue for future work is the investigation of how ecological, behavioural, or developmental barriers are maintaining genetic and phenotypic distinctiveness within the Caney Fork River system at such small scales, without clear geographic boundaries, and with apparent migration. Of particular interest is the asymmetric migration suggested by our demographic model comparisons, which can maintain genetic structure despite the movement of organisms between populations (Morrissey & de Kerckhove, 2009; Nosil, 2008). In the case of *E. akatulo*, this asymmetry might be a result of natural dispersal mechanisms; larvae are pelagic and flow downstream, especially from the Rocky River, whereas adults are unlikely to move if suitable habitat is available (Robinson et al., 2013; Simmons, 2004; Simmons et al., 2008).

Beyond practical implications for translocation of this endangered species, our results further highlight the remarkable microgeographic structure of this species. As predicted by Robinson et al. (2013), the phylogeography of *Etheostoma akatulo* revealed here precisely parallels that of the Corrugated Darter (*Etheostoma basilare*) in the same tributaries of the Caney Fork River system. *Etheostoma basilare* has extant eastern clades from the Rocky River, Calfkiller River, and Cane Creek that diverged from a western lineage in the Collins River approximately 5.4 Ma (Hollingsworth & Near, 2009). For both species, unknown barriers between tributaries appear sufficient to structure long-term, stable divergence between lineages at linear scales of less than 100 km. One surprising element of this result is the parallel phylogeographic histories of *E. akatulo* and *E. basilare* in the same river system show different timing. Western and eastern lineages diverged on the scale of hundreds of thousands of years ago in *E. akatulo* but at the magnitude of millions of years ago in *E. basilare*. Both histories, of course, long precede the impounding of the Caney Fork by Great Falls Dam in the early 20th century. While the reservoir may serve as a current barrier to dispersal (Layman & Mayden, 2009), it cannot explain divergence over hundreds of thousands of years in the case of *E. akatulo* or millions of years in the case of *E. basilare*. For *Etheostoma akatulo*, this old divergence has established a new obstacle to the coherent conservation of the species.

These patterns of freshwater population structure and diversification extend far beyond the Caney Fork River system. In North American darters alone, some cases of genetic subdivision are associated with human impacts such as dams and habitat degradation (Beneteau et al., 2009; Blanton et al., 2019; Fluker et al., 2019; Haponski et al., 2007; Sterling et al., 2012). In other cases, population structure has emerged from isolation-by-distance or drainage discontinuity (Argentina et al., 2018; Euclide & Marsden, 2018). Darter ecology and reproductive behaviour may also play a key role in maintaining population structure, as these fishes can have strong preferences for specific microhabitats (Alexander & Phillips, 2012) and sexual traits (Williams & Mendelson, 2011). Not all darters exhibit a strong signal of population structure, suggesting that the potential for diversification is a result of particular combinations of life historical, anthropogenic, and biogeographic conditions (Camak & Piller, 2018; Washburn et al., 2020). Here, the picture of freshwater fish diversification becomes local, complex, and contingent rather than general, unimodal, and singular (Mayden, 1988; Near & Keck, 2005). Bringing such a picture into focus will require a glimpse of the hidden boundaries which structure populations at scales far smaller than continents, glaciers, or plateaus.

ACKNOWLEDGEMENTS

We thank Daniel MacGuigan, Daemin Kim, Dylan Wainwright, Richard Harrington, and Ava Ghezelayagh for their support and discussions about laboratory work, scripts, and analyses. We also thank Dr. Cynthia Riginos and three anonymous reviewers for comments that refined the logic of our manuscript. We are indebted to Steven Layman for providing the raw meristic data from his systematic studies of *Etheostoma akatulo* and other closely related species. *Etheostoma akatulo* tissue samples were collected by Jeffrey W. Simmons under permit from the USFWS (#TE117405-3). This work was supported by a National Science Foundation Graduate Research Fellowship for Liam U. Taylor (#DGE1752134), the Peabody Museum Bingham Oceanographic Fund, and the Tennessee Valley Authority.

AUTHOR CONTRIBUTIONS

Liam U. Taylor collected genetic data, performed analyses, and wrote the first draft of the manuscript. Jeffrey W. Simmons conceived of the project, collected tissue samples, raised research questions, created maps, and edited the manuscript. Edgar Benavides collected genetic data, supported analyses, and edited the manuscript. Thomas J. Near conceived of the project, supported analyses, and contributed to the writing of the first draft.

DATA AVAILABILITY STATEMENT

Data and scripts for this study, including all raw sequence data, metadata, assembly, analysis, and visualization files have been made available at Dryad (https://doi.org/10.5061/dryad.3bk3j9khp). Sequence data is also available through the NCBI Sequence Read Archive (BioProject PRJNA713303).

