

Lacustrine speciation associated with chromosomal inversion in a lineage of riverine fishes

Daniel J. MacGuigan¹, Trevor J. Krabbenhoft^{1,2}, Richard C. Harrington³, Dylan K. Wainwright⁴, Nathan J. C. Backenstose¹, Thomas J. Near^{3,5}

¹Department of Biological Sciences, University at Buffalo, Buffalo, NY, United States

²RENEW Institute, University at Buffalo, Buffalo, NY, United States

³Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, United States

⁴Department of Biological Sciences, Purdue University, West Lafayette, IN, United States

⁵Yale Peabody Museum of Natural History, New Haven, CT, United States

Corresponding author: Department of Biological Sciences, University at Buffalo, 109 Cooke Hall, Buffalo, NY 14260, United States. Email: dmacguig@buffalo.edu

Abstract

Geographic isolation is the primary driver of speciation in many vertebrate lineages. This trend is exemplified by North American darters, a clade of freshwater fishes where nearly all sister species pairs are allopatric and separated by millions of years of divergence. One of the only exceptions is the Lake Waccamaw endemic *Etheostoma perlongum* and its riverine sister species *Etheostoma maculatiiceps*, which have no physical barriers to gene flow. Here we show that lacustrine speciation of *E. perlongum* is characterized by morphological and ecological divergence likely facilitated by a large chromosomal inversion. While *E. perlongum* is phylogenetically nested within the geographically widespread *E. maculatiiceps*, there is a sharp genetic and morphological break coinciding with the lake–river boundary in the Waccamaw River system. Despite recent divergence, an active hybrid zone, and ongoing gene flow, analyses using a de novo reference genome reveal a 9 Mb chromosomal inversion with elevated divergence between *E. perlongum* and *E. maculatiiceps*. This region exhibits striking synteny with known inversion supergenes in two distantly related fish lineages, suggesting deep evolutionary convergence of genomic architecture. Our results illustrate that rapid, ecological speciation with gene flow is possible even in lineages where geographic isolation is the dominant mechanism of speciation.

Keywords: speciation, structural variation, evolutionary genomics, population genetics, ecomorphology, *Etheostoma*

Introduction

Geographic isolation historically dominated the discussion of speciation mechanisms (Coyne & Orr, 2004). Genetic drift of incompatible alleles in geographically isolated lineages can drive speciation without the need to invoke natural selection (Orr, 1996). Models of ecological speciation that did involve selection typically included an initial “allopatric phase” accompanied by divergent ecological selection in generating barriers to gene flow (Schluter, 2001). Under this classic model of ecological speciation, geographic isolation initially limits gene flow and allows lineages to accumulate ecological differences, which ultimately confer reproductive isolation (Rundle & Nosil, 2005). In theory, speciation can occur without geographic barriers to gene flow; however, allopatry remained an important component of ecological speciation models. For instance, Ernst Mayr largely dismissed the idea that environmental breaks alone could lead to speciation, stating “no case is known to us that could be cited as irrefutable proof” (Mayr, 1942). This mindset percolated into many subdisciplines of biology, especially freshwater ichthyology where some concluded that there are no known examples of parapatric or sympatric speciation that show an “ecological mode,” in part because ichthyologists “tend to think ‘allopatrically’” (Wiley & Mayden, 1985).

Recently, a growing body of theoretical and empirical literature demonstrates that ecological speciation without geographic isolation limiting gene flow is possible, particularly in lakes where heterogeneous and highly structured environments can drive adaptive diversification (Seehausen & Wagner, 2014). Lakes serve as biogeographic “islands” that geographically and ecologically isolate emerging species from nearby freshwaters, such as an endemic minnow in Lake Biwa, Japan (Kakioka et al., 2015), a species complex of ricefishes in Sulawesi (Mokodongan et al., 2018), sailfin silversides in Lake Matano, Indonesia (Schwarzer et al., 2008), and Arctic Charr in many postglacial lakes (Jonsson & Jonsson, 2001). Despite myriad ecological speciation studies in lakes, taxonomy often lags behind. For instance, many lake-stream pairs of threespine sticklebacks are in the early stages of parapatric, ecological speciation and the diverse stickleback morphotypes are often considered a “complex of species” (Schluter & Conte, 2009), yet no taxonomic revision has ever been proposed. Discrepancy between speciation studies and taxonomy can have conservation consequences and lead to underestimates of lacustrine biodiversity (von Hippel, 2008).

One of the rare cases where taxonomy has outpaced speciation research is Lake Waccamaw, one of the largest Carolina

Received December 15, 2022; revisions received March 23, 2023; accepted April 20, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE).

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Bay lakes. This waterbody exemplifies the insular nature of lacustrine ecosystems. Lake Waccamaw is a shallow, sandy lake (max depth 3.3 m), with neutral pH, high alkalinity, clear waters, and open habitat compared with the surrounding acidic, highly structured blackwater habitats of the Waccamaw River basin (Frey, 1949, 1951; Shute et al., 1981; Stager & Cahoon, 1987). Although Lake Waccamaw is only 15,000–32,000 years old (Stager & Cahoon, 1987), it contains three phylogenetically disparate endemic fish species: the Waccamaw Silverside (*Menidia extensa*), the Waccamaw Killifish (*Fundulus waccamensis*), and the Waccamaw Darter (*Etheostoma perlongum*; Hubbs & Raney, 1946), as well as two endemic mussels and an endemic snail (Johnson et al., 2013; McCartney et al., 2016). Morphological differences between the Lake Waccamaw endemics and their closest riverine relatives suggest lacustrine adaptation. Each lake species has more vertebrae, more lateral line scales, and more slender, elongated body shapes than their riverine counterparts (Hubbs & Raney, 1946; Krabbenhoft et al., 2009).

Perhaps most unusual among the Lake Waccamaw endemics is the Waccamaw Darter, *E. perlongum*. Of the approximately 250 species of darters, 93% of sister species pairs are allopatric (Near et al., 2011), suggesting that geographic isolation is a key component of speciation in this clade. In contrast, there is no geographic isolation between *E. perlongum* and its closest relative, the riverine *Etheostoma maculataiceps* (formerly *E. olmstedii maculataiceps*) (MacGuigan et al., 2023). Additionally, while most darter species exclusively occupy riverine habitats, *E. perlongum* is the only entirely lacustrine species. Although some darter species exhibit intra-specific lake–river morphological differences (Oliveira et al., 2021), *E. perlongum* is the only potential case of lacustrine speciation in the clade. Lack of evidence for strong genetic differentiation within the Waccamaw River basin has led to questions about the distinctiveness of *E. perlongum* as a species (McCartney & Barreto, 2010; Williams et al., 1989).

In this study, we investigate the ecological, morphological, and genomic context of speciation between *E. perlongum* in Lake Waccamaw and *E. maculataiceps* in the adjacent Waccamaw River. We assemble and annotate a de novo chromosome-level genome for *E. perlongum* and analyze thousands of double digest restriction site-associated DNA (ddRAD) markers from individuals spanning a lake–river transect. Using this rich genomic data set, we determine if there are regions of the genome resistant to gene flow that potentially reveal the genomic architecture of speciation between *E. perlongum* and *E. maculataiceps*. We complement genomic analyses with morphometric analyses of body shape, analyses of ecomorphological osteology with microcomputed tomography (μ CT) imaging, and diet analysis of stomach contents.

