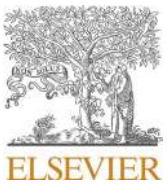
Molecular Phylogenetics and Evolution 178 (2023) 107645



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogeography, hybridization, and species discovery in the *Etheostoma nigrum* complex (Percidae: *Etheostoma*: *Boleosoma*)

Daniel J. MacGuigan [a,*], Oliver D. Orr [b,c], Thomas J. Near [b,c]

[a] Department of Biological Sciences, 109 Cooke Hall, University at Buffalo, Buffalo, NY 14260, USA
[b] Department of Ecology and Evolutionary Biology, 165 Prospect Street, Yale University, New Haven, CT 06520, USA
[c] Yale Peabody Museum of Natural History, 170 Whitney Ave, New Haven, CT 06520, USA

ARTICLE INFO

Keywords:
Phylogeography
Population structure
Allopatric speciation
Secondary contact
Gene flow
ddRAD

ABSTRACT

The history of riverine fish diversification is largely a product of geographic isolation. Physical barriers that reduce or eliminate gene flow between populations facilitate divergence via genetic drift and natural selection, eventually leading to speciation. For freshwater organisms, diversification is often the product of drainage basin rearrangements. In young clades where the history of isolation is the most recent, evolutionary relationships can resemble a tangled web. One especially recalcitrant group of freshwater fishes is the Johnny Darter (*Etheostoma nigrum*) species complex, where traditional taxonomy and molecular phylogenetics indicate a history of gene flow and conflicting inferences of species diversity. Here we assemble a genomic dataset using double digest restriction site associated DNA (ddRAD) sequencing and use phylogenomic and population genetic approaches to investigate the evolutionary history of the complex of species that includes *E. nigrum, E. olmstedi, E. perlongum,* and *E. susanae*. We reveal and validate several evolutionary lineages that we delimit as species, highlighting the need for additional work to formally describe the diversity of the *Etheostoma nigrum* complex. Our analyses also identify gene flow among recently diverged lineages, including one instance involving *E. susanae*, a localized and endangered species. Phylogeographic structure within the *Etheostoma nigrum* species complex coincides with major geologic events, such as parallel divergence in river basins during Pliocene inundation of the Atlantic coastal plain and multiple northward post-glacial colonization routes tracking river basin rearrangements. Our study serves as a nuanced example of how low dispersal rates coupled with geographic isolation among disconnected river systems in eastern North America has produced one of the world's freshwater biodiversity hotspots.

1. Introduction

Southeastern North America is a global hotspot of freshwater fish biodiversity underlain by patterns of allopatric endemism (Abell et al. 2000; Collen et al. 2014; Elkins et al. 2019). Across this region, the dendritic structure of rivers facilitates geographic isolation and diversification. Directional water flow and physical barriers limit dispersal (Selkoe et al. 2015; Davis et al. 2018; Lujan et al. 2020) and local adaptation to flow regimes, depths, or substrates can produce patchy distributions (Ward and Stanford 1983). In addition, river courses shift over long timescales due to erosion, glaciation, sedimentation, and other geologic processes which can isolate previously connected populations (Tight 1903; Bishop 1995). Allopatric speciation is so prevalent in freshwater fishes that it was codified as "Jordan's Law", which states that closely related organisms frequently tend to occupy "neighboring districts" separated by barriers (Jordan 1905; Allen 1907). Despite centuries of study, many North American fish species remain undiscovered and undescribed, creating an obstacle to conservation and study of this incredible biodiversity (Warren et al. 2000; April et al. 2011).

Darters are a clade of approximately 250 species which comprise a large portion of North American endemic fish diversity (Lundberg et al. 2000; Near et al. 2011). This clade exemplifies the predominance of allopatric speciation in freshwater fishes; 93 % of darter sister species pairs are geographically separated (Near et al. 2011; Page and Burr 2011). Allopatry in darters may be explained by a combination of limited dispersal capabilities, habitat specialization, and the fragmented nature of river basins (Etnier and Starnes 1993; Page 1983). Lack of gene flow between allopatric lineages promotes genetic and phenotypic

* Corresponding author.
E-mail address: dmacguig@buffalo.edu (D.J. MacGuigan).

https://doi.org/10.1016/j.ympev.2022.107645
Received 9 March 2022; Received in revised form 29 July 2022; Accepted 11 October 2022
Available online 15 October 2022

divergence which can make identifying new species straightforward. However, some darter species have large, continuous ranges which encompass substantial genetic and phenotypic variance (Page 1983). Delimiting widespread lineages is challenging due to complex histories of range fluctuation, secondary contact, and isolation by distance (Kim et al. 2022; Burbrink and Ruane 2021). Many widespread darter species contain undescribed biodiversity delimited by river basin structure, reflecting the broader allopatric patterns of darter diversification (Ceas and Page 1997; Ray et al. 2006; Piller et al. 2008; MacGuigan et al. 2021; Bossu and Near 2009; Hollingsworth and Near 2009).

One of the most geographically widespread darter lineages is the *Etheostoma nigrum* species complex. This group is phylogenetically nested in the clade *Boleosoma*, which also includes *E. vitreum, E. podostemone,* and *E. longimanum* (Heckman et al. 2009; Near et al. 2011; MacGuigan and Near 2019). The *Etheostoma nigrum* complex contains four currently recognized species: *E. nigrum* Rafinesque, *E. olmstedi* Storer, *E. susanae* (Jordan), and *E. perlongum* (Hubbs and Raney). Two of these species, *E. nigrum* and *E. olmstedi,* have continental-scale ranges with broad overlap in the James, Roanoke, Tar, and Neuse Rivers of Virginia and North Carolina, the Lake Ontario drainage basin in western New York, and the Ottawa River and St. Lawrence River basins in Ontario and Quebec (Fig. 1). Unlike many darter species, *Etheostoma nigrum* and *E. olmstedi* are habitat generalists, occupying freshwater systems with various substrate types and flow regimes (Leidy 1992). The other two species in the complex have small geographic ranges; *E. susanae* is restricted to a small portion of the upper Cumberland River and *E. perlongum* is endemic to Lake Waccamaw in North Carolina (Fig. 1). The only previous range-wide genetic assessment of the *Etheostoma nigrum* complex used two loci with poor phylogenetic resolution and identified several instances of mitochondrial introgression (Heckman et al. 2009). Lack of phylogenetic resolution may be due to recent divergence; the crown age of the *Etheostoma nigrum* species complex is approximately three million years before the present (Near and Keck 2013). Recent diversification coupled with a large geographic distribution provides an opportunity to investigate the early stages of geographic isolation and speciation.

In this study, we examine whether the *Etheostoma nigrum* species complex recapitulates allopatric diversification processes that broadly characterize North American freshwater fish diversity (Wiley and Mayden 1985; Mayden 1988; Near et al. 2003; April et al. 2011; Seehausen and Wagner 2014). We utilize double digest restriction site associated DNA (ddRAD) sequencing to assemble a large genomic dataset encompassing the range of all species in the *Etheostoma nigrum* complex. With this genomic context, we attempt to resolve the evolutionary history and species-level diversity in the clade. We compare analyses of genome-wide ddRAD markers to analyses of a single mtDNA marker commonly employed in phylogeographic studies (Heckman et al. 2009; Keck and Near 2010; Hayes and Piller 2018). In addition, we investigate the extent of hybridization and gene flow between lineages in the *Etheostoma nigrum* species complex. Our analyses identify several cases of peripheral isolation and genetic divergence within this geographically widespread species complex, which correlate with specific paleogeographic events. Although we find evidence of secondary contact and hybridization between closely related lineages, river basin structure limits the extent of introgression. Our results demonstrate that the fragmented geography of freshwater environments can curb the homogenizing effects of gene flow and prevent fusion of widespread lineages following secondary contact. The *Etheostoma nigrum* species complex provides a window into the allopatric diversification process that produced much of North American fish diversity.

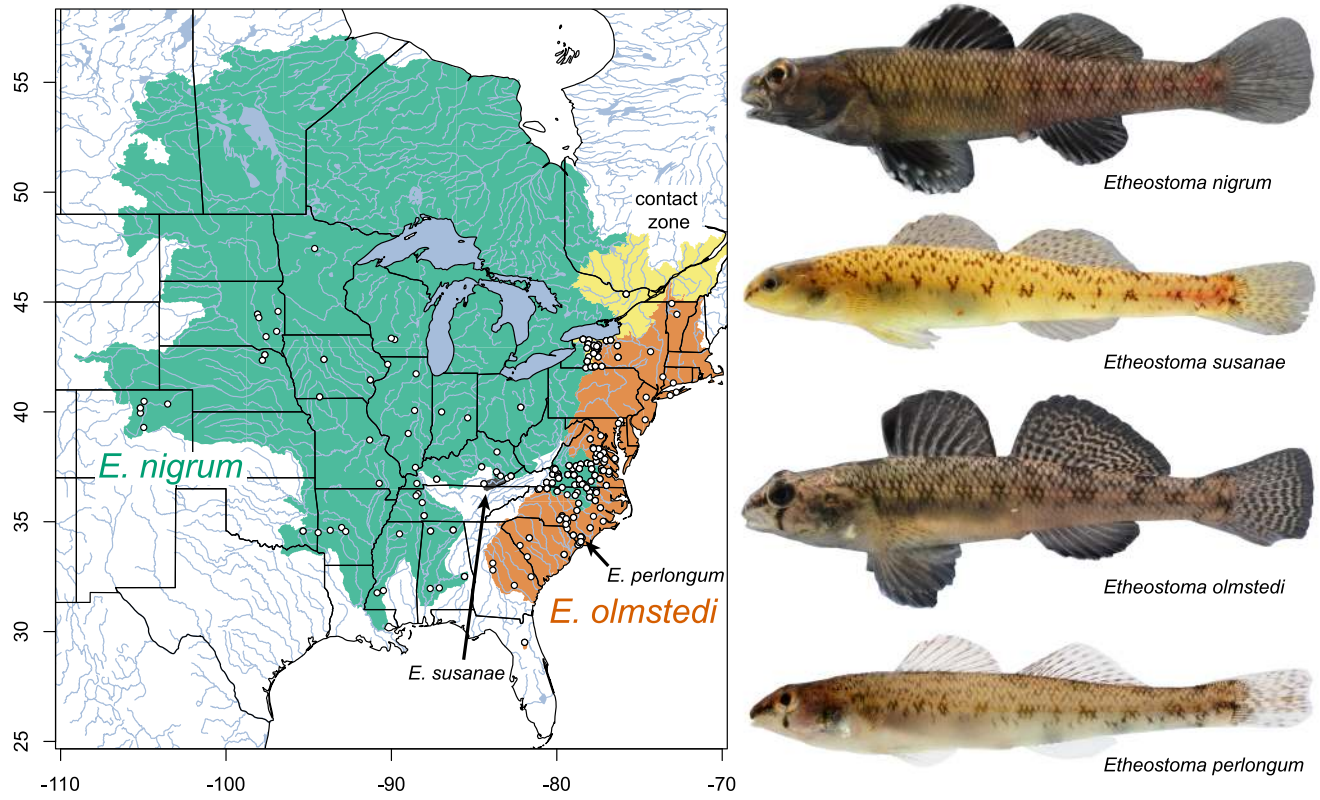


Fig. 1. Range map for current taxonomy of the *Etheostoma nigrum* species complex with sampling localities marked. Yellow indicates the area of range overlap between *E. nigrum* and *E. olmstedi*. *Etheostoma susanae* image courtesy of Richard C. Harrington. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Methods

2.1. Sample collection and sequencing

We collected tissues from all seven recognized species of *Boleosoma*: *Etheostoma vitreum*, *E. podostemone, E. longimanum,* and the *Etheostoma nigrum* species complex consisting of *E. olmstedi, E. perlongum, E. nigrum,* and *E. susanae* (Heckman et al. 2009, Near et al. 2011). In addition, we sampled *E. mariae* as an outgroup taxon. Our final dataset included 449 total specimens: 443 specimens of the *Etheostoma nigrum* species complex from 189 unique localities, 5 specimens of the other *Boleosoma* species, and 1 outgroup specimen (Fig. 1; Table S1). With all appropriate state scientific collection permits and following IACUC protocol number 2018-10681, fish were euthanized in MS-222 before tissue samples were taken from the right pectoral fin. Tissues were preserved in 95 % ethanol at −20 °C. DNA was extracted from tissues using a standard Qiagen DNEasy Blood and Tissue Kit. Prior to library preparation, we cleaned DNA extractions using an ethanol precipitation: 3 M sodium acetate (pH = 5.2) was added equal to 10 % of the total volume of the DNA extraction, followed by 100 % ethanol equal to 2.5 times the total volume of DNA. After mixing, the extractions were incubated for 10 min at −80 °C. Samples were centrifuged for 30 min at 8,000 RCF, the supernatant was discarded, and the DNA pellet was washed with 250 uL of cold 70 % ethanol. Samples were centrifuged again for five minutes at 8,000 RCF and the supernatant was discarded. We allowed the pellet to air dry for ~15 min, then resuspended with the desired amount of DNAse-free water.