ORCID

Liam U. Taylor https://orcid.org/0000-0001-8454-1856

REFERENCES

Alexander, M. L., & Phillips, C. T. (2012). Habitats used by the endangered Fountain Darter (*Etheostoma fonticola*) in the San Marcos River, Hays County. Texas. *the Southwestern Naturalist*, *57*(4), 449–452. https://doi.org/10.1894/0038-4909-57.4.449

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Andreasen, L., & Springer, C. (2001). Hatcheries promote fish recovery. *Endangered Species Update*, *18*(3), S32–S33.

Argentina, J. E., Angermeier, P. L., Hallerman, E. M., & Welsh, S. A. (2018). Spatial extent of analysis influences observed patterns of population genetic structure in a widespread darter species (Percidae). *Freshwater Biology*, *63*(10), 1185–1198. https://doi.org/10.1111/fwb.13106

Beneteau, C. L., Mandrak, N. E., & Heath, D. D. (2009). The effects of river barriers and range expansion of the population genetic structure and stability in Greenside Darter (*Etheostoma blennioides*) populations. *Conservation Genetics*, *10*(2), 477–487. https://doi.org/10.1007/s10592-008-9627-9

Blanton, R. E., Cashner, M. F., Thomas, M. R., Brandt, S. L., & Floyd, M. A. (2019). Increased habitat fragmentation leads to isolation among and low genetic diversity within populations of the imperiled Kentucky Arrow Darter (*Etheostoma sagitta spilotum*). *Conservation Genetics*, *20*(5), 1009–1022. https://doi.org/10.1007/s10592-019-01188-y

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Bouzat, J. L., Johnson, J. A., Toepfer, J. E., Simpson, S. A., Esker, T. L., & Westemeier, R. L. (2009). Beyond the beneficial effects of translocations as an effective tool for the genetic restoration of isolated populations. *Conservation Genetics*, *10*(1), 191–201. https://doi.org/10.1007/s10592-008-9547-8

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Camak, D. T., & Piller, K. R. (2018). Going with the flow: Testing the role of habitat isolation among three ecologically divergent darter species. *Copeia*, *106*(2), 375–387. https://doi.org/10.1643/CG-17-623

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Conant, S. (1988). Saving endangered species by translocation: Are we tinkering with evolution? *BioScience*, *38*(4), 254–257. https://doi.org/10.2307/1310848

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Oxford University Press.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, *15*(7), 290–295. https://doi.org/10.1016/S0169-5347(00)01876-0

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Doll, J. C., Etchison, L., & Owensby, D. (2020). Population estimate of the state and federally threatened Spotfin Chub using underwater observations. *North American Journal of Fisheries Management*, *40*(2), 342–353. https://doi.org/10.1002/nafm.10414

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

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eaton, D. A. R., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, *66*(3), 399–412. https://doi.org/10.1093/sysbio/syw092

Euclide, P., & Marsden, J. E. (2018). Role of drainage and barriers in the genetic structuring of a tessellated darter population. *Conservation Genetics*, *19*(6), 1379–1392. https://doi.org/10.1007/s10592-018-1107-2

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Fluker, B., Jones, K., & Kuhajda, B. (2019). Genetic structure and diversity of the Blueface Darter *Etheostoma cyanoprosopum*, a microendemic freshwater fish in the southeastern USA. *Endangered Species Research*, *40*, 133–147. https://doi.org/10.3354/esr00986

George, A. L., Kuhajda, B. R., Williams, J. D., Cantrell, M. A., Rakes, P. L., & Shute, J. R. (2009). Guidelines for propagation and translocation for freshwater fish conservation. *Fisheries*, *34*(11), 529–545. https://doi.org/10.1577/1548-8446-34.11.529

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Graham, C. F., Boreham, D. R., Manzon, R. G., Stott, W., Wilson, J. Y., & Somers, C. M. (2020). How "simple" methodological decisions affect interpretation of population structure based on reduced representation library DNA sequencing: A case study using the lake whitefish. *PLoS One*, *15*(1), e0226608. https://doi.org/10.1371/journal.pone.0226608

Haponski, A. E., Marth, T. A., & Stepien, C. A. (2007). Genetic divergence across a low-head dam: A preliminary analysis using logperch and greenside darters. *Journal of Great Lakes Research*, *33*, 117–126.

Hollingsworth, P. R., & Near, T. J. (2009). Temporal patterns of diversification and microendemism in Eastern Highland endemic Barcheek Darters (Percidae: Etheostomatinae). *Evolution: International Journal of Organic Evolution*, *63*(1), 228–243. https://doi.org/10.1111/j.1558-5646.2008.00531.x

IUCN/SSC (2013). *Guidelines for reintroductions and other conservation translocations. Version 1.0*. IUCN/SSC.