Our analyses reveal genomic, morphological, and ecological differentiation between darters from Lake Waccamaw and the Waccamaw River. We detect substantial, ongoing gene flow between *E. perlongum* and *E. maculataiceps* facilitated by an active hybrid zone in the Waccamaw River just downstream of Lake Waccamaw. Despite gene flow, we find elevated genomic divergence between *E. perlongum* and *E. maculataiceps* localized in a 9 Mb region of chromosome 9 with high levels of linkage disequilibrium, indicative of a chromosomal inversion. Gene annotations suggest that this inversion may be a supergene involved in retinal development, olfactory signaling, and circadian regulation. We hypothesize

that the chromosome 9 inversion is subject to differential selective pressures in lake and river environments driving ecological and morphological divergence. This darter inversion is similar in gene content to supergene inversions previously identified in two distantly related fish species, suggesting that deep evolutionary convergence of chromosomal inversions can repeatedly facilitate local adaptation. Our results reveal that in a clade where diversification is primarily the product of geographic isolation, modifications to genomic architecture can enable ecological divergence and speciation in the face of gene flow.

Materials and methods

Sampling and ddRAD sequencing

We collected tissue samples for 93 individuals along a transect encompassing five sites in Lake Waccamaw and seven sites in the Waccamaw River basin (Figure 1A, Supplementary Table S1). In addition, we sampled 43 individuals from 28 localities across the range of *E. maculataiceps*, the closest relative of *E. perlongum*, and two individuals from each of two outgroup species, *E. olmstedii* and *E. nigrum* (Supplementary Table S1). Following Yale IACUC protocol number 2018–10681, fish were euthanized by immersion in 250 mg/L buffered MS-222 solution before collecting tissue samples, typically from the right pectoral fin. Tissues were preserved in 95% ethanol at -20°C . We performed a modified version of a ddRAD (Peterson et al., 2012; Poland et al., 2012). Details of DNA extraction, library preparation, and assembly may be found in the Supplementary Material and Supplementary Table S2.

A de novo darter genome

We sequenced and assembled a de novo genome for *E. perlongum*. One *E. perlongum* individual (YFTC 37806, YPM ICH 033096) was euthanized with MS-222 and immediately dissected. We isolated fin, muscle, gill, liver, brain, eye, heart, gut, and testes tissue. Tissues were immediately frozen with liquid nitrogen and stored at -80°C . For genome sequencing, we extracted DNA from muscle and gill tissue using a QIAGEN Genomic-tip 500/G kit. We used Oxford Nanopore long-read technology to sequence the *E. perlongum* genome. Prior to sequencing, we performed a size selection using the Circulomics Short Read Eliminator Kit. Sequencing libraries were prepared using 1 μg of gill or muscle genomic DNA and the Oxford Nanopore LSK-109 ligation sequencing kit. We also performed short-read Illumina sequencing with a 10x Genomics Chromium library preparation, allowing us to polish error-prone Nanopore reads with accurate short reads. The Nanopore data were assembled into contigs and scaffolded using Hi-C chromatin conformation capture sequencing (see Supplementary Material for details).

To generate evidence for gene annotation, we performed transcriptome sequencing using eight different tissues (gill, fin, gut, heart, liver, brain, eye, and testes). We combined equimolar amounts of RNA from each tissue type for transcriptome sequencing performed by NovoGene (<https://en.novogene.com/>). We used the Maker v.2.3.10 pipeline to assign putative functional annotations to the *E. perlongum* genome assembly (Holt & Yandell, 2011). For annotation evidence, we used the transcriptome generated in this study, exon sequences from a close relative (Daane et al., 2019), and protein sequence data from eight other teleost fishes. We performed four iterative rounds of annotation with Maker:

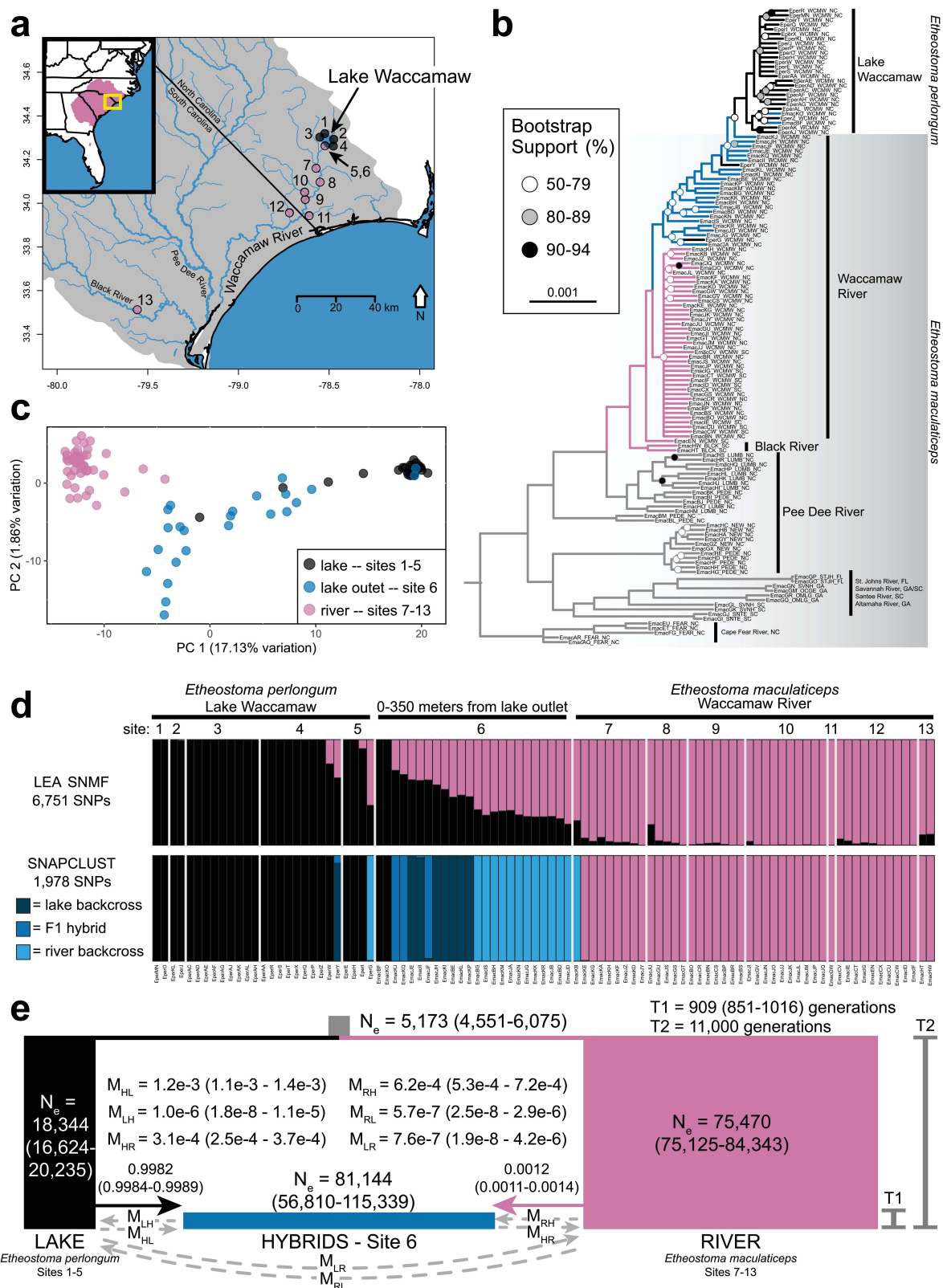


Figure 1. Evolutionary history of lake–river divergence. (A) Map of sampling localities within the Waccamaw and Pee Dee River systems. Site 5 is located within Lake Waccamaw immediately above the outlet, site 6 is 0–350 meters downstream of the lake. Insert map shows range of *Etheostoma maculatiiceps* in pink with sampling area highlighted by the yellow box. Gray area on the full map indicates the Pee Dee River basin. (B) Maximum likelihood ddRAD phylogeny of *E. maculatiiceps* and *E. perlongum*. Outgroup taxa are not shown. Nodes with <50% bootstrap support are collapsed; nodes with >95% bootstrap support are unlabeled. Branches colored according to hybrid classification results. (C) PCA of 10,628 SNPs with <10% missing data. (D) Genetic clustering analyses. Each vertical bar represents a sample, bar colors represent admixture coefficients (SNMF) or assignment probabilities (SNAPCLUST). Sampling sites are delimited by vertical white bars. (E) Parameter estimates from model with the best Akaike information criterion score (Supplementary Figure S4, Model 2, 11,000 generations of lake–river divergence). Nonparametric bootstrap 95% HDIs are indicated in parentheses.

one round of evidence-based annotation and three rounds of ab initio gene prediction. Putative functions were assigned to the predicted genes by identifying potential matches to the UniProt Swiss-Prot database and searching for protein domains with InterProScan v.5 (Jones et al., 2014). The CoGe SynMap tool (<https://genomeevolution.org/coge/>) was used to visualize synteny between the annotated *E. perlongum* genes two relatives in the family Percidae: *Etheostoma spectabile* and *Perca flavescens*. Details of the genome annotation pipeline are provided in the [Supplementary Material](#).

Linkage disequilibrium and inversion detection

Using the ddRAD assembly, we estimated linkage disequilibrium (LD) between all pairs of single-nucleotide polymorphisms (SNPs) for the largest genomic scaffolds representing the 24 chromosomes. We created a reduced data set using VCFTools v.0.1.15 (Danecek et al., 2011), retaining only biallelic SNPs (“--max-alleles 2”), SNPs with less than 50% missing data (“--max-missing 0.5”), and SNPs with minor allele frequency > 5% (“--maf 0.05”). We estimated LD using PLINK v.1.90 (Chang et al., 2015), then averaged LD estimates in 1 Mb windows across each scaffold for visualization. Additionally, following Matschiner et al. (2022), we calculated an individual SNP LD metric as the cumulative sum of distance (in base pairs) between other intrachromosomal SNPs in high LD ($R^2 > 0.8$).

We used LDna v.2.0 (Kempainen et al., 2015) to identify clusters of ddRAD loci in high LD that could represent chromosomal rearrangements. We focused on “single outlier clusters” (SOCs) of high LD, as “compound outlier clusters” comprised multiple SOCs can be the product of different evolutionary forces and are difficult to interpret (Kempainen et al., 2015). LDna analyses were performed separately for the 24 largest scaffolds. To further examine SOC-1 on chromosome 9 (see Results), we recalculated LD as described above separately for samples with at least 75% of SOC-1 SNPs homozygous for the reference (lake) or alternate (river) allele. LDna details are provided in the [Supplementary Material](#).

Phylogeny and population structure

To understand the broader evolutionary context of *E. perlongum*, we performed concatenated phylogenetic analyses of the ddRAD data using IQTree v.1.6.12 (Nguyen et al., 2015). To determine the impact of missing data on phylogenetic inference, we assembled alignments with 30%, 20%, 10%, and 5% missing data ([Supplementary Table S4](#)). We used a GTR + gamma nucleotide substitution model for all analyses. We ran each analysis until 100 unsuccessful tree search iterations were completed. To assess topological support, we also performed 1,000 ultrafast bootstrap replicates (Hoang et al., 2018).

We assembled a set of filtered SNPs for 93 individuals from the Waccamaw River system plus two individuals from the Black River (a small tributary near the mouth of the Waccamaw River) to assess population structure. Using VCFTools, we retained only biallelic SNPs (“--max-alleles 2”) and removed sites that contained more than 10% missing data (“--max-missing 0.9”). Since SNP singletons can confound inference of population structure (Linck & Battey, 2019), we also removed SNPs with minor allele frequency < 5% (“--maf 0.05”). Finally, we thinned our data set to include only a single SNP per 10,000 bp window (“--thin 10000”) to minimize the effects of physical linkage on our inference of population

structure. We used PCA to examine general population structure patterns and sparse non-negative matrix factorization (sNMF) to estimate individual ancestry coefficients with the R package LEA v.2.6.0 (Frichot & François, 2015). Based on the genetic clustering results from PCA and LEA, we used the *snaphclust* function in the R package adegenet v.2.1.3 (Beugin et al., 2018; Jombart, 2008) to classify individuals as hybrids or backcrosses. Lastly, we ran the same PCA and sNMF analysis using two different input data sets: only SNPs within the chromosome 9 inversion and all SNPs except those found in the chromosome 9 inversion (see Results). Additional details are available in the [Supplementary Material](#).

Demographic modeling

Demographic modeling using coalescent simulations was used to compare different evolutionary scenarios in the Waccamaw River basin. We used fastsimcoal2 v.2.6.0.3 (Excoffier et al., 2013) to estimate parameters from the observed site frequency spectrum. We compared three evolutionary models ([Supplementary Figure S4A](#)). All models contained three populations: *E. perlongum* in Lake Waccamaw, the hybrid population at the Lake Waccamaw outlet, and *E. maculatiiceps* in the Waccamaw River. In each model, the hybrid population is formed by an admixture event after initial divergence between Lake Waccamaw and the Waccamaw River. The models varied in the number of migration edges connecting each population. In Model 1, no gene flow was allowed between the three populations ([Supplementary Figure S4A](#)). Model 2 allows for migration only after the formation of the hybrid zone, a scenario of secondary contact following allopatric divergence ([Supplementary Figure S4A](#)). Finally, Model 3 allows continuous migration following initial divergence between Lake Waccamaw and the Waccamaw River, a scenario of parapatric divergence with gene flow ([Supplementary Figure S4A](#)). We compared model fit by calculating AIC scores. For the best-fit model, we generated 100 nonparametric bootstrap replicates and calculated the 95% highest density interval for each parameter. Details of model specification are provided in the [Supplementary Material](#).