We performed a modified version of the double digest restriction site associated DNA (ddRAD) sequencing protocol outlined in Peterson et al. (2012) and Poland et al. (2012). We digested 200 ng of DNA from each sample for 8 h at 37 °C using the restriction enzymes *PstI* and *SbfI*. Samples were visually checked using a 2.5 % agarose gel to ensure complete digestion. We ligated custom adapters using T4 DNA ligase and an incubation period of three hours at 22 °C. The adapters contained a set of 96 unique 8–10 bp barcodes. Barcode sequences were generated using https://www.deenabio.com/services/gbs-adapters with at least two mutational differences between any pair of barcodes to avoid misassignment during demultiplexing. Samples were then pooled and cleaned using a QIAquick PCR purification kit. We performed 12 rounds of PCR, each cycle with 30 s at 98 °C, 30 s at 62 °C, and 30 s at 72 °C. Following the 12 cycles, we held for 10 min at 17 °C. After PCR, sets of 96 samples with unique barcodes were pooled to perform a 300–500 base pair size selection using a Blue Pippen 2 % agarose gel cassette. Size selection was confirmed using an ABI Bioanalyzer High Sensitivity DNA assay. Size selection and fragment analysis were performed at the Yale DNA Analysis Facility on Science Hill. We sequenced each 96-sample genomic library using one lane of Illumina HiSeq 4000 with 100 or 150 base pair single-end reads at the University of Oregon Genomics and Cell Characterization Core Facility. We inspected raw read quality using FastQC (Andrews, 2010).

We used the *Etheostoma perlongum* genome (MacGuigan 2020, NCBI BioProject PRJNA682188) to map and assemble the ddRAD dataset with iPyrad v.0.9.81 (Eaton and Overcast 2020). After demultiplexing the Illumina reads in iPyrad step 1, we trimmed raw reads using CutAdapt (Martin 2011). Reads were trimmed to 100 bp and three rounds of trimming were used to remove adapter contamination and restriction cut sites ("-b TGCAG", "-a TGCAG", and "-a CCGAGATCGGAAGAGC"). Trimmed reads were then used as input for iPyrad steps 2–7 (see Table S2 for assembly parameters).

In addition to ddRAD sequencing, we also amplified the complete protein-coding region of the mitochondrial cytochrome *b* (*cytb*) gene using polymerase chain reaction (PCR) with previously described primers and annealing conditions (Near et al. 2000). Amplified double-stranded DNA was purified using 20 % polyethylene glycol DNA precipitation. Forward and reverse DNA strands were sequenced at the Yale Keck DNA Sequencing Facility using Big Dye Terminator Reaction Kits (Applied Biosystems, Foster City, CA). Sequencing was carried out on an Applied Biosystems 3730xL (Applied Biosystems, Foster City, CA). We trimmed and assembled reads into contiguous sequences using Geneious v.8.0.5 (https://www.geneious.com). We then aligned these sequences using the program MUSCLE v.3.8.425 (Edgar 2004). We checked each alignment of protein-coding sequences by eye for erroneous insertions or deletions.

2.2. Phylogenetic analyses

We performed concatenated maximum likelihood phylogenetic analyses of the ddRAD dataset and inferred a maximum likelihood *cytb* gene tree using IQTree v.2.1.1 (Minh et al. 2020). To determine the impact of missing data on the phylogenies inferred using ddRAD data, we used three datasets with different minimum specimens-*per*-locus thresholds: 270 of 449 specimens (<40 % missing data), 314 of 449 specimens (<30 % missing data), and 360 of 449 specimens (<20 % missing data). We used a GTR + gamma nucleotide substitution model for all ddRAD analyses. For analyses of the *cytb* gene sequences, we partitioned by codon position and performed substation model optimization ("-m MFP + MERGE") with ModelFinder (Kalyaanamoorthy et al. 2017). We ran each tree search until 100 unsuccessful tree search iterations were completed. To assess topological support, we also performed 1,000 ultrafast bootstrap replicates (Hoang et al. 2018).

In addition to concatenated phylogenetic analyses, we inferred a species tree from the ddRAD dataset using the multispecies coalescent model implemented in the SNAPP package for BEAST v.2.5.2 (Bryant et al. 2012; Bouckaert et al., 2019). To achieve reasonable runtimes, we assembled four smaller datasets that sampled one to five individuals from each of the major lineages identified in concatenated phylogenetic analyses (Table S3). One SNAPP dataset contained all major lineages from the *Etheostoma nigrum* clade, while the other three SNAPP datasets focused on the *Etheostoma olmstedi* clade. To maximize the number of SNPs in the SNAPP analyses, we selected individuals with the most loci before cross-sample alignment in iPyrad step 6. Using VCFTools v.0.1.15 (Danecek et al. 2011), we retained only biallelic SNPs ("–max-alleles 2"), excluded singletons, and removed SNPs with missing data ("–max-missing 1.0"). Additionally, we retained only one SNP per 10,000 bp window ("–thin 10000") to avoid using loci in tight physical linkage.

We used the Ruby script provided by Stange et al. (2018) to prepare SNAPP input files and specify a time calibration on the root node. We used a normally distributed root age prior (mean = 13.23 Ma, sd = 1.42 Ma) based on previous time-calibrated darter phylogenies (Near et al. 2011; Near and Keck 2013). We also constrained the monophyly of the *Etheostoma nigrum* species complex. All other priors and parameter settings were left at default values. For the SNAPP analyses focusing on the *Etheostoma olmstedi* clade, we performed four independent SNAPP analyses, each running for 500,000 MCMC generations. Due to issues with convergence for the *Etheostoma nigrum* clade dataset, we performed two independent SNAPP analyses for 1,000,000 MCMC generations. We discarded the first 25 % of each MCMC as burnin and assessed convergence and mixing of parameters using Tracer v.1.7.1 (Bouckaert et al., 2019). Posterior tree distributions were combined using LogCombiner v.2.5.2 and summarized using TreeAnnotator v.2.5.2 as a maximum clade credibility tree with median node age estimates (Bouckaert et al., 2019). All phylogenies were visualized using FigTree v.1.4.4 and the R package ggtree v.2.0.1 (Yu et al. 2017).

2.3. Population structure analyses

We assembled several sets of SNPs for population genetic analyses. Using VCFTools, we retained only biallelic SNPs ("–max-alleles 2") and removed sites that contained more than 10 % missing data ("–max-missing 0.9"). We also removed true singletons and private doubletons identified by the "–singletons" flag in VCFTools since these markers can confound inference of population structure (Linck and Battey 2019).

Finally, we thinned our dataset to include a single SNP per 10,000 base pair window ("–thin 10000") to minimize the effects of physical linkage on our inference of population structure. While the above filtering removed SNPs with large amounts of missing data, some individuals in our dataset still contained large proportions of missing data. Therefore, we used the *impute* function in the R package LEA to replacing missing genotypes with a random value weighted by the observed genotype probabilities (Frichot et al. 2014).

We first performed population structure analyses using all samples from the *Etheostoma nigrum* species complex. However, our analyses of the entire species complex identified many genetic clusters, making interpretation difficult. Therefore, we also performed population structure analyses separately for the three major clades within the *Etheostoma nigrum* complex: the *Etheostoma nigrum* clade (139 specimens, 70 localities), the *Etheostoma maculaticeps* clade (58 specimens, 32 localities), and the *Etheostoma olmstedi* clade (246 specimens, 87 localities). We also assembled a dataset consisting of the two Atlantic coast clades, *E. olmstedi* and *E. maculaticeps*. For each dataset, we performed filtering and missing data imputation as described above.

We deployed three population clustering approaches for each dataset, incrementing the number of genetic clusters (K) from one to a maximum of seventy (for analyses of the entire *E. nigrum* complex) or fifty (for analyses of the subclades). All three methods are considered robust to departures from population genetic model assumptions such as Hardy-Weinberg equilibrium (Frichot et al. 2014). First, we utilized the *snmf* function in the R package LEA v.3.4.0 (Frichot and François 2015). For these analyses, we performed 10 repetitions for each K value. We set the regularization parameter (alpha) to 10, the tolerance to 0.00001, the maximum number of algorithm iterations to 1,000, and the proportion of masked genotypes for cross-entropy validation of K values to 0.05. For each value of K, the replicate with the minimum cross-entropy score was selected for further visualization and comparison with other clustering methods. A cross-validation approach using cross-entropy criteria was performed to determine the optimal value of K.

The second population clustering approach utilized the R package TESS3R v.1.1 (Caye et al. 2016), which also uses sparse non-negative matrix factorization to estimate ancestry coefficients, but includes a spatial regularization parameter (lambda). We used the default value of lambda = 1 for our analyses, which indicates that nearby individuals should have more similar ancestry coefficients, as expected under an isolation-by-distance model (Caye et al. 2018). We performed 10 replicate runs for each K value with a maximum of 1,000 algorithm iterations. All other parameters were set to default values. Cross-validation using cross-entropy criteria was performed to determine the optimal value of K.

The final population clustering approach utilized discriminant analyses of the principal components (DAPC) of genetic variation with the R package adegenet v.2.1.4 (Jombart 2008). First, we assigned individuals to genetic clusters using K-means clustering of genetic principal components as implemented by the *find.clusters* function. We retained all principal component axes for K-means clustering, used 1,000 random starting centroids, and performed 10,000,000 iterations for each value of K. All other parameters were left at default settings. We then used the groups identified by K-means clustering to perform DAPC, implemented by the *dapc* function. The proportion of variance retained in discriminant analysis can affect the stability of group membership probabilities (Jombart and Collins 2015); therefore, we compared discriminant analysis results retaining principal components that accounted for 70 %, 80 %, 90 %, and 95 % of the cumulative variance in dataset. To determine the optimal number of genetic clusters, we calculated BIC scores for each value of K.

To visualize the spatial context of genetic structure, we used the *plot* function in TESS3R to construct interpolated surfaces based on Q-matrices of ancestry coefficients (LEA and TESS3R) or the discriminant analysis assignment proportions (DAPC). We used thin plate spline regression ("interpol = FieldsTpsModel()") with a grid resolution of 1,000. The union of all interpolated surfaces was plotted ("method = 'map.max'") and restricted to species ranges using custom shapefiles.

2.4. Tree-based gene flow estimation

We used the program TreeMix v.1.13 as implemented in iPyrad v.0.9.81 (Pickrell and Pritchard 2012; Eaton and Overcast 2020) to infer the history of gene flow between lineages identified by phylogenetic and population structure analyses. We included one to five samples from every major lineage in the *Etheostoma nigrum* and *Etheostoma olmstedi* clades plus five samples from the *Etheostoma maculaticeps* clade (Table S3). VCFTools was used to assemble a dataset of biallelic SNPs ("–max-alleles 2"), excluding singletons ("—singletons") and SNPs with more than 20 % with missing data ("–max-missing 0.8"). SNPs were also required to be present in at least half of the samples for each lineage. To avoid biases due to physical linkage, iPyrad was used to sample one random SNP per 10,000 base pair window ("–thin 10000"). To account for the heterogeneity of gene flow across the genome, we ran TreeMix using 100 randomly sampled SNP subsets. For each SNP subset, we ran several sets of TreeMix analyses with 0 to 10 allowed migration edges. We summarized the results using a custom R script, retaining migration edges detected in greater than 20 % of SNP subsets for each set of analyses.

3. Results

3.1. Phylogenetic inference

We sequenced 443 specimens of the *E. nigrum* complex (Table S1) sampled from 189 unique localities, with a median of two specimens per locality (minimum = 1, maximum = 10). All individuals in the ddRAD dataset are represented by at least 187,000 reads (mean = 3.7e6 reads, sd = 2.4e6 reads) and 22.6X average read depth per locus. iPyrad identifies 382,413 ddRAD loci mapping to the reference genome which are used to assemble all subsequent datasets. Additionally, we assembled a 1,229 bp alignment of the complete coding region of the mtDNA gene *cytb* with flanking tRNA genes for 757 specimens.

We inferred maximum likelihood (ML) phylogenies using three concatenated ddRAD alignments: 46,564 loci with <40 % missing samples per locus, 28,210 loci with <30 % missing samples per locus, and 12,822 loci with <20 % missing samples per locus. The total percent of missing sites in each concatenated ddRAD alignment is 42.1 %, 38.1 %, and 32.3 %, respectively. Each of the ddRAD phylogenies have a high frequency of nodes with strong node support and resolve *Etheostoma vitreum* as the sister lineage of all other species of *Boleosoma* (Fig. S1). In the ddRAD phylogenies, the two colorful species *E. podostemone* and *E. longimanum* form a clade sister to the *Etheostoma nigrum* species complex (Fig. S1). Theses relationships within *Boleosoma* are also strongly supported in species tree analyses of ddRAD SNPs (Fig. 2). However, in the mtDNA phylogeny, *E. longimanum* is sister to the *E. olmstedi* clade, *E. vitreum* is sister to *E. olmstedi* plus *E. longimanum*, and *E. podostemone* is nested within the *E. nigrum* clade (Fig. S2).