Jamieson, I. G. (2011). Founder effects, inbreeding, and loss of genetic diversity in four avian reintroduction programs. *Conservation Biology*, *25*(1), 115–123. https://doi.org/10.1111/j.1523-1739.2010.01574.x

Jelks, H. L., Walsh, S. J., Burkhead, N. M., Contreras-Balderas, S., Diaz-Pardo, E., Hendrickson, D. A., Lyons, J., Mandrak, N. E., McCormick, F., Nelson, J. S., Platania, S. P., Porter, B. A., Renaud, C. B., Schmitter-Soto, J. J., Taylor, E. B., & Warren, M. L. (2008). Conservation status of imperiled North American freshwater and diadromous fishes. *Fisheries*, *33*(8), 372–407. https://doi.org/10.1577/1548-8446-33.8.372

Johnson, M. S. (2000). Measuring and interpreting genetic structure to minimize the genetic risks of translocations. *Aquaculture Research*, *31*(1), 133–143. https://doi.org/10.1046/j.1365-2109.2000.00396.x

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Keck, B. P., & Near, T. J. (2010). Geographic and temporal aspects of mitochondrial replacement in Nothonotus darters (Teleostei: Percidae: Etheostomatinae). *Evolution*, *64*(5), 1410–1428. https://doi.org/10.1111/j.1558-5646.2009.00901.x

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Laikre, L., Schwartz, M. K., Waples, R. S., & Ryman, N. (2010). Compromising genetic diversity in the wild: Unmonitored large-scale release of plants and animals. *Trends in Ecology & Evolution*, *25*(9), 520–529. https://doi.org/10.1016/j.tree.2010.06.013

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLOS Genetics*, *8*(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Layman, S. R., & Mayden, R. L. (2009). A new species of the darter subgenus *Doration* (Percidae: Etheostoma) from the Caney Fork River system. *Tennessee. Copeia*, *2009*(1), 157–170. https://doi.org/10.1643/CI-07-129

Layman, S. R., Simons, A. W., & Wood, R. M. (1993). Status of the Dirty Darter, Etheostoma olivaceum, and Bluemask Darter, Etheostoma (Doration) sp., with notes on fishes of the Caney Fork River system, Tennessee. *Journal of the Tennessee Academy of Science*, *68*(2), 65–70.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M.-L., Koskiniemi, J., & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120. https://doi.org/10.1002/ece3.4905

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Lundberg, J. G., Kottelat, M., Smith, G. R., Stiassny, M. L. J., & Gill, A. C. (2000). So many fishes, so little time: An overview of recent ichthyological discovery in continental waters. *Annals of the Missouri Botanical Garden*, *87*(1), 26–62. https://doi.org/10.2307/2666207

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Biology*, *37*(4), 329–355. https://doi.org/10.1093/sysbio/37.4.329

McFerrin, L. (2013). *HDMD: Statistical analysis tools for high dimension molecular data (HDMD) (Version 1.2) [R]*. Retrieved from https://CRAN.R-project.org/package=HDMD

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Moritz, C. (1995). Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *349*(1327), 113–118. https://doi.org/10.1098/rstb.1995.0097

Moritz, C. (1999). Conservation units and translocations: Strategies for conserving evolutionary processes. *Hereditas*, *130*(3), 217–228. https://doi.org/10.1111/j.1601-5223.1999.00217.x

Morrissey, M. B., & de Kerckhove, D. T. (2009). The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *The American Naturalist*, *174*(6), 875–889. https://doi.org/10.1086/648311

Near, T. J., Bossu, C. M., Bradburd, G. S., Carlson, R. L., Harrington, R. C., Hollingsworth, P. R., Keck, B. P., & Etnier, D. A. (2011). Phylogeny and temporal diversification of darters (Percidae: Etheostomatinae). *Systematic Biology*, *60*(5), 565–595. https://doi.org/10.1093/sysbio/syr052

Near, T. J., & Keck, B. P. (2005). Dispersal, vicariance, and timing of diversification in Nothonotus darters. *Molecular Ecology*, *14*(11), 3485–3496. https://doi.org/10.1111/j.1365-294X.2005.02671.x

Near, T. J., & Keck, B. P. (2013). Free from mitochondrial DNA: Nuclear genes and the inference of species trees among closely related darter lineages (Teleostei: Percidae: Etheostomatinae). *Molecular Phylogenetics and Evolution*, *66*(3), 868–876. https://doi.org/10.1016/j.ympev.2012.11.009Get

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x

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. https://doi.org/10.1371/journal.pone.0037135

Petty, M. A., Simmons, J. W., Ruble, C., Shute, J. R., & Rakes, P. L. (2020). *Propagation and Reintroduction of Bluemask Darter, Etheostoma akatulo, in the Calfkiller River, White and Putnam Counties, Tennessee. Progress Report for 2019—May 2020*. Submitted to the U.S. Fish and Wildlife Service, Cookeville, Tennessee Field Office, Conservation Fisheries, Inc., and Tennessee Valley Authority.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of high-density genetic maps for Barley and Wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

Prevosti, A., Ocana, J., & Alonso, G. (1975). Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, *45*(6), 231–241.