Genetic geographic clines

Genetic clustering analyses reveal broad population patterns, but different regions of the genome may have experienced different levels of gene flow between *E. perlongum* and *E. maculatiiceps*. Therefore, we examined geographic clines in allele frequencies for 10,480 SNPs. We used the same VCFTools filtering parameters as for the population structure analyses but we did not thin SNPs by genomic position. Allele frequency clines were fit using the R package HZAR v.0.2.5 (Derryberry et al., 2014). For each SNP, we compared a null model (no difference in allele frequency between the lake and river) and three geographic cline models: a logistic curve with empirical maximum and minimum allele frequencies, a logistic curve with estimated maximum and minimum allele frequencies, and a logistic curve with estimated maximum and minimum allele frequencies and exponential decay curves at both ends. We compared these models using AICc scores. Model details are provided in the [Supplementary Material](#).

Genetic outlier detection

Genetic loci under strong selection or with reduced gene flow during speciation may manifest as strongly differentiated outliers relative to the genomic background. We examined the

ddRAD data for the presence of outlier loci using several different approaches. First, we assembled an SNP data set that excluded individuals from site 6, since this locality consisted almost entirely of admixed hybrid individuals. The remaining individuals were classified as either lake (sites 1–5, $n = 27$) or river (sites 7–13, $n = 44$). We used VCFTools to retain only biallelic SNPs (“--max-alleles 2”), remove sites that contained more than 10% missing data (“--max-missing 0.9”), and remove SNPs with minor allele frequency < 5% (“--maf 0.05”). After filtering, the data set contained 12,214 SNPs.

We first identified SNP outliers using a PCA-based approach implemented in the R package pcadapt v.4.3.3 (Luu et al., 2017). Additionally, we used two approaches to estimate the fixation index, F_{ST} , and identify outliers of high differentiation for each SNP (Wright, 1950). First, we used AMOVA F_{ST} estimation and identified outliers above the 99th percentile of differentiation (Catchen et al., 2013). Second, we used Bayesian inference to estimate F_{ST} and identify outliers after correcting for multiple comparisons (Foll & Gaggiotti, 2008). Of these methods, we expected simple percentile-based outlier detection to be the most liberal, while the Bayesian method should be the most conservative. Since variation in sequencing depth can lead to imprecise F_{ST} estimates for individual SNPs (Beissinger et al., 2015; Cruickshank & Hahn, 2014), we also estimated relative divergence (F_{ST}) and absolute divergence (D_{XY}) in 1 Mb sliding windows with a slide size of 100 kb. For the sliding window estimates, regions above the 99th percentile of differentiation were designated as outliers. Additionally, we estimated nucleotide diversity (π) using the same sliding window approach. See [Supplementary Material](#) for details.

Gene ontology enrichment analysis

To determine whether any gene ontologies (GO) were over-represented in regions of elevated divergence, we performed GO enrichment analysis in three genomic windows (F_{ST} plateau, D_{XY} peak 1, and D_{XY} peak 2, [Figure 4A](#)). First, we assembled a reference database of amino acid sequences from three closely related species: *Larimichthys crocea* (Ensembl L_crocea_2.0), *Gasterosteus aculeatus* (Ensembl BROAD S1), and *Cottoperca gobio* (Ensembl fCotGob3.1). We used BlastX (NCBI) to perform a search against this reference database using the *E. perlongum* predicted transcriptome as the query sequence. We then used the GOATOOLS *find_enrichment.py* script to identify GO terms that were overrepresented in genes found within outlier regions (Klopfenstein et al., 2018). REVIGO (Supek et al., 2011) was used to cluster redundant GO terms with 50% allowed SimRel similarity. Additional details in the [Supplementary Material](#).

Morphology

We used geometric morphometrics to assess the degree of body shape divergence between populations from Lake Waccamaw and the Waccamaw River. In total, we digitized a series of 16 landmarks and 24 sliding semilandmarks ([Figure 2A](#)) for 240 specimens using tpsDIG v.2.31 (Rohlf, 2006). We performed principal component analysis of the landmark data using the PCA function in the R package Momocs v.1.2.9 (Bonhomme et al., 2014) and visualized morphospace with the Momocs *plot.PCA* function. In addition to the landmark analysis, we took four linear morphological measurements related to feeding ecology, swimming performance, and predator avoidance (Carroll et al., 2004; Gosline, 1997): head length, lower jaw

length, body depth, and caudal length ([Figure 2A](#)). We analyzed geographic variation in the second principal component axis ([Figure 2B](#)), body depth, caudal length, head length, and upper jaw length ([Figure 2A](#)) using a cline-fitting approach in the R package HZAR. The last four traits were corrected for body size by taking the residual versus standard length ([Figure 2A](#)). We performed cline fitting as described for the SNP data.

To further investigate ecomorphological patterns among river, lake, and hybrid-zone populations, we made linear measurements of the skeleton from μ CT scans of a subset of 18 individuals (eight from the lake, six from the river, and four from the hybrid-zone). We made 10 measurements that have been previously used to study darter comparative morphology (Carlson & Wainwright, 2010) and are expected to differ with changes in feeding ecology and diet. We also used our measurements to calculate two ratios, including jaw opening-lever ratio. Finally, we collected a series of meristic trait measurements that are traditionally used to discover and delimit species in ichthyology. We compared morphometric, osteological, and meristic trait distributions among the lake, lake outlet, and river populations using ANOVA. For meristic and osteological traits, we also examined the correlation with lake–river genomic ancestry estimated from the SNP data set. Additional details of measurements and analyses are found in the [Supplementary Material](#).

Diet

We assessed the trophic niche of Lake Waccamaw and Waccamaw River populations by collecting diet data from gut contents of specimens of *E. perlongum* ($n = 22$), the hybrid zone ($n = 5$), and *E. maculaticeps* ($n = 27$). Contents of the foregut were identified to ordinal or family level, where possible. We estimated multivariate trophic ecospace using non-metric multidimensional scaling (NMDS) and visualized the distribution of prey items with a bipartite network. Additional details are provided in the [Supplementary Material](#).

Results

Genome assembly

We produced a de novo chromosome-level genome assembly of *E. perlongum* using Nanopore long reads polished with high-quality Illumina short reads and scaffolded with a Hi-C contact map ([Supplementary Table S3](#)). We generated 29.85 Gb of Nanopore sequence data, with a median read length of 2,634 bp and a read length N50 of 10,459 bp. One percentage of the Nanopore reads are longer than 100 kbp, with a maximum length of 301,807 bp. Based on the genome assembly size of another *Etheostoma* species (Moran et al., 2020), our Nanopore sequence data represents approximately 35X genomic coverage for *E. perlongum*. The total length of the 150 bp Illumina paired-end reads used for error correction is 111.26 Gb, representing an additional approximately 130X genomic coverage.