The ddRAD phylogenies resolve three strongly supported clades within the *Etheostoma nigrum* species complex. The *Etheostoma nigrum* clade comprises most populations currently recognized as *E. nigrum*, excluding populations in rivers on the Atlantic Seaboard in North Carolina and Virginia (Fig. 2, Fig. S1). The *Etheostoma maculaticeps* clade comprises all populations currently identified as *E. olmstedi* south of the Neuse River basin, as well as *E. perlongum* from Lake Waccamaw, North Carolina (Fig. 2, Fig. S2). The *Etheostoma olmstedi* clade comprises *E. olmstedi* populations north of the Cape Fear River basin plus all populations in Virginia and North Carolina currently recognized as *E. nigrum* (Fig. 2, Fig. S2). While the *Etheostoma olmstedi* and *Etheostoma nigrum* clades are sister lineages in the ddRAD phylogenies (Fig. 2, Fig. S1), the mtDNA phylogeny resolves the *Etheostoma maculaticeps* clade sister to the *Etheostoma nigrum* clade (Fig. S2).

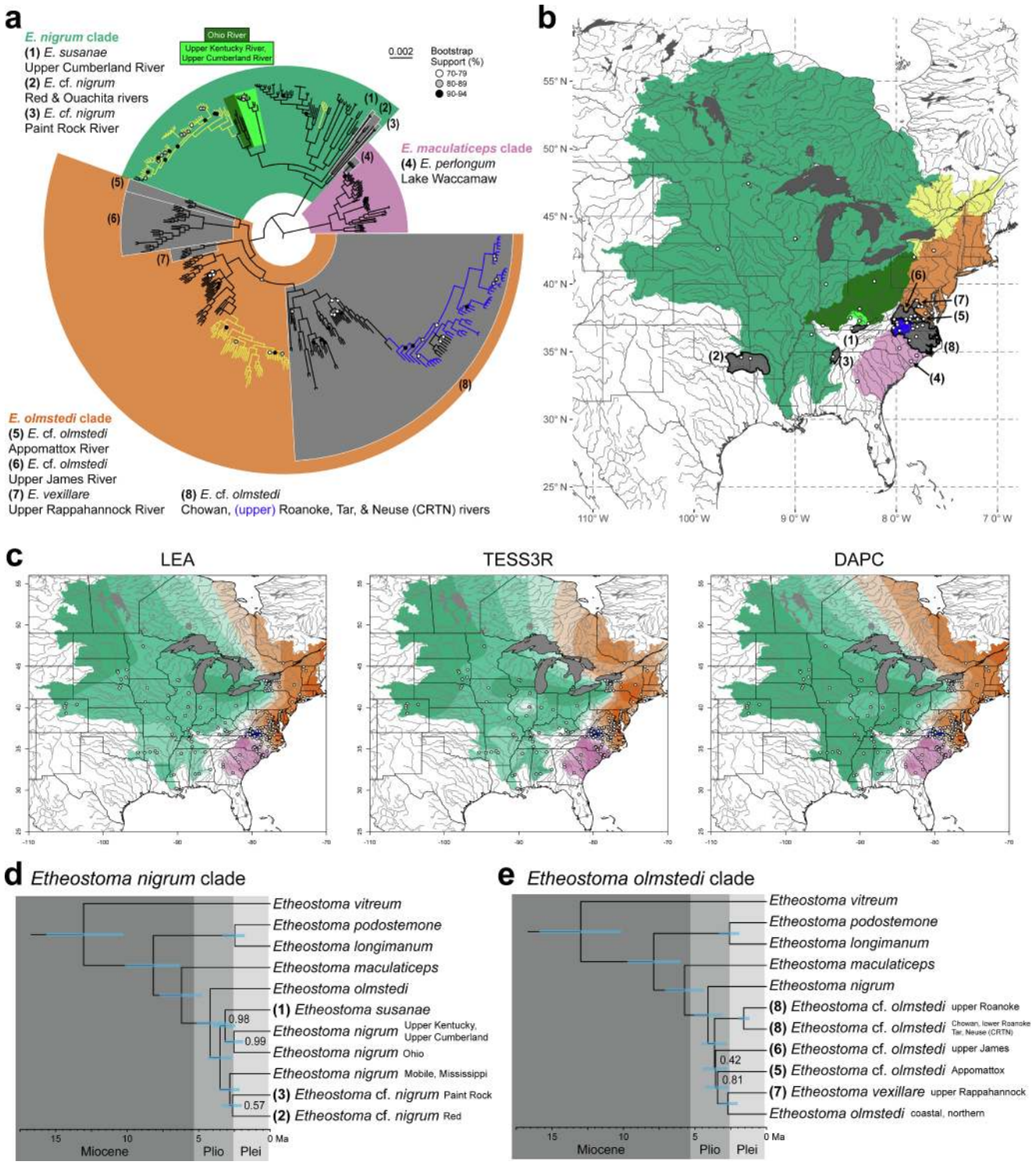


Fig. 2. a) Maximum likelihood concatenated ddRAD phylogeny inferred from 46,564 loci with < 40 % missing samples per locus, branch lengths proportional to the expected number of substitutions per site. Nodes with bootstrap support >=95 % are unlabeled. Nodes with bootstrap support < 70 % are collapsed. *Etheostoma mariae, E. vitreum, E. podostemone,* and *E. longimanum* are pruned from the phylogeny. Yellow branches indicate samples from the area of overlap between *E. nigrum* and *E. olmstedi*, blue branches indicate samples from the upper Roanoke River. b) Range map of clades in the *Etheostoma nigrum* species complex. Yellow indicates *E. nigrum* and *E. olmstedi* zone of contact, blue indicates the upper Roanoke River. Green shading indicates specific *E. nigrum* lineages. White points indicate localities included in species tree and TreeMix analyses. c) Geographic interpolation of ancestry coefficients (LEA and TESS3R) or assignment probabilities (DAPC) for four genetic clusters. White points indicate all sampling localities. d-e) SNAPP time-calibrated species trees for *Boleosoma*, with expanded sampling for the *E. nigrum* clade or the *E. olmstedi* clade. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.1. *Etheostoma nigrum* Phylogeography

ddRAD and mtDNA phylogenies resolve three early diverging lineages within the *Etheostoma nigrum* clade. The two earliest diverging lineages occupy the Paint Rock River system of Tennessee and Alabama and the Red River system of Oklahoma and Arkansas (Fig. 2, Fig. S3). The ddRAD phylogenies resolve these as sister lineages, while the mtDNA gene tree resolves the Red River lineage sister to all other *E. nigrum* populations (Fig. S4). Additionally, the ddRAD phylogeny resolves one individual from the Ouachita River basin sister to individuals from the Red River system (Fig. S3); in the mtDNA phylogeny, this individual is deeply nested within the more inclusive *E. nigrum* clade (Fig. S4). The third early diverging lineage in the ddRAD phylogenies is *E. susanae*, endemic to the upper Cumberland River in Kentucky (Fig. 2, Fig. S3). However, in the mtDNA phylogeny, *E. susanae* is sister to populations of *E. nigrum* from the upper Cumberland River and the adjacent upper Kentucky River system (Fig. S4).

In the ddRAD phylogenies, most of the sampled populations of *Etheostoma nigrum* resolve into two geographically widespread clades. The first clade contains individuals from the Ohio River basin and from the contact zone between *E. olmstedi* and *E. nigrum* in western New York (clade 1, Fig. S3). Within this clade, *E. nigrum* populations from the upper Kentucky River and the upper Cumberland River (clade 3, Fig. S3) are deeply split from all other Ohio River populations (clade 4, Fig. S3). The other widespread clade within *E. nigrum* is distributed as far north as Wisconsin, as far west as Colorado, and as far south as Alabama (clade 2, Fig. S3). In both widespread clades, the northernmost populations tend to be nested within paraphyletic lineages comprised of more southerly populations.

In the mtDNA phylogeny there are also two widespread, well-supported clades that generally correspond to the widespread ddRAD clades. There are a few notable exceptions: three individuals from tributaries of the lower Tennessee River are nested within the Ohio River clade (clade 1, Fig. S4), *Etheostoma podostemone* is nested within the Mississippi River clade in the mtDNA tree (clade 2, Fig. S4), and individuals from the upper Cumberland and Kentucky Rivers resolve sister to *E. susanae* (clade 3, Fig. S4).

3.1.2. *Etheostoma maculaticeps* Phylogeography

The ddRAD phylogenies resolve the northernmost populations of *Etheostoma maculaticeps* from the Cape Fear River as the sister lineage of all other populations of *E. maculaticeps* (Fig. S5). Populations from the Savannah, Santee, and Altamaha basins form a strongly supported clade that also includes a geographically disjunct population from the St. Johns River in Florida (Fig. S5). Samples from the Little River (New River-Ohio River system) resolve sister to a geographically proximate population of *E. maculaticeps* in the Atlantic watershed Pee Dee River basin (Fig. S5). The mtDNA phylogeny has poor support for relationships within *E. maculaticeps* and does not resolve the same broad phylogeographic patterns (Fig. S6). The Lake Waccamaw endemic *Etheostoma perlongum* resolves as a clade nested within *E. maculaticeps* populations from the Waccamaw River in the ddRAD phylogenies (Fig. 2, Fig. S5). In the mtDNA phylogeny, *E. perlongum* does not form a clade and is intermixed with individuals from the Waccamaw River (Fig. S6).

3.1.3. *Etheostoma olmstedi* Phylogeography

The *Etheostoma olmstedi* clade exhibits complex phylogeographic patterns. All populations in the James, Appomattox, Chowan, Roanoke, Tar, and Neuse river basins upstream of the Fall Line (the boundary between the Coastal Plain and the Piedmont geologic provinces) are currently recognized as *E. nigrum.* However, mtDNA and ddRAD phylogenies reveal that these populations are more closely related to *E. olmstedi* (Fig. 2, Fig S7). In the ddRAD phylogenies, individuals from the James River upstream of the Fall Line and the Appomattox River resolve as sister clades, which together comprise the earliest diverging lineage in the *Etheostoma olmstedi* clade (Fig. 2, Fig. S7). Individuals from the Chowan, Roanoke, Tar, and Neuse (CRTN) basins resolve in a strongly supported clade (Fig. 2, Fig S7). The population from Big Reed Island Creek in the New River basin, a presumed non-native introduction, is deeply nested within the CRTN clade and most closely related to individuals from the headwaters of the Roanoke River (Fig. S7). Finally, populations above the Fall Line in the Rappahannock River resolve as a clade that corresponds to the subspecies *E. olmstedi vexillare,* which we treat as a distinct species *Etheostoma vexillare* (Jordan) (Fig. 2, Fig. S7). All other populations of *E. olmstedi* form a widespread clade ranging from Virginia to Ontario (Fig. S7).

Relationships between the CRTN clade, *Etheostoma vexillare*, and widespread *Etheostoma olmstedi* clade are not strongly supported in the ML ddRAD phylogenies. There is moderate node support in two of the ddRAD phylogenies for a sister relationship between the CRTN clade and the widespread *Etheostoma olmstedi* clade, while the third ddRAD phylogeny infers a strongly supported sister relationship between *E. vexillare* and the widespread *Etheostoma olmstedi* clade (Fig. S7). The northernmost populations in the widespread clade are nested within a paraphyletic grade consisting of more southerly populations (Fig. S7). The mtDNA phylogeny has poor overall phylogeographic resolution of *E. olmstedi* lineages (Fig. S8).

3.1.4. *Etheostoma olmstedi x. E. nigrum contact zone*

We sampled several populations in the western New York contact zone between *Etheostoma olmstedi* and *E. nigurm*. ddRAD phylogenies reveal that one *E. olmstedi* lineage and two *E. nigrum* lineages are present in the contact zone (Fig. 2a). The first *E. nigrum* lineage is widespread in the Lake Ontario basin and is most closely related to upper Ohio River populations (Fig. S3). The second *E. nigrum* lineage is found only in Red Creek, a small tributary of Lake Ontario. This lineage is most closely related to populations from the Lake Erie basin and the upper Mississippi River basin (Fig. S3). Phylogenetic analysis of mtDNA haplotypes reveals widespread, bidirectional introgression within the Lake Ontario drainage basin (Fig. S3, S4, S7, S8). However, we do not find introgressed mtDNA haplotypes in any individuals outside of the Lake Ontario contact zone in New York (Fig. S3, Fig. S7). We reserve detailed examination of hybridization and genomic introgression in western New York for a future study.

3.2. *Species tree inference*

Deeper relationships in the SNAPP species tree are congruent with the concatenated ML phylogenies (Fig. 2). *Etheostoma maculaticeps* is consistently resolved as the sister lineage of all other species in the *Etheostoma nigrum* complex. However, species tree relationships within the *E. nigrum* clade differ significantly from the ML trees (Fig. 2D). The Red River and Paint Rock River lineages are poorly supported as sister taxa in a clade with the widespread Mobile and Mississippi basin *E. nigrum* clade (Fig. 2D). *Etheostoma susanae* resolves in a clade with the remaining *E. nigrum,* including those from the Kentucky and upper Cumberland rivers (Fig. 2D). The second most common posterior tree topology resolves Paint Rock *E. nigrum* sister to the Mobile and Mississippi *E. nigrum* lineage (Fig. S9). All *E. nigrum* lineages originate in the mid-late Pliocene or early Pleistocene, approximately 2.5–3.5 Ma.