R Core Team (2020). *R: A language and environment for statistical computing (Version 3.6.3)*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Robinson, J. D., Simmons, J. W., Williams, A. S., & Moyer, G. R. (2013). Population structure and genetic diversity in the endangered Bluemask Darter (*Etheostoma akatulo*). *Conservation Genetics*, *14*(1), 79–92. https://doi.org/10.1007/s10592-012-0427-x

Seddon, P. J. (2010). From reintroduction to assisted colonization: Moving along the conservation translocation spectrum. *Restoration Ecology*, *18*(6), 796–802. https://doi.org/10.1111/j.1526-100X.2010.00724.x

Shute, J. R., Rakes, P. L., & Shute, P. W. (2005). Reintroduction of four imperiled fishes in Abrams Creek. Tennessee. *Southeastern Naturalist*, *4*(1), 93–110.

Simmons, J. W. (2004). Selected life history aspects of the endangered Bluemask Darter Etheostoma (Doration) sp. (MS Thesis). Tennessee Technological University, Cookeville.

Simmons, J. W., Layzer, J. B., & Smith, D. D. (2008). Age-class structure and variability of two populations of the Bluemask Darter *Etheostoma (Doration) sp*. *The American Midland Naturalist*, *160*(2), 300–309.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. *Systematic Biology*, *67*(4), 681–699. https://doi.org/10.1093/sysbio/syy006

Sterling, K. A., Reed, D. H., Noonan, B. P., & Warren, M. L. (2012). Genetic effects of habitat fragmentation and population isolation on Etheostoma raneyi (Percidae). *Conservation Genetics*, *13*(3), 859–872. https://doi.org/10.1007/s10592-012-0335-0

Storfer, A. (1999). Gene flow and endangered species translocations: A topic revisited. *Biological Conservation*, *87*(2), 173–180. https://doi.org/10.1016/S0006-3207(98)00066-4

Sunde, J., Yıldırım, Y., Tibblin, P., & Forsman, A. (2020). Comparing the performance of microsatellites and RADseq in population genetic studies: Analysis of data for Pike (*Esox lucius*) and a synthesis of previous studies. *Frontiers in Genetics*, *11*, 218. https://doi.org/10.3389/fgene.2020.00218

Tallmon, D. A., Luikart, G., & Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, *19*(9), 489–496. https://doi.org/10.1016/j.tree.2004.07.003

U. S. Fish and Wildlife Service (1993). Final decision on relict and blue-mask darters. *Federal Register*, *58*, 68480–68486.

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Washburn, B. A., Cashner, M. F., & Blanton, R. E. (2020). Small fish, large river: Surprisingly minimal genetic structure in a dispersal-limited, habitat specialist fish. *Ecology and Evolution*, *10*(4), 2253–2268. https://doi.org/10.1002/ece3.6064

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., Coates, D. J., Eldridge, M. D. B., Sunnucks, P., Breed, M. F., James, E. A., & Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Williams, T. H., & Mendelson, T. C. (2011). Female preference for male coloration may explain behavioural isolation in sympatric darters. *Animal Behaviour*, *82*(4), 683–689. https://doi.org/10.1016/j.anbehav.2011.06.023

Wright, B. R., Grueber, C. E., Lott, M. J., Belov, K., Johnson, R. N., & Hogg, C. J. (2019). Impact of reduced-representation sequencing protocols on detecting population structure in a threatened marsupial. *Molecular Biology Reports*, *46*(5), 5575–5580. https://doi.org/10.1007/s11033-019-04966-6

Yamamoto, S., Morita, K., Koizumi, I., & Maekawa, K. (2004). Genetic differentiation of White-Spotted Charr (*Salvelinus leucomaenis*) populations after habitat fragmentation: Spatial–temporal changes in gene frequencies. *Conservation Genetics*, *5*(4), 529–538. https://doi.org/10.1023/B:COGE.0000041029.38961.a0

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, *186*(Suppl 1), S74–89. https://doi.org/10.1086/682405

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T.-T.-Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zieritz, A., Hoffman, J. I., Amos, W., & Aldridge, D. C. (2010). Phenotypic plasticity and genetic isolation-by-distance in the freshwater mussel *Unio pictorum* (Mollusca: Unionoida). *Evolutionary Ecology*, *24*(4), 923–938. https://doi.org/10.1007/s10682-009-9350-0

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Taylor LU, Benavides E, Simmons JW, Near TJ. Genomic and phenotypic divergence informs translocation strategies for an endangered freshwater fish. *Mol Ecol*. 2021;00:1–14. https://doi.org/10.1111/mec.15947