The total size of the scaffolded *E. perlongum* genome assembly is 788 Mb, slightly smaller than other assembled percid genomes (958 Mb, *Perca fluviatilis* [Ozerov et al., 2018]; 877 Mb, *P. flavescens* [Feron et al., 2020]; 855 Mb, *Etheostoma spectabile* [Moran et al., 2020]). The *E. perlongum* genome assembly comprised 2,095 contigs (N50 = 2.3 Mb, L50 = 100 contigs) grouped into 1,179 scaffolds

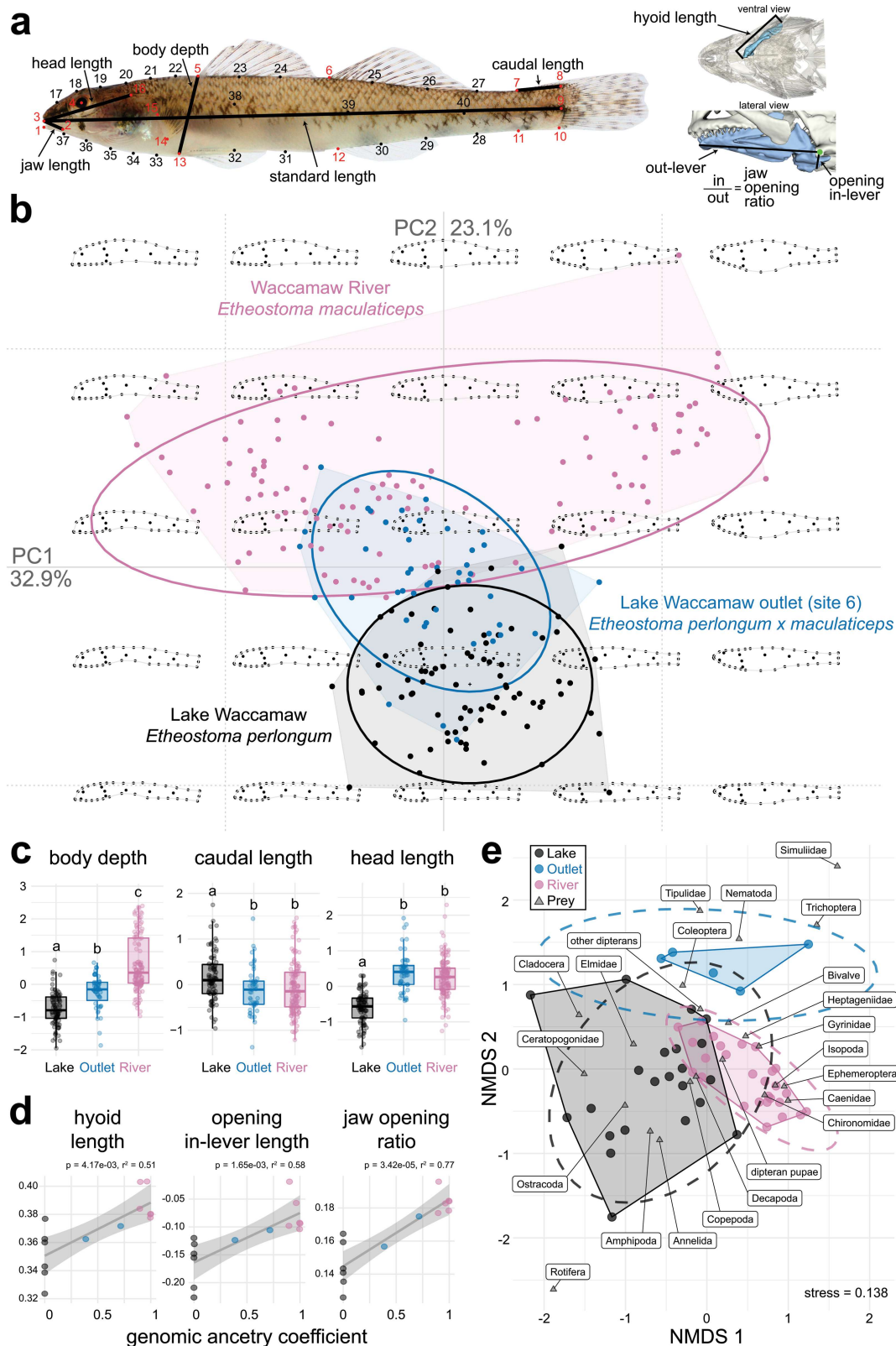


Figure 2. Morphological and ecological divergence associated with genomic divergence. (A) Position of 16 landmarks (red) and 24 sliding semilandmarks (black) and the linear measurements. (B) Morphospace of first and second principal components of the geometric morphometric analysis. Filled regions are convex hulls with 95% credible interval ellipsoids around the mean for each group. Gray outlines represent body shapes in different regions of morphospace. (C) Boxplots of PC2 scores plus linear regression residuals of caudal length and head length versus standard length. Boxplots center line shows the median; box limits show 25th and 75th quartiles; whiskers show 1.5× interquartile range; all data points shown. Compact letter displays of Tukey HSD tests indicated above each plot. (D) Linear regressions of osteological traits versus genomic ancestry coefficients estimated by SNMF (Figure 1D). Points are colored by sampling location (lake, outlet, or river). (E) Results of non-metric multidimensional scaling (NMDS) with fish stomach contents. Circles indicate fish specimens (colored by sampling location), while triangles represent prey categories. Filled regions are convex hulls with 95% credible interval ellipsoids around the mean for each group.

(N50 = 30.8 Mb, L50 = 11 scaffolds). The 24 longest scaffolds account for 92.4% of the total assembly length and exhibit a 1:1 correspondence with *P. flavescens* and *E. spectabile* chromosomes (Supplementary Figure S1). Hereafter, we refer to the 24 longest *E. perlongum* scaffolds as chromosomes. The *E. perlongum* genome assembly exhibits conserved synteny with the *P. flavescens* genome (Feron et al., 2020) and, to a lesser extent, the *E. spectabile* genome (Moran et al., 2020; Supplementary Figure S1).

96.7% of BUSCO Actinopterygii orthologs are represented as complete sequences in the *E. perlongum* genome assembly, with an additional 1% represented as partial sequences. The annotation pipeline identifies 42% (328.6 Mb) of the *E. perlongum* assembly as repetitive DNA, a larger proportion than the 30%–33% of repetitive DNA identified in other percid genomes (Feron et al., 2020; Moran et al., 2020; Ozerov et al., 2018). Total GC content of the *E. perlongum* genome is 40.5%. The *E. perlongum* transcriptome consists of 260,143 transcripts for 138,096 genes, with a contig N50 of 1,731 bp. 71.5% of BUSCO Actinopterygii orthologs are represented as complete sequences in the transcriptome, with an additional 12.2% represented as partial sequences. Using the *E. perlongum* transcriptome and several other lines of evidence, ab initio genome annotation with the Maker pipeline identifies 28,034 putative protein-coding genes. 80% (22,534) of these predicted genes are assigned putative functional annotation based on homology with the UniProt Swiss-Prot database. 75% (16,850) of matches with the UniProt Swiss-Prot database are longer than 30 amino acids and have >50% sequence similarity. Additionally, InterProScan identifies 41,636 functional domains for 19,948 of the predicted genes.

Phylogenetics and population genetics

We used the *E. perlongum* genome to map and assemble ddRAD sequence data from 140 individuals of *E. perlongum* and *E. maculiceps*. All individuals in the ddRAD data set are represented by at least 461,000 reads, with a substantially larger mean number of reads per sample (mean = 3.7×10^6 , $SD = 2.1 \times 10^6$). The assembly pipeline identifies 267,237 orthologous ddRAD loci shared by at least four samples. Across all individuals, the mean read depth per locus was $22.5 \times$ (minimum = $6 \times$, $SD = 8.4$). In the filtered data set containing only biallelic SNPs with minor allele frequency greater than 5% and fewer than 10% missing genotypes, 98% (10,628/10,752) of the SNPs are located on the 24 longest genomic scaffolds.