Relationships within the *Etheostoma olmstedi* clade inferred by species tree analyses also differ from the ML phylogenies. The Appomattox River and upper James River lineages are not resolved as sister lineages in the species trees (Fig. 2D, E, Fig. S10), though this relationship is frequently observed in the posterior tree distribution (Fig. S9). Depending on taxon sampling, *Etheostoma vexillare* is strongly supported as sister to the coastal, northern *E. olmstedi* lineage (Fig. 2E, Fig. S10B) or sister to the coastal, northern lineage and the CRTN lineage (Fig. S10A). The upper Roanoke and CRTN lineages are strongly supported as sister when both are included in the species tree analysis (Fig. 2E). However, when the CRTN lineage is excluded, the upper Roanoke lineage resolves with moderate support as sister to *E. nigrum* (Fig. S10B). The upper James River lineage, Appomattox River lineage, and *E. vexillare*

originate from the mid Pliocene to the early Pleistocene (Fig. 2E, Fig. S10). In contrast, the divergence time of the upper Roanoke lineage varies from early Pliocene to mid Pleistocene depending on taxon sampling (Fig. 2E, Fig. S10).

3.3. Population genetic structure

Model selection identifies 16–31 distinct genetic clusters as optimal to explain genetic variation within the *Etheostoma nigrum* species complex (Fig. S11). At K = 4, all population genetic methods identify the same four genetic clusters (Fig. 2C) which correspond to the three major clades (Fig. 2A) plus a cluster in the upper Roanoke River. To aid interpretation of population structure, we examine genetic clustering results independently for each clade, grouping upper Roanoke River populations with the *Etheostoma olmstedi* clade. In addition, we examine genetic clustering results for a dataset consisting of all populations on the Atlantic coast (the *E. olmstedi* and *E. maculaticeps* clades). For most datasets, TESS3R identifies the largest optimal K value, while K-means clustering identifies the lowest optimal K value (Fig. S11). The optimal K value is thus strongly influenced by the choice of method and model comparison. For our assessment of population structure within the *Etheostoma nigrum* species complex, we discuss results for a range of K values but focus on the smallest optimal K value identified across each set of analyses for detailed interpretation.

3.3.1. Etheostoma nigrum population structure

LEA, TESS3R, and K-means clustering identify 12, 15, and 11 clusters, respectively as optimal to explain the genetic variation within the *Etheostoma nigrum* clade (Fig. S11). For LEA and K-means clustering, a large range of K values result in similar cross-entropy or BIC scores. All three analyses identify similar genetic clusters up to K = 4 (Fig. S12). There is some correspondence between phylogenetic clades and genetic clusters. For example, at K = 4, one genetic cluster corresponds to the upper Cumberland River and Kentucky River clade, another cluster matches the upper Ohio River clade, while the remaining two genetic clusters comprise the widespread Mobile and Mississippi clade (Fig. S12, S3). Above K = 4, clustering methods identify slightly different groups. For instance, each method identifies six similar clusters at K = 11: an upper Mississippi and Missouri River cluster (cluster 1, Fig. 3A), a lower Mississippi River cluster (cluster 2, Fig. 3A), an upper Kentucky River and Cumberland River cluster (cluster 3, Fig. 3A), a cluster corresponding to *E. susanae* (cluster 4, Fig. 3A), a cluster in the lower Tennessee River (cluster 5, Fig. 3A), and a Mobile River and Ouachita River cluster (cluster 6, Fig. 3A). However, LEA and TESS3R identify the Paint Rock River as a distinct cluster (cluster 7, Fig. 3A), while TESS3R and DAPC identify a cluster in the lower Ohio River (cluster 10, Fig. 3A). Overall, clustering results identify more *E. nigrum* genetic structure in the south, with many small pockets of distinct genetic ancestry.

3.3.2. Etheostoma maculaticeps populations structure

Analyses of the *Etheostoma maculaticeps* clade reveal less population structure than the *E. nigrum* or *E. olmstedi* clade. Model selection with LEA, TESS3R, and K-means clustering identifies 6, 45, or 4 clusters as optimal to describe the genetic variation (Fig. S11). All three methods identify similar genetic clusters up to K = 6, although K-means clustering generally has higher individual assignment probabilities (Fig. 3B). At K = 4, we identify four genetic clusters that match clades in the ddRAD phylogenies (Fig. S5): the Cape Fear River basin (cluster 1, Fig. 3B) the upper Pee Dee and New rivers (cluster 2, Fig. 3B), the lower Pee Dee River, Waccamaw River, and Lake Waccamaw (cluster 3, Fig. 3B), and the Santee, Savannah, Altamaha, and St. Johns rivers (cluster 4, Fig. 3B). Despite its geographic isolation, the Florida population in the St. Johns River does not form a distinct genetic cluster (Fig. 3B). Although we sampled four populations in the Waccamaw River, including one population only 20 km downstream of Lake Waccamaw, *E. perlongum* from Lake Waccamaw was identified as a distinct genetic cluster in some analyses (Fig. S13).

3.3.3. Etheostoma olmstedi population structure

Despite having a smaller range size than *Etheostoma nigrum*, we detect a similarly high degree of genetic structure in the *E. olmstedi* clade. LEA, TESS3R, and K-means clustering model selection infer 19, 36, and 9 clusters to optimally describe genetic variation in the *Etheostoma olmstedi* clade (Fig. S11). Up to K = 5, all methods produce consistent clustering results (Fig. S14). At K = 9, genetic structure is also consistent across methods with one exception. LEA identifies two genetic clusters in the James River (clusters 7 and 9, Fig. 3C), while TESS3R and DAPC identify a cluster restricted to the Appomattox River (cluster 9, Fig. 3C). Like the *E. nigrum* clade, there is substantially more genetic structure in southern populations of *E. olmstedi*. Populations of *Etheostoma olmstedi* from the Virginia coastal plain north to Ontario are comprised of only two genetic clusters (clusters 1 and 2, Fig. 3C). The remaining seven genetic clusters occur in rivers of Virginia and North Carolina, described in detail below.

Atlantic Seaboard Population Structure – The *Etheostoma olmstedi* and *Etheostoma maculaticeps* clades occupy coastal plain rivers of the Atlantic Seaboard that share many freshwater connections. Population structure analyses with LEA, TESS3R, and K-means reveal 25, 32, and 12 clusters as optimal to describe genetic variation within these clades (Fig. S11). At K = 12, three of the genetic clusters are restricted to the *Etheostoma maculaticeps* clade (Fig. 4A, Fig. S15) and are identical to genetic structure inferred when analyzing *E. maculaticeps* alone (Fig. S13). Likewise, the remaining nine genetic clusters of the *Etheostoma olmstedi* clade (Fig. 4A) are identical to genetic structure inferred when analyzing *E. olmstedi* alone (Fig. 3). Signals of admixture are weak and inconsistent between the *Etheostoma maculaticeps* and *Etheostoma olmstedi* clades despite their abutting ranges. One noteworthy exception is the single *E. maculaticeps* specimen sampled from the upper Cape Fear River, which has approximately 50 % ancestry derived from *E. olmstedi* ("EmacFG_FEAR_NC", Fig. S15).

Several genetic clusters within the *Etheostoma olmstedi* clade are delimited by the Fall Line, the boundary between the Coastal Plain and the Piedmont geologic provinces. The Rappahannock River upstream of the Fall Line (*E. vexillare*) is comprised of a single genetic cluster with no admixture (Fig. 4B). Our analyses detect some admixture from *E. vexillare* in Rappahannock populations downstream of the Fall Line (Fig. 4C). One individual from a locality just downstream of the Fall Line has approximately 60 % ancestry derived from the *E. vexillare* genetic cluster ("EolmUF_RAPP_VA", Fig. 4C). Similarly, populations in the James River and Appomattox River upstream of the Fall Line form distinct genetic clusters, but with little downstream admixture (Fig. 4B). Surprisingly, the York River exhibits no such genetic break at the Fall Line (Fig. 4B).

The Chowan, Roanoke, Tar, and Neuse (CRTN) river basins contain four genetic clusters (Fig. 4B). However, unlike populations in the Rappahannock, James, and Appomattox, CRTN genetic clusters are not delimited by the Fall Line. Three genetic clusters in the Chowan River and Roanoke River basins exhibit clinal variation from upstream to downstream with no hard breakpoints (Fig. 4B). The fourth genetic cluster is largely restricted to the Tar and Neuse basins. In some CRTN populations, LEA and TESS3R detect admixture derived from the widespread coastal and northern *Etheostoma olmstedi* genetic clusters (Fig. 4B, Fig. S15). These admixed populations are limited to the Tar, Neuse, and Chowan basins, plus the middle and lower reaches of the Roanoke River. Populations in the Roanoke River upstream of the Fall Line exhibit little or no admixture with the northern *E. olmstedi* genetic clusters (Fig. 4B, Fig. S15).

3.4. Tree-based migration estimation

We used TreeMix to examine signals of gene flow in 100 random sets of 10,871 unlinked SNPs, retaining migration edges detected in greater

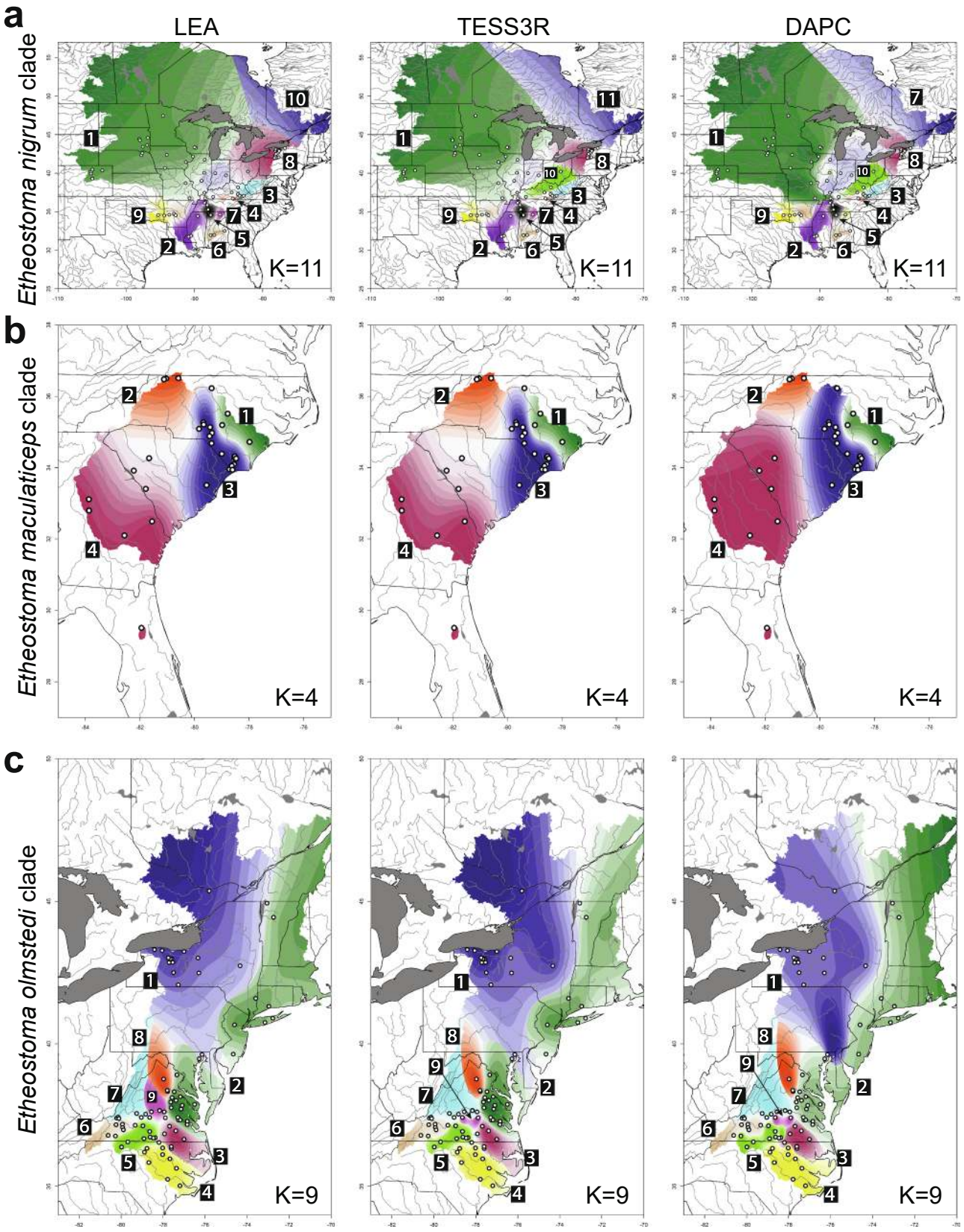


Fig. 3. Population structure for each clade in the *Etheostoma nigrum* species complex. Geographic interpolation of ancestry coefficients (LEA and TESS3R) or assignment probabilities (DAPC) for a) the *E. nigrum* clade, b) the *E. maculaticeps* clade, and c) the *E. olmstedi* clade. Each clade displays the number of genetic clusters (K) with the lowest BIC score from K-mean clustering (Fig. S10). Genetic clusters are labeled on each map. For the LEA and DAPC analyses of the *E. nigrum* clade, no individuals were strongly assigned to the 11th genetic cluster. Sampling localities are marked on each map.

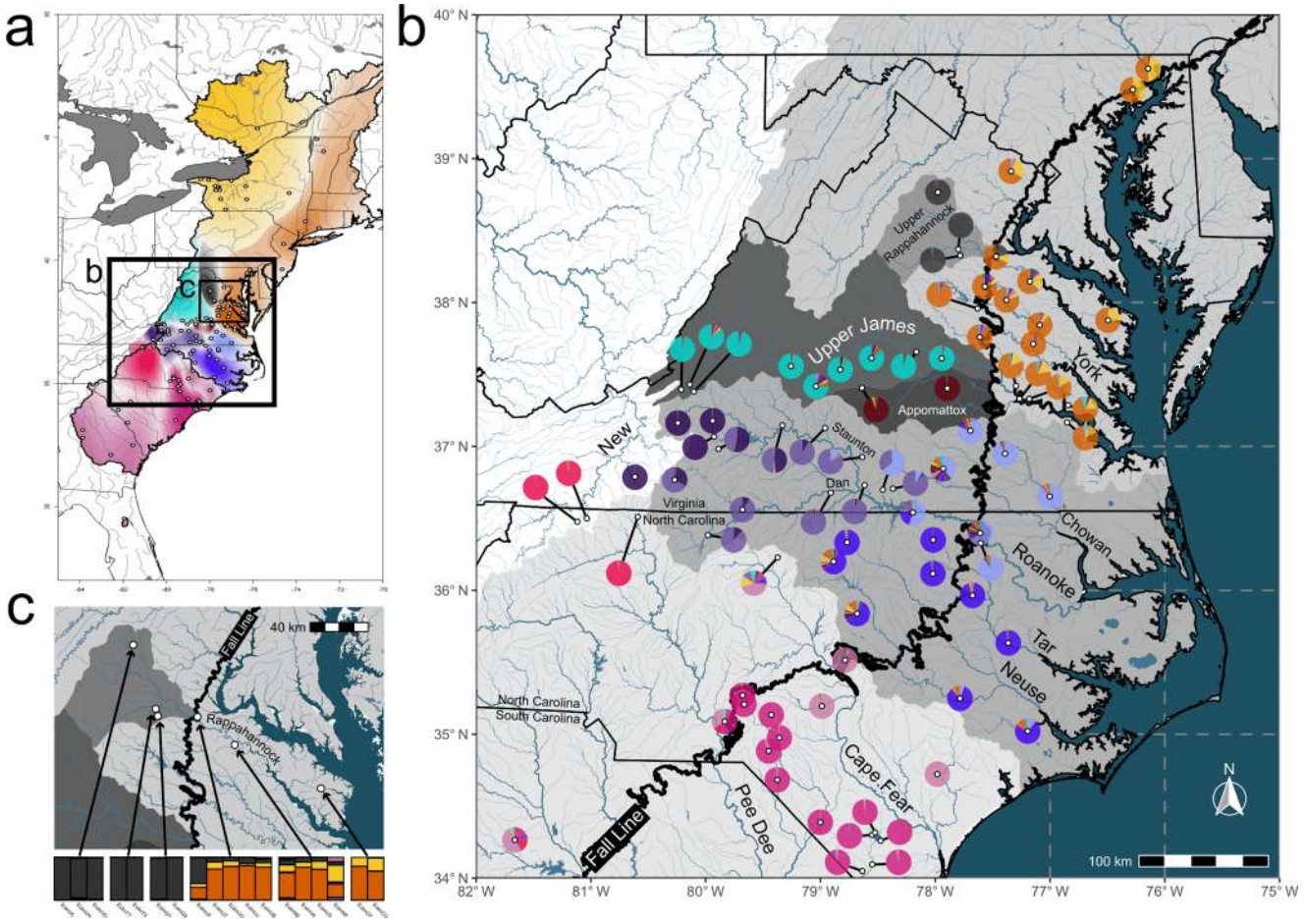


Fig. 4. Population structure for the *Etheostoma olmstedi* and *E. maculaticeps* along the Atlantic Seaboard. a) Geographic interpolation of LEA ancestry coefficients for 12 genetic clusters. b) Small white points indicate sampling localities. Pie charts display ancestry coefficients averaged across all individuals at each locality. c) Bars represent individual ancestry coefficients within the Rappahannock River basin.

than 20 % of SNP subsets. TreeMix likelihood scores plateau for analyses with more than four migration edges (Fig. S16A) and the same four migration edges are consistently detected in >75 % of SNP subsets (Fig. S16B). Across all analyses with more than four migration edges, no other migration edge is supported by more than 50 % of SNP subsets (Fig. S16B). Therefore, we focus on the TreeMix analyses with four allowed migration edges. With four migration edges, the bifurcating TreeMix topology is nearly identical the ML phylogenies (Fig. 2, Fig. S1) except for the upper Roanoke River *Etheostoma olmstedi* lineage, which is resolved as sister to the *Etheostoma nigrum* clade (Fig. 5A), a result also observed in some species tree analyses (Fig. S10B).

All SNP subsets identify a migration edge from *Etheostoma susanae* to the upper Kentucky River and Cumberland River populations of *E. nigrum* (Fig. 5A,B). This edge has a mean migration weight of 0.20 (Fig. 5C), indicating that about 20 % of *E. nigrum* genomic ancestry in the upper Kentucky and Cumberland rivers is derived from gene flow with *E. susanae* (Fig. 5A). There is considerable variance in the estimated migration weight for this edge (sd = 0.05) compared to the other three migration edges (sd = 0.007, 0.005, and 0.007). The spatial extent of *E. susanae* gene flow is limited; TreeMix does not detect gene flow from *E. susanae* to *E. nigrum* populations further downstream in the Kentucky River or Ohio River basins (Fig. S16B).

The next three strongest signals of gene flow involve lineages on the Atlantic Seaboard. 98 % of SNP subsets identify a migration edge from the upper Roanoke *Etheostoma olmstedi* lineage into the geographically adjacent upper James *E. olmstedi* lineage with a mean migration weight of 0.10 (Fig. 5). 97 % of SNP subsets also identify a migration edge from the upper Roanoke *E. olmstedi* lineage into the Chowan, lower Roanoke, Tar, and Neuse (CRTN) lineage with a migration weight of 0.24 (Fig. 5). 91 % of SNPs subsets identify a final migration edge from *E. maculaticeps* into the CRTN lineage with a migration weight of 0.08 (Fig. 5). Although we observe mtDNA introgression within the western New York contact zone between *E. olmstedi* and *E. olmstedi,* TreeMix analyses do not infer a migration edge between these lineages (Fig. S16B).

4. Discussion

Phylogenetic and population structure analyses uncover an intricate history of geographic isolation and secondary contact in the *Etheostoma nigrum* species complex. Our results provide a robust phylogenetic hypothesis of relationships within this group. We reveal that the *Etheostoma nigrum* complex is a microcosm of the pervasive pattern of allopatric speciation observed in darters and other clades of freshwater fishes (Near et al. 2011, Seehausen and Wagner 2014). We highlight unrecognized species-level diversity, both validating and revising the taxonomy of species within the *Etheostoma nigrum* complex. This study illustrates the power of genomic data to decipher tangled evolutionary histories at the species boundary.

4.1. A revived southern species: Etheostoma maculaticeps

Current taxonomy recognizes four species within the *Etheostoma nigrum* complex: *E. nigrum* Rafinesque, *E. olmstedi* Storer, *E. susanae* (Jordan), and *E. perlongum* (Hubbs and Raney). *Etheostoma nigrum* and

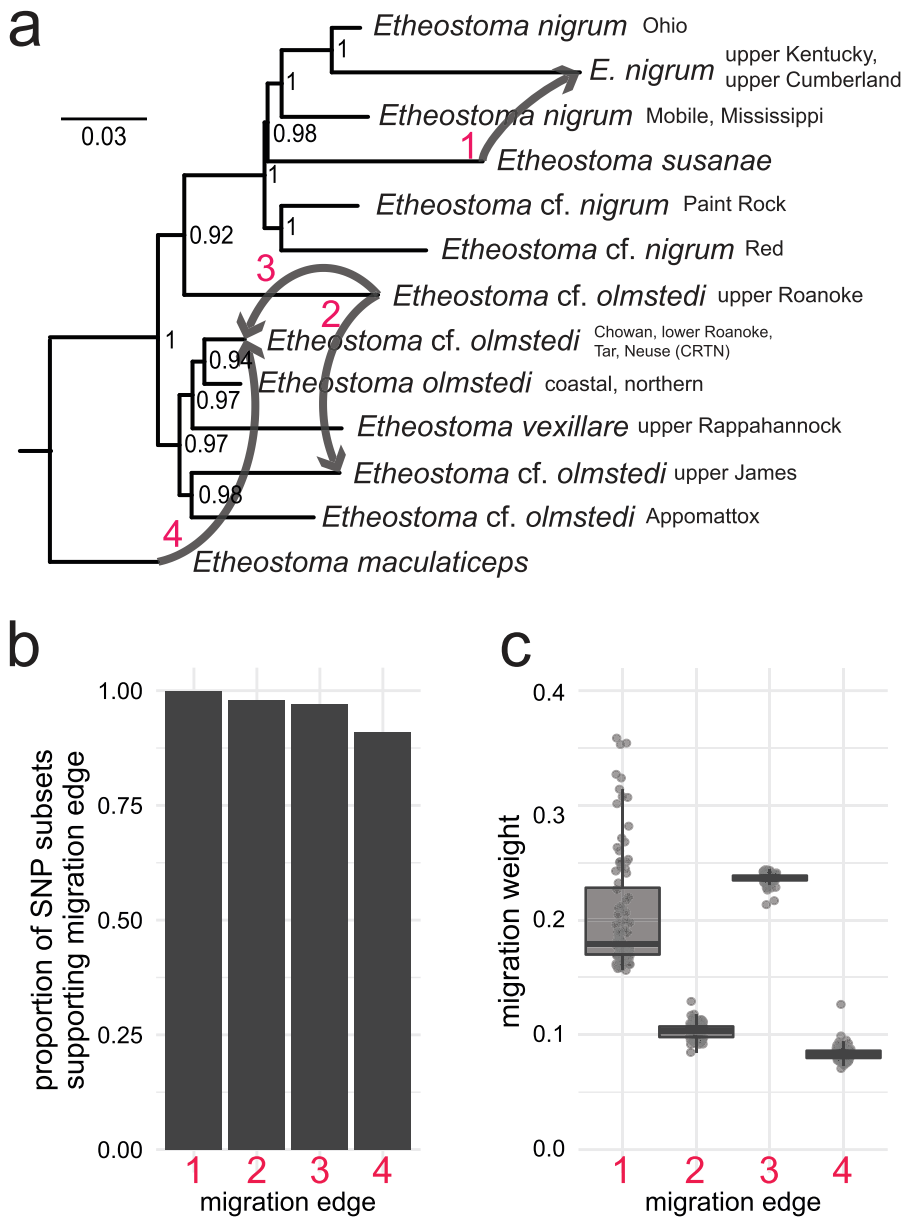


Fig. 5. TreeMix estimation of gene flow with four allowed migration edges. a) TreeMix network topology. Numbers represent the proportion of SNP subsets that inferred the given clade. Branch lengths are proportional to the TreeMix drift parameter. Arrows indicate inferred migration edges. b) Proportion of 100 SNP subsets that infer each migration edge. c) Distribution of migration weights inferred by 100 SNP subsets for each migration edge.

E. olmstedi are widespread and largely isolated on opposite sides of the Appalachian Mountains (Fig. 1). However, phylogenetic analyses of mtDNA and ddRAD data reveal that *E. olmstedi* is not monophyletic. All populations of *E. olmstedi* south of the Neuse River basin resolve as the sister clade of northern *E. olmstedi* plus *E. nigrum* (Fig. 2).

The southern *Etheostoma olmstedi* clade did not escape the notice of earlier taxonomists. Cope (1870) described *Boleosoma maculaticeps* from the Catawba River in North Carolina as a distinct species largely based on the presence of two anal fin spines, compared to *Etheostoma olmstedi* which has only a single anal fin spine. *Boleosoma maculaticeps* was later treated as a subspecies (Jordan 1888). However, Cole (1957) also found that 77 % of individuals (n = 1,206) from southern populations (*E. o. maculaticeps*) had two anal fin spines while 91 % of northern *E. olmstedi* individuals had one anal fin spine (n = 2,752). Other meristic trait distributions largely overlap between the two species (Cole 1957, 1967). Additional phenotypic differences remain unstudied, particularly male breeding pigmentation patterns. Based on the genetic distinctiveness of southern *E. olmstedi* populations and the diagnostic character of two anal fin spines, we treat *Etheostoma maculaticeps* (Cope) as a distinct species in the darter clade *Boleosoma*.