Despite considerable variation in the percent of missing data (Supplementary Table S4), analyses of all alignments produce well-resolved phylogenies with congruent relationships between populations of *E. perlongum* and *E. maculiceps*. We therefore report only results from the phylogenetic analysis of the data set with the largest number of ddRAD loci. All individuals from the Waccamaw River and Lake Waccamaw resolve in a strongly supported clade that is nested within a lineage comprised of populations from the lower Pee Dee River system (Figure 1B).

However, genetic differentiation exists at fine spatial scales between Lake Waccamaw and the Waccamaw River (Figure 1C and D). Within the Waccamaw basin, the first principal component (PC) axis of genetic variation separates individuals from Lake Waccamaw and the Waccamaw River. Likewise, cross-entropy scores estimated by sNMF exhibit an “elbow” at $K = 2$, indicating that two ancestry clusters best explain the genetic data (Supplementary Figure S2A). Analyses with $K >$

2 do not reveal additional population substructure within the Waccamaw basin (Supplementary Figure S2B).

In addition to the two genetic ancestry clusters, we find evidence of an active hybrid zone in the river outlet immediately downstream of Lake Waccamaw (site 6, Figure 1A). Most individuals at this site have intermediate PC1 scores (Figure 1C) and exhibit admixed ancestry (Figure 1D). The snapclust algorithm classifies all but two individuals from this site as putative hybrids or backcrosses (Figure 1D). Outside of locality 6, only three other individuals are classified as hybrids or backcrosses.

Population structure analyses excluding SNPs within the chromosome 9 inversion (see below) reveal virtually identical results as the analyses of all SNPs (Supplementary Figure S3). However, population structure analyses of only chromosome 9 inversion SNPs reveal that a much larger proportion of genetic variance (64.2%) is explained by PC1, with individuals segregating into three clusters along PC1. Cross-entropy scores from sNMF analyses using only inversion SNPs exhibit an “elbow” at $K = 3$, with most individuals strongly assigned to one of the three ancestry clusters (Supplementary Figure S3).

Demographic modeling

To date the origin of *E. perlongum* and examine the history of gene flow between *E. perlongum* and *E. maculiceps*, we compared three different demographic scenarios across a range of plausible divergence times (Supplementary Figure S4a). The first scenario models allopatric divergence with a single instance of secondary contact to form a hybrid zone, but no subsequent gene flow. The second scenario also models allopatric divergence with secondary contact but allows gene flow between the two species after the formation of a hybrid zone. The third scenario models parapatric divergence by allowing continuous gene flow between the two lineages immediately after their initial split.

The scenarios of secondary contact with gene flow and parapatric divergence have consistently better Akaike information criterion (AIC) scores than the model with no gene flow after secondary contact (Supplementary Figure S4b). Support for these two models is mixed across a range of initial divergence times ($\Delta AIC < 10$, Supplementary Figure S4c, Supplementary Table S5). The most strongly supported demographic models indicate that divergence between *E. perlongum* in Lake Waccamaw and *E. maculiceps* in the Waccamaw River began 11,000–15,000 generations ago (Supplementary Table S5). With an Akaike weight of 0.89, the best fit of all examined models is the scenario of allopatric divergence between *E. perlongum* and *E. maculiceps* 11,000 generations ago with secondary contact and gene flow in the last 909 generations (Figure 1E).

Demographic inference suggests that a spatially narrow hybrid zone facilitates gene flow between *E. perlongum* and *E. maculiceps* (Figure 1E). Under the best-fit model, direct migration rate estimates between *E. perlongum* and *E. maculiceps* are very low (lake to river N_m [expected number of migrants per generation] ≈ 0.06 , river to lake $N_m \approx 0.01$). However, there are high migration rates from *E. maculiceps* in the Waccamaw River into the hybrid population ($N_m \approx 50$), as well as a slightly lower migration rate in the reverse direction ($N_m \approx 23$). The opposite is true for *E. perlongum* in Lake Waccamaw, with a migration rate from the hybrid population into *E. perlongum* ($N_m \approx 22$) that is much