There is little evidence of contemporary admixture between *Etheostoma olmstedi* and *E. maculaticeps*; however, we found one *E. maculaticeps* specimen from the upper Cape Fear River that exhibits admixed ancestry suggestive of gene flow with *E. olmstedi*. Populations of *E. maculaticeps* in the upper Cape Fear River are morphologically similar to *E. olmstedi*, including an interrupted infraorbital canal (Cole 1957, 1967, 1971). In addition, TreeMix analyses identify a minor signal of gene flow from *E. maculaticeps* to *E. olmstedi* populations in the Chowan, lower Roanoke, Tar, and Neuse rivers, which we discuss in detail below. Further genetic and morphological investigation of upper Cape Fear River *E. maculaticeps* is needed to explore the possibility of introgression with *E. olmstedi*.

Unlike *Etheostoma nigrum* and *E. olmstedi*, we find no evidence of widespread secondary contact between *E. maculaticeps* and *E. olmstedi*

even though these taxa occupy coastal rivers that are connected during flooding and periods of low sea level (Cole 1967; Zorach 1971). Indeed, the Cape Fear-Neuse drainage divide is a biogeographic dispersal barrier for many taxa with disparate ecologies (Hocutt et al. 1986, Engle and Summers 1999). The geologic history of the region may explain this biogeographic boundary. The Cape Fear arch is an area of uplift roughly centered along the Cape Fear-Neuse divide and extending onto the continental shelf (Markewich 1985; Wheeler 2006; Van De Plassche et al. 2014), with recent uplift starting in the mid-Pliocene (Harris et al. 1979, Hack 1982; Markewich 1985). Piedmont rivers on either side of the Cape Fear arch differ in gradient, shape, direction, and paleodrainage connections due to underlying geologic formations (Hack 1982, Baum et al. 1979). Recent uplift caused downcutting and deeper incision of the Cape Fear River during the late Cenozoic, likely reducing opportunities for inter-basin dispersal via stream capture or flooding events (Markewich 1985). We hypothesize that uplift along the Cape Fear arch created a strong geographic dispersal barrier preventing widespread secondary contact and gene flow between *E. olmstedi* and *E. maculaticeps*.

4.2. Comments on the Lake Waccamaw Endemic, Etheostoma perlongum

Etheostoma perlongum, endemic to Lake Waccamaw, North Carolina, has a more elongate body, more vertebrae, and more lateral line scales than *E. maculaticeps* (Hubbs and Raney 1946; Krabbenhoft et al. 2009). The streamlined morphology of *E. perlongum* is hypothesized to be an adaptive response to higher predation pressure in an open lake environment (Hubbs and Raney 1946), making *E. perlongum* the only known case of lacustrine ecological speciation within darters. We find that *E. perlongum* is phylogenetically nested within *E. maculaticeps* and most closely related to *E. maculaticeps* from the Waccamaw River (Fig. S5). Although we sampled several localities in the Waccamaw River, *Etheostoma perlongum* from Lake Waccamaw forms a distinct genetic cluster in population structure analyses at higher K values (Fig. S13). This genetic differentiation is remarkable considering that Lake Waccamaw formed only 15,000–32,000 years ago (Stager and Cahoon 1987). It is unlikely that genetic differentiation would be observed over such a small spatial scale if gene flow is unrestricted between Lake Waccamaw and the Waccamaw River. Rapid lacustrine speciation is common in fishes, exemplified by the 500–1,000 cichlid species endemic to Lake Victoria that diversified in the past 14,600 years (Seehausen 2006). We consider *E. perlongum* a distinct species and suggest thorough genomic sampling of the Waccamaw basin will uncover the evolutionary history of this unusual lineage.

4.3. Undescribed biodiversity at the range periphery of Etheostoma nigrum

There are many peripheral, geographically isolated populations in the *Etheostoma nigrum* species complex. However, geographic isolation is not always indicative of evolutionary distinctiveness. For example, the St. Johns River in Florida contains a disjunct population of *E. maculaticeps* separated by about 270 km from the closest populations in Georgia. Likewise, the Platte River basin in Colorado and Wyoming contains isolated populations of *E. nigrum* that are about 400 km upstream of the closest populations in Kansas and Nebraska. Despite their geographic isolation, neither of these populations are genetically distinct (Fig. S3).

However, in some cases, peripheral isolation does lead to evolutionary divergence. We observe a striking pattern in the *Etheostoma nigrum* clade: the three earliest diverging lineages are all peripheral isolates. The oldest lineages in ddRAD and mtDNA phylogenies are found in the Red River and Paint Rock River at range periphery of *E. nigrum* (Fig. 2, Fig. S2). Species tree inference suggests that the Paint Rock River and Red River lineages originated in the late Pliocene, ~2.6 Ma (Fig. 2D). Mitochondrial DNA introgression is common in darters (Near et al. 2011), so persistence of old mtDNA lineages unique to the Paint Rock River and Red River suggests a lack of gene flow with other *E. nigrum* populations. However, one *E. nigrum* individual from the upper Ouachita River resolves as sister to the Red River lineage in the ddRAD phylogenies ("EngrBK_OCHI_AR", Fig. S3). Although the Ouachita River basin is adjacent to the Red River basin, they only share a freshwater connection far downstream near the Mississippi River delta. We hypothesize that there was historical gene flow from the isolated Red River lineage into upper Ouachita River *E. nigrum* populations through a paleodrainage connection (Matthews and Robison 1982; Mayden 1985).

The Red River and Paint Rock River *E. nigrum* lineages may represent distinct and morphologically cryptic species, but a thorough analysis of phenotypic variation is necessary. Unfortunately, there has been no morphological study of *E. nigrum* that included specimens from the Paint Rock River. Individuals from the Red River basin were included in an unpublished master's thesis, but no diagnostic morphological characteristics were identified (Krotzer 1990).

4.4. Etheostoma susanae genetic divergence and gene flow

Etheostoma susanae is endemic to the Cumberland River upstream of the Cumberland Falls, a significant dispersal barrier. *Etheostoma nigrum* is largely absent downstream of the Cumberland Falls, with only a handful of confirmed records despite intensive sampling throughout the river basin (Etnier and Starnes 1993). Our study does not include specimens from these rare populations. Previous studies of darter mtDNA haplotypes suggested a close relationship between *E. susanae* and upper Kentucky River *E. nigrum* (Strange 1998; Heckman et al. 2009). Additionally, Strange (1998) found a population of *E. nigrum* in the Poor Fork of the upper Cumberland River that was thought to be occupied exclusively by *E. susanae*. Our comprehensive mtDNA sampling confirms these earlier results and resolves *E. susanae* sister to upper Kentucky and Cumberland River *E. nigrum* (Fig. S4). *Etheostoma nigrum* has expanded further downstream in the Cumberland River than previously suspected.. We identified an *E. nigrum* specimen (EngrEK_POFK_KY, YPM ICH.027390) from a locality 22 km downstream of the original Poor Fork locality reported in Strange (1998).

In the mtDNA gene tree, a clade comprising *Etheostoma nigrum* populations from the upper Kentucky and Cumberland rivers and *E. susanae* is resolved as the sister lineage of all other *E. nigrum* except for the Paint Rock River and Red River lineages (Fig. S4). In contrast, ML ddRAD analyses resolve *E. susanae* as the sister lineage of all other populations of *E. nigrum* except for the Red River and Paint Rock River (Fig. 2, Fig. S3), while populations from the upper Kentucky and Cumberland rivers are nested within *E. nigrum* and most closely related to populations in the Ohio River basin (clade 4, Fig. S3). Species tree inference supports a different topology but still resolves upper Kentucky and Cumberland populations as the sister lineage of populations in the Ohio River basin (Fig. 2D). Thus, it appears there is complete replacement of *E. nigrum* mtDNA haplotypes in the upper Kentucky and Cumberland rivers with *E. susanae* mtDNA haplotypes. This mtDNA introgression is likely not recent, considering there is 1.9 % sequence divergence between upper Kentucky River and Cumberland River *E. nigrum* haplotypes and *E. susanae* haplotypes. There are no *E. susanae* mtDNA haplotypes in *E. nigrum* populations further downstream in the Kentucky River, indicating that historical mtDNA introgression was spatially limited (Fig. S4).

Previous biogeographic studies hypothesized that headwater exchange explains the shared fauna of the upper Cumberland River and Kentucky River systems (Kuehne and Bailey 1961; Jenkins et al. 1972; Starnes and Starnes 1978, 1979). For instance, the darter species *Etheosoma baileyi* is restricted to the upper Cumberland and Kentucky basins, while the darter sister species pair *E. sagitta* and *E. spilotum* are separated in the upper Cumberland and upper Kentucky basins (Kuehne and Bailey 1961; Near et al. 2011). Our analyses strongly support this hypothesized headwater connection. In addition to the mtDNA

introgression, TreeMix analyses reveal that about 20 % of *E. nigrum* genomic ancestry in the upper Kentucky and Cumberland rivers derives from gene flow with *E. susanae* (Fig. 5). This gene flow may have also produced morphological intergradation between *E. susanae* and upper Kentucky River *E. nigrum* (Starnes and Starnes 1979).

Etheostoma susanae has a contentious taxonomic history (Starnes and Starnes 1979; O'Bara 1991; Strange 1998). Morphological differences thought to distinguish *E. susanae* from *E. nigrum* include an interrupted preoperculomandibular (POM) canal, lack of scales on the nape, operculum, and midbelly, and an interrupted preorbital stripe of dark pigmentation (Jordan and Swain 1883; Starnes and Starnes 1979). However, many *E. nigrum* populations show variation in the degree of squamation, and other *E. nigrum* populations also exhibit interrupted POM canals (Underhill 1963; Cole 1971; Krotzer 1990). Pigmentation differences remain the only reliable diagnostic character. Our analyses reveal that *E. susanae* is a deeply divergent lineage that originated in the late Pliocene (Fig. 2D), validating the continued recognition of *E. susanae* as a distinct species. *Etheostoma susanae* is listed as endangered under the United States Endangered Species Act and has experienced dramatic population declines due to habitat degradation from regional mining operations (Starnes and Starnes 1979; O'Bara 1991; Strange 1998). Range expansion of *E. nigrum* in the Poor Fork of the Cumberland River could facilitate additional gene flow between these lineages and poses a threat to *E. susanae*, highlighting the need for a detailed population-level study to inform conservation action.

4.5. Atlantic Fall Line biogeography

The Atlantic Fall Line marks the boundary between the Coastal Plain and the Piedmont physiographic provinces and is spanned by many parallel rivers flowing from mountain headwaters to the Atlantic Ocean. The Coastal Plain is a low gradient, low elevation region with deep sedimentary formations (Fenneman 1914, Fenneman 1938) and sluggish, marshy rivers that often have low pH blackwater conditions (Jenkins and Burkhead 1993). In contrast, the Piedmont region comprises an eroded portion of the Appalachian Mountains (Fenneman 1914, 1938) with variable freshwater habitats, ranging from slow, sandy streams near the Fall Line to moderate-gradient, rocky streams further inland (Jenkins and Burkhead 1993). The Fall Line itself is a distinct geologic feature where metamorphic Piedmont bedrock increases in slope, resulting in a series of waterfalls or rapids when rivers cross the boundary (Renner 1927; Carlston 1969).

Uplift of the Piedmont combined with regular inundation of the Coastal Plain throughout the Cenozoic make the Fall Line an important biogeographic feature. Indeed, many organisms have distributions influenced by the Fall Line, including turtles (Tinkle 1959), salamanders (Pauly et al. 2007; Beamer and Lamb 2008), trees (Shankman and Hart 2007), and freshwater fishes (Collette 1962; Williams and Robins 1970; Lang and Echelle 2011). Many freshwater fish distributions span the Fall Line, and it is suggested to play a role in population structure of *Micropterus cataractae* (Taylor et al. 2018) and *Gambusia holbrooki* (Hernandez-Martich and Smith 1997). However, there has been no range-wide assessment of genomic variation for a species distributed above and below the Fall Line.

Etheostoma olmstedi presents an ideal system to study the biogeographic role of the Fall Line. Our analyses reveal that populations upstream of the Fall Line in the James River, Appomattox River, and Rappahannock River comprise the three earliest diverging lineages in the *Etheostoma olmstedi* clade (Fig. 2). Population structure analyses identify a genetic break at the Fall Line in these river basins with evidence of downstream gene flow across the Fall Line in only the Rappahannock River (Fig. 3). A fourth early diverging lineage of *E. olmstedi* straddles the Fall Line in the Chowan, Roanoke, Tar, and Neuse basins (Fig. 2). Some populations in the Chowan, lower Roanoke, Tar, and Neuse (CRTN) basins are admixed with the coastal, northern *E. olmstedi* clade (Fig. 3). Species tree and TreeMix analyses also infer a close phylogenetic relationship between the CRTN populations and coastal, northern *E. olmstedi* (Fig. 5, Fig. S10B). However, Roanoke River populations above the Fall Line are genetically distinct and are not admixed with coastal, northern *E. olmstedi,* though there is a strong signal of downstream gene flow from the upper Roanoke lineage into other CRTN populations (Fig. 5). In some analyses, the upper Roanoke lineage resolves as sister to the *E. nigrum* clade (Fig. S10B). There is a weak signal of gene flow from the upper Roanoke lineage into the upper James lineage that may indicate historical stream capture between these basins (Fig. 5).