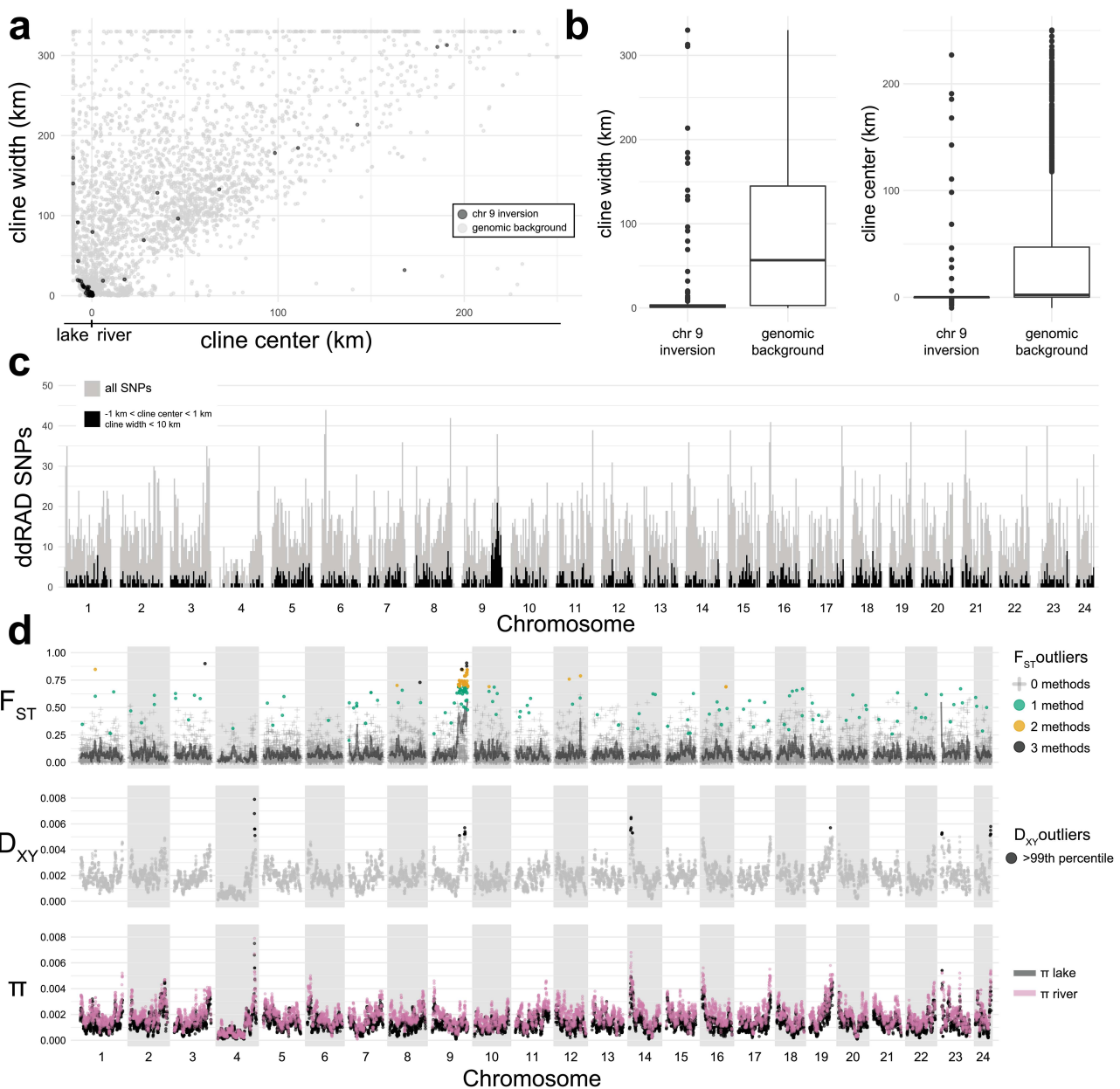


Figure 3. Genomic outliers of lake–river divergence are concentrated on chromosome 9. (A) Scatter plot of estimated cline centers versus cline widths for each ddRAD SNP. The outlier of Lake Waccamaw is located at 0 km on the x-axis. Black points represent SNPs within the inversion. (B) Boxplots of cline center and cline width for SNPs within the chromosome 9 inversion versus the genomic background. (C) SNP density along the 24 largest *Etheostoma perlongum* scaffolds, histogram bins in 1 Mbp windows. Gray bars show the distribution of all analyzed SNPs. Black bars show the distribution of SNPs with estimated cline centers within 1 km of the lake–river boundary and with estimated cline widths < 10 km. (D) Genome-wide estimates of F_{ST} (relative divergence), D_{xy} (absolute divergence), and π (nucleotide diversity) between Lake Waccamaw (*E. perlongum*) and the Waccamaw River (*E. maculataiceps*). F_{ST} estimates for individual SNPs are indicated by dots or crosses, sliding window F_{ST} estimates are indicated by the black line. Point color indicates how many methods identified a particular SNP as an outlier. D_{xy} plot shows sliding window estimates; black points indicate outliers. π plot shows sliding window estimates for Lake Waccamaw (black) and the Waccamaw River (pink). All sliding window estimates used a 1 Mbp window size with a 100 kbp step size.

higher than the migration rate from *E. perlongum* into the hybrid population ($N_e m \approx 0.08$).

Ecological and morphological divergence

We examined several lines of evidence for lake–river morphological and ecological divergence. Geometric morphometric analyses of body shape reveal that the first PC axis largely corresponds to a concave or convex shape of the abdominal region, which can be explained by the presence of gravid females in the data set (Supplementary Figure S5). However,

PC2 differentiates Lake Waccamaw *E. perlongum* from Waccamaw River *E. maculataiceps* (Figure 2B). Lake individuals have more slender bodies and elongate caudal regions compared with individuals from the river (Figure 2B). There is also an associated change in mouth position and head shape, with lake fish having more terminal mouths than their river counterparts (Figure 2B). Individuals from the hybrid zone downstream of Lake Waccamaw exhibit intermediate body shapes (Figure 2B). Body depth contributes strongly to this intermediacy; individuals from the hybrid zone are