When admixed CRTN populations are excluded, time-calibrated species tree analyses infer that the upper Roanoke (4.3 Ma), James (3.7 Ma), Appomattox (3.4 Ma), and Rappahannock (2.6 Ma) lineages all originated in the mid to late Pliocene (Fig. S10B). The mid-Pliocene (2.9–3.3 Ma) is notable for considerably warmer global temperatures (Raymo et al. 1996) that were accompanied by elevated sea levels, extending the shoreline as far inland as the Fall Line (Dowsett and Cronin 1990; Miller et al. 2005, 2012; Krantz 1991; Kominz et al. 2008; Rovere et al. 2015). We hypothesize that Pliocene inundation provided conditions for the initial allopatric isolation and subsequent divergence of *E. olmstedi* lineages above the Fall Line.

The Coastal Plain was reopened to colonization by freshwater species after sea levels receded during the early Pleistocene. While the upper James and Appomattox lineages evidently remained isolated upstream of the Fall Line, there are signals of downstream gene flow from the upper Rappahannock and upper Roanoke lineages into the Coastal Plain. We hypothesize that the Rappahannock River contains an active hybrid zone, discussed in detail below. However, gene flow in the Roanoke River basin is more complex. Populations in the CRTN basins have ancestry derived from upper Roanoke *E. olmstedi* and coastal, northern *E. olmstedi*. Additionally, TreeMix detects a minor signal of gene flow from *E. maculaticeps* into the CRTN populations (Fig. 5). We hypothesize that the CRTN basins were colonized from three different directions when sea levels fell in the Pleistocene, producing admixed populations that may represent the remnants of an old hybrid swarm.

4.6. Unrecognized species diversity within Etheostoma olmstedi

Deep divergence time estimates suggest the four Fall Line lineages of *Etheostoma olmstedi* are distinct species (Fig. 2E). Populations in the upper Rappahannock River are inconsistently treated as a subspecies *E. olmstedi vexillare* (Jordan 1880; Cole 1957; Page 1983; Near et al. 2011). Cole (1967) noted the presence of morphological intermediacy between *E. o. vexillare* and *E. olmstedi* from the upper Pamunkey River of the York River basin. While we included samples from the Pamunkey River above the Fall Line, these individuals are not closely related to *E. o. vexillare* in the molecular phylogenetic analyses (Fig S7). However, in the Rappahannock River just below the Fall Line there is evidence for secondary contact between *E. olmstedi* and *E. o. vexillare*, with genetic admixture (Fig. 4C) and morphological intermediacy (Cole 1967). Unidirectional downstream gene flow is limited to the Rappahannock River system (Fig. 4B,C) and Cole (1967) detected only a single morphologically intermediate individual out of 50 specimens from just downstream of the Fall Line. Based on our genetic analyses and the morphological distinctiveness described by Cole (1967), we treat *Etheostoma vexillare* (Jordan) as a distinct species.

Populations of *Etheostoma olmstedi* in the upper James and Appomattox Rivers are genetically distinct with no evidence of secondary contact or gene flow downstream of the Fall Line (Fig. 4B). In contrast, populations in the upper Roanoke River exhibit extensive downstream gene flow into populations of the Chowan, lower Roanoke, Tar, and Neuse (CRTN) rivers (Fig. 4B). These three lineages have a complex taxonomic history. Cole (1957) described minor morphological differences and suggested subspecies status for the lineages above the Fall Line, but later considered them to be only "racially distinct" members of *E. nigrum* due to the supposedly diagnostic trait of interrupted

infraorbital canals (Cole 1971). While the infraorbital canal is consistently interrupted in *E. nigrum*, other cephalic canals show considerable geographic variation in completeness (Krotzer 1990). Additionally, presence of a complete infraorbital canal is not a fixed character in *E. olmstedi,* with 13 % of individuals from a range-wide sample exhibiting an incomplete canal (n = 1,459) (Cole 1957). Therefore, features of the cephalic canal system are not reliable morphological characters to diagnose species of the *Etheostoma nigrum* complex.

Instead, we hypothesize that *Etheostoma olmstedi* lineages upstream of the Fall Line are paedomorphic. Some features of the cephalic system form late during development in other darter species (Collette 1962). *Etheostoma olmstedi* lineages upstream of the Fall Line generally attain smaller maximum body size and reduced numbers of lateral line scales compared to downstream populations, possibly indicating a truncated developmental period in colder, higher elevation streams (Cole 1967). Taken together, our results suggest that these lineages above the Fall Line are distinct species, neither *E. nigrum* nor *E. olmstedi*. A detailed re-evaluation of the morphology of these lineages is needed, but we suggest that the upper James River, Appomattox River, and upper Roanoke River lineages are distinct and undescribed species.

5. Conclusions

We present a range-wide investigation of genomic variation in a geographically widespread and taxonomically complex group of freshwater fishes. Our results show that high levels of genetic structure predispose the *Etheostoma nigrum* species complex to allopatric speciation. This genetic structure includes unrecognized species-level biodiversity, namely 1) *E. maculaticeps* ranging along the Atlantic Coast from North Carolina to Florida, 2) several peripherally isolated *E. nigrum* lineages including *E. susanae* and two new putative species in the Red River and Paint Rock River, and 3) three old *E. olmstedi* lineages and *E. vexillare* in Virginia and North Carolina that have diverged in parallel upstream of the Fall Line. Divergence time estimates for these four Fall Line lineages suggest that Pliocene uplift and inundation along the Atlantic slope provided a geologic mechanism for initial geographic isolation and subsequent genetic differentiation. Lastly, despite a widespread contact zone, introgression between *E. nigrum* and *E. olmstedi* is limited to the Lake Ontario basin in western New York. Patterns of allopatry and secondary contact within the *Etheostoma nigrum* species complex provide a window into the process of speciation in darters and other lineages of freshwater fishes.

CRediT authorship contribution statement

Daniel J. MacGuigan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Oliver D. Orr:** Investigation, Data curation, Writing – review & editing. **Thomas J. Near:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

NCBI SRA accession numbers for the ddRAD data are available in Table S1. NCBI GenBank accession numbers for the *cytb* sequences are availale in Table S1. Alignments, VCF files, and treefiles are available via Dryad https://doi.org/10.5061/dryad.k0p2ngfb1.

Acknowledgements

The authors thank D. Kim, A. Ghezelayagh, C.E. Parker, L.L. Bowman, E. France, M. Correa, C. Tamburri, A. Dornburg, M. Lambert, A. Mossman, C. Gottsegen, G.G. Morris, J.P. Morris, and R.J. MacGuigan for assistance in the field. G. Watkins-Colwell helped secure collecting permits and provided museum collection support. F. Janzen (NMC), S. Tessier (NMC), C. Pasbrig (SD GFP), J. Olson (IA DNA), M. Wagner (MMNS), D. Schumacher (NE DEQ), S. Schainost (NE G&P), C. Williams (VCSU), B. Tracy (NC DEQ), G. Hogue (NCSM), and M. Pinder (VA DGIF) provided tissue samples for this work. Finally, the authors thank members of the Near, Donoghue, and Muñoz labs for helpful feedback.

Funding

This work was supported by the Bingham Oceanographic Fund maintained by the Yale Peabody Museum; the Yale Peabody Museum Ginsberg Fund; and the Yale Institute for Biospheric Studies Small Grants Program. DJM was supported by the Yale Training Program in Genetics [National Institute of Health grant number T32 GM007499].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107645.

References

Abell, R.A., Olson, D.M., Dinerstein, E., Eichbaum, W., Hurley, P., Diggs, J.T., Walters, S., Wettengel, W., Allnutt, T., Loucks, C.J. and Hedao, P., 2000. Freshwater Ecoregions of North America:A Conservation Assessment, vol. 2. Island Press.

Allen, J.A., 1907. Mutations and the geographic distribution of nearly related species in plants and animals. Am. Nat. 41, 653–655.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

April, J., Mayden, R.L., Hanner, R.H., Bernatchez, L., 2011. Genetic calibration of species diversity among North America's freshwater fishes. Proc. Natl. Acad. Sci. 108, 10602–10607.

Baum, G.R., Harris, W.B., Zullo, V.A., 1979. Structural and stratigraphic framework for the Coastal Plain of North Carolina. F. Trip Guid. Carolina Geol. Soc.

Beamer, D.A., Lamb, T., 2008. Dusky salamanders (Desmognathus, Plethodontidae) from the Coastal Plain: Multiple independent lineages and their bearing on the molecular phylogeny of the genus. Mol. Phylogenet. Evol. 47, 143–153.

Bishop, P., 1995. Drainage rearrangement by river capture, beheading and diversion. Prog. Phys. Geogr. 19, 449–473.

Bossu, C.M., Near, T.J., 2009. Gene trees reveal repeated instances of mitochondrial DNA Introgression in orangethroat darters (percidae: etheostoma). Syst. Biol. 58, 114–129.

Bouckaert, R., Vaughan, T.G., Fourment, M., Gavryushkina, A., Heled, J., Denise, K., De Maio, N., Matschiner, M., Ogilvie, H., Plessis, L., Popinga, A., 2019. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 15.

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. Mol. Biol. Evol. 29, 1917–1932.

Burbrink, F.T., Ruane, S., 2021. Contemporary philosophy and methods for studying speciation and delimiting species. Ichthyol. Herpetol. 109, 874–894.

Carlston, C.W., 1969. Longitudinal slope characteristics of rivers of the midcontinent and the Atlantic east gulf slopes. Int. Assoc. Sci. Hydrol. Bull. 14, 21–31.

Caye, K., Deist, T.M., Martins, H., Michel, O., François, O., 2016. TESS3: Fast inference of spatial population structure and genome scans for selection. Mol. Ecol. Resour. 16, 540–548.

Caye, K., Jay, F., Michel, O., François, O., 2018. Fast inference of individual admixture coefficients using geographic data. Ann. Appl. Stat. 12, 586–608.

Ceas, P.A., Page, L.M., 1997. Systematic studies of the Etheostoma spectabile complex (Percidae; subgenus Oligocephalus), with descriptions of four new species. Copeia 496–522.

Cole, C.F., 1957. The taxonomy of the percid fishes of the genus *Etheostoma*, subgenus *Boleosoma*, of eastern United States. Cornell University.

Cole, C.F., 1967. A study of the Eastern Johnny Darter, *Etheostoma olmstedi* Storer (Teleostei, Percidae). Chesap. Sci. 8.

Cole, C.F., 1971. Status of the Darters, *Etheostoma nigrum, E. longimanum,* and *E. podostemone* in Atlantic Drainages (Teleostei, Percidae, Subgenus *Boleosoma*). In: The Distributional History of the Biota of the Southern Appalachians. Part III: Vertebrates. Research Division Monograph 4, Virginia Polytechnic Institute, Blacksburg, Virginia, pp. 119–138.

Collen, B., Whitton, F., Dyer, E.E., Baillie, J.E.M., Cumberlidge, N., Darwall, W.R.T., Pollock, C., Richman, N.I., Soulsby, A.M., Böhm, M., 2014. Global patterns of freshwater species diversity, threat and endemism. Glob. Ecol. Biogeogr. 23, 40–51.
Collette, B.B., 1962. The swamp darters of the subgenus *Hololepis* (Pisces, Percidae). Tulane Stud. Zool.
Cope, E.D., 1870. On Some Etheostomine Perch from Tennessee and North Carolina. Proc. Am. Philos. Soc. 11, 261–270.
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, 2156–2158.
Davis, C.D., Epps, C.W., Flitcroft, R.L., Banks, M.A., 2018. Refining and defining riverscape genetics: How rivers influence population genetic structure. Wiley Interdiscip. Rev. Water. 5, e1269.
Dowsett, H.J., Cronin, T.M., 1990. High eustatic sea level during the middle Pliocene: Evidence from the southeastern U.S. Atlantic Coastal Plain. Geology 18, 435–438.
Eaton, D.A.R., Overcast, I., 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36, 2592–2594.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Elkins, D., Sweat, S.C., Kuhajda, B.R., George, A.L., Hill, K.S., Wenger, S.J., 2019. Illuminating hotspots of imperiled aquatic biodiversity in the southeastern US. Glob. Ecol. Conserv. 19, e00654.
Engle, V., Summers, K., 1999. Latitudinal gradients in benthic community composition in Western Atlantic estuaries. J. Biogeogr. 26, 1007–1023.
Etnier, D.E., Starnes, W.C., 1993. The Fishes of Tennessee. The University of Tennessee Press, Knoxville, USA.
Fenneman, N.M., 1914. Physiographic Boundaries within the United States. Ann. Assoc. Am. Geogr. 4, 84–134.
Fenneman, N.M., 1938. Physiography of the eastern United States, illustrated. McGraw-Hill Book Company, New York and London.
Frichot, E., François, O., 2015. LEA: An R package for landscape and ecological association studies. Methods Ecol. Evol. 6, 925–929.
Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., François, O., 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196, 973–983.
Hack, J.T., 1982. Physiographic divisions and differential uplift in the Piedmont and Blue Ridge. USGPO.
Harris, W.B., Zullo, V.A., Baum, G.R., 1979. Tectonic effects on Cretaceous, Paleogene, and early Neogene sedimentation, North Carolina. Struct. Stratigr. Framew. Coast. Plain North Carolina Carolina Geol. Soc. F. Trip Guideb., pp. 17–29.
Hayes, M.M., Piller, K.R., 2018. Patterns of diversification in a North American endemic fish, the Blackbanded Darter (Perciformes, Percidae). Zool. Scr. 47, 477–485.
Heckman, K.L., Near, T.J., Alonzo, S.H., 2009. Phylogenetic relationships among *Boleosoma* darter species (Percidae: Etheostoma). Mol. Phylogenet. Evol. 53, 249–257.
Hernandez-Martich, J.D., Smith, M.H., 1997. Downstream gene flow and genetic structure of *Gambusia holbrooki* (eastern mosquitofish) populations. Heredity 79, 295–301.
Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522.
Hocutt, C.H., Jenkins, R.E., Stauffer, J.R., 1986. Zoogeography of the fishes of the Central Appalachians and central Atalantic Coastal Plain. In: The Zoogeography of North American Freshwater Fishes, pp. 161–211.
Hollingsworth Jr, P.R., Near, T.J., 2009. Temporal patterns of diversification and microendemism in Eastern Highland endemic barcheek darters (Percidae: Etheostomatinae). Evolution 63, 228–243.
Hubbs, C.L., Raney, E.C., 1946. Endemic Fish Fauna of Lake Waccamaw, North Carolina. Misc. Publ. Univ. Michigan Museum Zool, pp. 65.
Jenkins, R.E., Burkhead, N.M., 1993. Freshwater Fishes of Virginia. American Fisheries Society, Bethesda, MD.
Jenkins, R.E., Schwartz, F.J., Lachner, E. A., 1972. Fishes of the central Appalachian drainages: their distribution and dispersal. In: Hold, P.C. (Ed.), Distributional History of the Biota of the Southern Appalachians, Part 3, Vertebrates. Virginia Polytechnic Institute and State University, pp. 436–117.
Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Jombart, T., Collins, C., 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet. Rvignette. https://doi.org/10.1038/72708.
Jordan, D.S., 1880. Description of new species of North American fishes. Proc. United States Natl. Museum 2, 235–241.
Jordan, D.S., 1888. Bullet. United States Fish Comm. 8, 104–105.
Jordan, D.S., 1905. Guide to the Study of Fishes, vol. I. Henry Hold and Company, New York, USA.
Jordan, D.S., Swain, J.S., 1883. List of fishes collected in the Clear Fork of the Cumberland, Whitley County, Kentucky, with descriptions of three new species. Proc. United States Natl. Museum 6, 248–251.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
Keck, B.P., Near, T.J., 2010. A young clade repeating an old pattern: Diversity in *Nothonotus* darters (Teleostei: Percidae) endemic to the Cumberland River. Mol. Ecol. 19, 5030–5042.
Kim, D., Bauer, B.H., Near, T.J., 2022. Introgression and Species Delimitation in the Longear Sunfish *Lepomis megalotis* (Teleostei: Percomorpha: Centrarchidae). Syst. Biol. 71, 273–285.
Kominz, M.A., Browning, J.V., Miller, K.G., Sugarman, P.J., Mizintseva, S., Scotese, C.R., 2008. Late Cretaceous to Miocene sea-level estimates from the New Jersey and Delaware coastal plain coreholes: An error analysis. Basin Res. 20, 211–226.
Krabbenhoft, T.J., Collyer, M.L., Quattro, J.M., 2009. Differing evolutionary patterns underlie convergence on elongate morphology in endemic fishes of Lake Waccamaw, North Carolina. Biol. J. Linn. Soc. 98, 636–645.
Krantz, D.E., 1991. A chronology of Pliocene sea-level fluctuations: The U.S. Middle Atlantic Coastal Plain record. Quat. Sci. Rev. 10, 163–174.
Krotzer, M.J., 1990. Variation and systematics of *Etheostoma nigrum* Rafineque. The Johnny Darter (Pices: Percidae). University of Tennessee.
Kuehne, R.A., Bailey, R.M., 1961. Stream Capture and the Distribution of the Percid Fish *Etheostoma sagitta*, with Geologic and Taxonomic Considerations. Copeia 1661, 1–8.
Lang, N.J., Echelle, A.A., 2011. Novel Phylogeographic Patterns in a Lowland Fish, *Etheostoma proeliare* (Percidae). Southeast. Nat. 10, 133–144.
Leidy, R.A., 1992. Microhabitat selection by the johnny darter, *Etheostoma nigrum* Rafinesque, in a Wyoming stream. Great Basin Natl. 68–74.
Linck, E., Battey, C.J., 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol. Ecol. Resour. 19, 639–647.
Lujan, N.K., Weir, J.T., Noonan, B.P., Lovejoy, N.R., Mandrak, N.E., 2020. Is Niagara Falls a barrier to gene flow in riverine fishes? A test using genome-wide SNP data from seven native species. Mol. Ecol. 29, 1235–1249.
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. Ann. Missouri Bot. Gard. 87, 26–62.
MacGuigan, D.J., 2020. Speciation and Hybridization Across Multiple Temporal Scales in Darters (Teleostei: Percidae). Yale University. Doctoral dissertation.
MacGuigan, D.J., Near, T.J., 2019. Phylogenomic Signatures of Ancient Introgression in a Rogue Lineage of Darters (Teleostei: Percidae). Syst. Biol. 68, 329–346.
MacGuigan, D.J., Hoagstrom, C.W., Domisch, S., Hulsey, C.D., Near, T.J., 2021. Integrative ichthyological species delimitation in the Greenthroat Darter complex (Percidae: Etheostomatinae). Zool. Scr. 50, 707–733.
Markewich, H., 1985. Geomorphic evidence for Pliocene-Pleistocene uplift in the area of the Cape Fear Arch, North Carolina. In: Morisawa, M., Hack, J.T. (Eds.), Tectonic Geomorphology. Allen and Unwin, Boston, pp. 279–297.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10–12.
Matthews, W.J., Robison, H.W., 1982. Addition of *Etheostoma collettei* (Percidae) to the Fish Fauna of Oklahoma and of the Red River Drainage in Arkansas. Southwest. Nat. 27, 215–216.
Mayden, R.L., 1985. Biogeography of Ouachita Highland Fishes. Southwest. Nat. 30, 195–211.
Mayden, R, 1988. Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. Syst. Zool. 37 (4), 329–355.
Miller, K.G., Kominz, M.A., Browning, J.V., Wright, J.D., Mountain, G.S., Katz, M.E., Sugarman, P.J., Cramer, B.S., Christie-Blick, N., Pekar, S.F., 2005. The phanerozoic record of global sea-level change. Science 310, 1293–1298.
Miller, K.G., Wright, J.D., Browning, J.V., Kulpecz, A., Kominz, M., Naish, T.R., Cramer, B.S., Rosenthal, Y., Peltier, W.R., Sosdian, S., 2012. High tide of the warm Pliocene: Implications of global sea level for Antarctic deglaciation. Geology 40, 407–410.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.
Near, T.J., Kassler, T.W., Koppelman, J.B., Dillman, C.B., Philipp, D.P., 2003. Speciation in North American black basses, Micropterus (Actinopterygii: Centrarchidae). Evolution 57, 1610–1621.
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). Syst. Biol. 60, 565–595.
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). Mol. Phylogenet. Evol. 66, 868–876.
Near, T., Porterfield, J., Page, L., McEachran, J., 2000. Evolution of cytochrome b and the molecular systematics of *Ammocrypta* (Percidae: Etheostomatinae). Copeia 3, 701–711.
O'Bara, C.J., 1991. Current Distribution and Status of the Upper Cumberland River Johnny Darter, Etheostoma nigrum susanae. J. Tennessee Acad. Sci. 66.
Page, L.M., Burr, B.M., 2011. Peterson Field Guide to Freshwater Fishes of North America North of Mexico, Second Edition. Houghton Mifflin Harcourt Publishing Company, New York, New York.
Page, L.M., 1983. Handbook of Darters. T.F.H Publications, Neptune City, NJ.
Pauly, G.B., Piskurek, O., Shaffer, H.B., 2007. Phylogeographic concordance in the southeastern United States: The flatwoods salamander, *Ambystoma cingulatum*, as a test case. Mol. Ecol. 16, 415–429.
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, e37135.
Pickrell, J.K., Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. Nat Prec. 1-1.
Piller, K.R., Bart, H.L., Hurley, D.L., 2008. Phylogeography of the greenside darter complex, Etheostoma blennioides (Teleostomi: Percidae): A wide-ranging polytypic taxon. Mol. Phylogenet. Evol. 46, 974–985.
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, e32253.

Ray, J.M., Wood, R.M., Simons, A.M., 2006. Phylogeography and post-glacial colonization patterns of the rainbow darter, Etheostoma caeruleum (Teleostei: Percidae). J. Biogeogr. 33, 1550–1558.
Raymo, M.E., Grant, B., Horowitz, M., Rau, G.H., 1996. Mid-Pliocene warmth: Stronger greenhouse and stronger conveyor. Mar. Micropaleontol. 27, 313–326.
Renner, G.T., 1927. The Physiographic Interpretation of the Fall Line. Geogr. Rev. 17.
Rovere, A., Hearty, P.J., Austermann, J., Mitrovica, J.X., Gale, J., Moucha, R., Forte, A. M., Raymo, M.E., 2015. Mid-Pliocene shorelines of the us Atlantic coastal plain - an improved elevation database with comparison to earth model predictions. Earth Sci. Rev. 145, 117–131.
Seehausen, O., 2006. African cichlid fish: A model system in adaptive radiation research. Proc. Royal Soc. B: Biol. Sci. 273, 1987–1998.
Seehausen, O., Wagner, C.E., 2014. Speciation in Freshwater Fishes. Annu. Rev. Ecol. Evol. Syst. 45, 621–651.
Selkoe, K.A., Scribner, K.T., Galindo, H.M., 2015. Waterscape Genetics - Applications of Landscape Genetics to Rivers, Lakes, and Seas. P. In: Landscape Genetics: Concepts, Methods, Applications, vol. 1, pp. 264.
Shankman, D., Hart, J.L., 2007. The Fall Line: a Physiographic-Forest Vegetation Boundary. Geogr. Rev. 97, 502–519.
Stager, J.C., Cahoon, L.B., 1987. The age and trophic history of Lake Waccamaw, North Carolina. J. Elisha Mitchell Sci. Soc. 103, 1–13.
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. Syst. Biol. 67, 681–699.
Starnes, W.C., Starnes, L.B., 1978. A New Cyprinid of the Genus *Phoxinus* Endemic to the Upper Cumberland River Drainage. Copeia 508–516.
Starnes, W.C., Starnes, L.B., 1979. Taxonomic status of the percid fish *Etheostoma nigrum susanae*. Copeia 1979, 430.
Strange, R.M., 1998. Mitochondrial DNA Variation in Johnny Darters (Pisces: Percidae) from Eastern Kentucky Supports Stream Capture for the Origin of Upper Cumberland River Fishes. Am. Midl. Nat. 140, 96–102.
Taylor, A.T., Tringali, M.D., Sammons, S.M., Ingram, T.R., O'Rouke, P.M., Peterson, D.L., Long, J.M., 2018. Genetic population structure of Shoal Bass within their native range. North Am. J. Fish. Manag. 38, 549–564.
Tight, W.G., 1903. Drainage modifications in southeastern Ohio and adjacent parts of West Virginia and Kentucky. Washington, D.C.
Tinkle, D.W., 1959. The Relation of the Fall Line to the Distribution and Abundance of Turtles. Copeia 1959, 167–170.
Underhill, J.C., 1963. Distribution in Minnesota of the Subspecies of the Percid Fish *Etheostoma nigrum*, and of Their Intergrades. Am. Midl. Nat. 470–478.
Van De Plassche, O., Wright, A.J., Horton, B.P., Engelhart, S.E., Kemp, A.C., Mallinson, D., Kopp, R.E., 2014. Estimating tectonic uplift of the Cape Fear Arch (south-eastern United States) using reconstructions of Holocene relative sea level. J. Quat. Sci. 29, 749–759.
Ward, J.V., Stanford, J.A., 1983. Serial Discontinuity Concept of Lotic Ecosystems.
Warren Jr., M.L., Burr, B.M., Walsh, S.J., Bart Jr., H.L., Cashner, R.C., Etnier, D.A., Freeman, B.J., Kuhajda, B.R., Mayden, R.L., Robison, H.W., Ross, S.T., Starnes, W.C., 2000. Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. Fisheries 25, 7–31.
Wheeler, R.L., 2006. Quaternary tectonic faulting in the Eastern United States. Eng. Geol. 82, 165–186.
Wiley, A.E.O., Mayden, R.L., 1985. Species and Speciation in Phylogenetic Systematics, with Examples from the North American Fish Fauna. Ann. Missouri Bot. Gard. 72, 596–635.
Williams, J.D., Robins, C.R., 1970. Variation in Populations of the fish *Cottus carolinae* in the Alabama River System with Description of a New Subspecies from below the Fall Line. Am. Midl. Nat. 368–381.
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 Ecol. Evol. 8, 28–36.
Zorach, T., 1971. Taxonomic Status of the Subspecies of the Tessellated Darter, *Etheostoma olmstedi* Storer, in Southeastern Virginia. Chesapeake Sci. 12, 254–263.