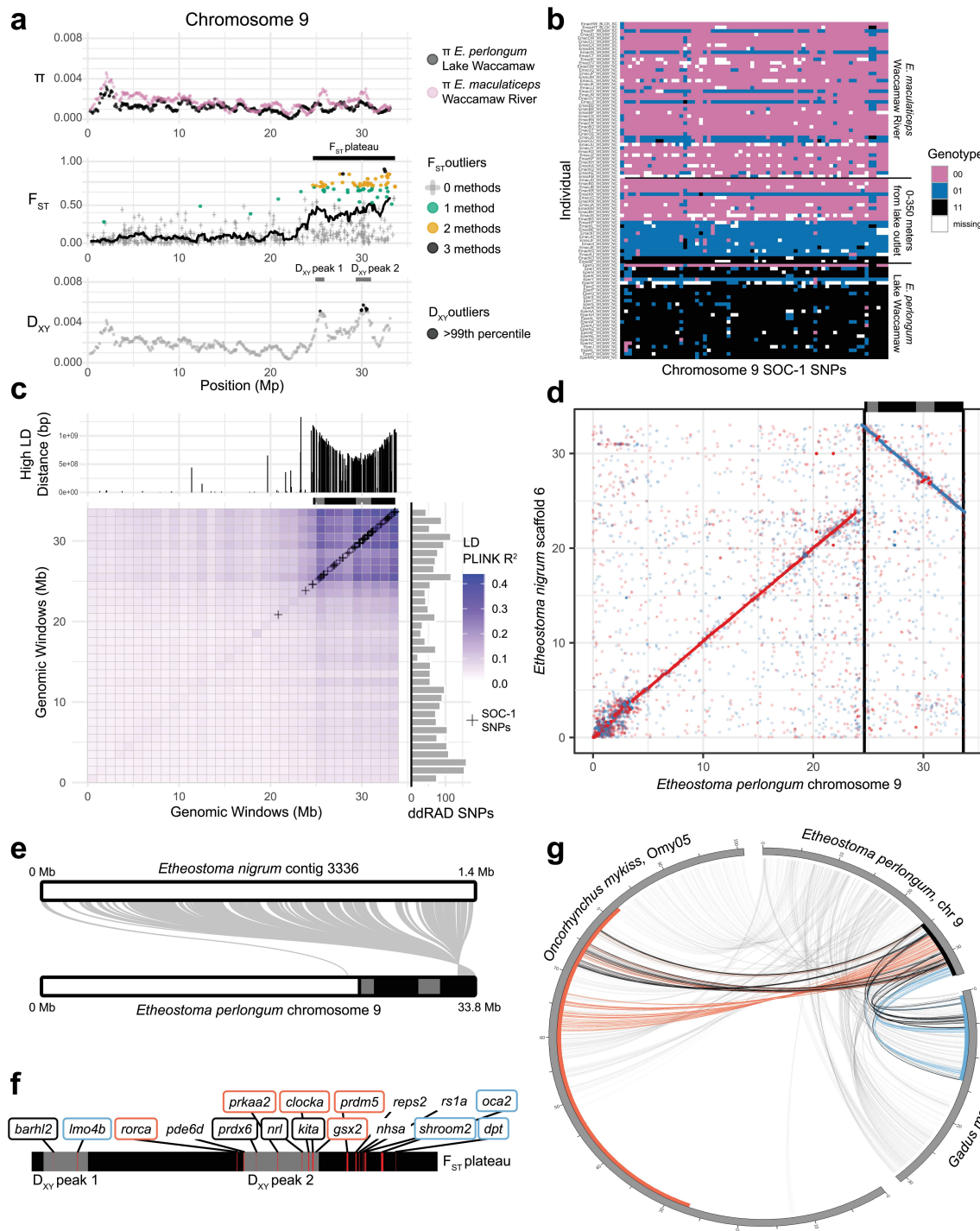


Figure 4. *Etheostoma perlongum* chromosome 9 contains a likely inversion supergene. (A) Chromosome 9 patterns of nucleotide diversity (π) and genomic differentiation (F_{ST} and D_{XY}). F_{ST} estimates for individual SNPs are indicated by dots or crosses; sliding window F_{ST} estimates are indicated by the black line. Point color indicates how many methods identified a particular SNP as an outlier. D_{XY} and π plots show sliding window estimates with a 1 Mbp window size with a 100 kbp step size. The F_{ST} plateau (black bar) and D_{XY} peaks (gray bars) indicate outlier regions used for gene ontology enrichment analyses. (B) Genotypes of SNPs in the chromosome 9 single outlier cluster of high LD (SOC-1). Individuals are ordered by increasing downstream distance from the north shore of Lake Waccamaw. (C) Linkage disequilibrium (LD) heatmap for 1 Mbp windows along chromosome 9. Locations of SNPs in SOC-1 are noted. Bar plot on the top margin shows per-SNP linkage metric calculated as the sum of distances between SNPs in high LD ($R^2 > .8$). Histogram on the right margin shows the distribution of ddRAD SNPs in 1 Mbp windows along chromosome 9. (D) Nucmer alignment dot plot of *E. perlongum* chromosome 9 versus *E. nigrum* scaffold 6. Red indicates forward strand alignments; blue indicates reverse strand alignments. (E) Alignment of *E. nigrum* contig 3336 spanning the *E. perlongum* inversion. (F) Position of vision, olfaction, and circadian genes within the chromosome 9 F_{ST} plateau. (G) Circos plot of gene synteny between *Etheostoma perlongum* chromosome 9, *Gadus morhua* linkage group 12, and *Oncorhynchus mykiss* chromosome 5. Red, blue, and black regions on each chromosome represent inversion/supergene location. Each link represents a syntenic gene pair. Gray links are genes that are not syntenic between two inversions. Orange links indicate genes syntenic between the *E. perlongum* and *O. mykiss* inversions, blue links indicate genes syntenic between the *E. perlongum* and *G. morhua* inversions, and black links indicate genes syntenic between *E. perlongum* and both *O. mykiss* and *G. morhua* inversions.

distinct from both *E. perlongum* in Lake Waccamaw and the Waccamaw River population of *E. maculaticeps* (Figure 2C). Body depth, head length, and PC axis two exhibit sharp geographic clines centered at the lake–river boundary (Supplementary Figure S6).

μCT imaging reveals that *E. perlongum* and *E. maculaticeps* differ in the number of vertebrae and several ecologically relevant measures such as the length of the hyoid and jaw opening in-lever (Figure 2D, Supplementary Figure S7). In addition, meristic traits that are traditional ichthyological taxonomic characters differ between *E. perlongum* and populations of *E. maculaticeps* in the Waccamaw River (Supplementary Figure S8). Several osteological and meristic traits that differ between the lake and river are strongly correlated with genomic ancestry estimates, and admixed individuals have intermediate trait values (Figure 2D, Supplementary Figures S7 and S8).

To further explore ecological lake–river divergence, we examined the gut contents of *E. perlongum* and Waccamaw River *E. maculaticeps*. We identified the gut contents from 25 specimens of *E. perlongum*. Two of these specimens were gravid females with empty stomachs and were excluded from further analysis. The three most frequent prey items to occur in *E. perlongum* stomachs are Amphipods, Chironomidae midge larvae, and Ostracods, each appearing in more than 50% of examined stomachs. Copepods, Cladocera, and Ceratopogonidae larvae also occur in greater than 10% of stomachs examined. A variety of benthic invertebrates occur in two or fewer individual stomachs, including Ephemeroptera larvae, Elmidae larvae, Diptera pupae, Annelidae, Rotifera, and a Decapoda.

We identified contents of 28 stomachs from *E. maculaticeps* specimens (one empty and excluded from analysis) and 6 specimens from the contact zone in the Waccamaw River, just below Lake Waccamaw (one with empty stomach and excluded from analysis). Chironomidae larvae are present in all *E. maculaticeps* stomachs, and Isopods (52%) and Ephemeroptera larvae (48%) are the other most frequently consumed prey. Trichoptera and Coleoptera larvae, Cladocera, Copepods, Amphipods, and Ostracods are consumed by 10%–20% of riverine individuals.

Etheostoma perlongum and *E. maculaticeps* consume many of the same prey categories, but their strength of association with these common prey differs (Supplementary Figure S9). The percent frequency occurrence of Chironomidae larvae is high for both *E. perlongum* and *E. maculaticeps*, but stomachs of *E. maculaticeps* tend to contain more chironomid larvae (mean 16.37, maximum 44) than *E. perlongum* (mean of 5.19, maximum 18). Microcrustacea (Amphipods, Ostracods, Copepods) are consumed by both species, but occur in *E. perlongum* stomachs more frequently and in higher numbers. Bipartite network analysis associates Ephemeroptera, Trichoptera, Coleoptera, Chironomidae, Bivalves, and Isopoda with *E. maculaticeps*, while the other prey categories (e.g., Ceratopogonidae, Amphipoda, Ostracoda, Copepoda) are more strongly associated with *E. perlongum* (Supplementary Figure S9).

Outlier SNPs

We examined thousands of SNPs to identify regions of the genome that are resistant to gene flow between *E. maculaticeps* and *E. perlongum*. Geographic cline analyses reveal that 14.6% of the 10,752 SNPs are best described by a model with a sharp allele frequency cline less than 10 km wide and

centered within 1 km of the lake–river boundary (Figure 3A). While SNPs are broadly distributed across the genome, SNPs with the narrowest clines centered on the lake–river boundary are concentrated on chromosome 9 (Figure 3B and C). Thirty-three percentage of the SNPs on chromosome 9 exhibit narrow geographic clines centered within 1 km of the lake–river boundary, while no other chromosome has more than 20% of SNPs with such narrow clines. This pattern is most pronounced within the last 9 Mb of chromosome 9, where 57% (110 of 195) of SNPs exhibit clines narrower than 10 km centered within 1 km of the lake–river boundary (Figure 3A and C).

Although SNPs with high lake–river differentiation are scattered across the genome, outlier detection methods reveal an overabundance of highly differentiated SNPs within the last 9 Mb of chromosome 9 (Figure 3D). Forty-five percentage (86 of 193) of PCAdapt outliers, 85% (53 of 62) of SNP F_{ST} outliers, and 60% (3 of 5) of BayeScan outliers are found on chromosome 9. The abundance of high F_{ST} SNPs on chromosome 9 is not the result of biased sex-ratio sampling; separate analyses of males and females qualitatively reveal the same patterns (Supplementary Figure S10). Outlier SNPs are not evenly distributed across chromosome 9; 94% (81 of 86) of chromosome 9 outlier SNPs occur within a 9 Mb plateau of elevated divergence with no obvious peak (Figure 4A). Sliding window estimates reveal a similar pattern: 86% (31 of 36) of all outlier F_{ST} windows create a plateau of divergence along the last 9 Mb of chromosome 9 (Figure 4A). Within the 9 Mb plateau of high F_{ST} , there are two smaller outlier peaks of elevated D_{XY} (Figure 4A). The 9 Mb region of chromosome 9 is the only genomic region that exhibits both elevated relative (F_{ST}) and absolute (D_{XY}) lake–river divergence (Figure 3D).

Evidence for an inversion on chromosome 9

To look for signals of structural genomic variation, we examined LD using several SNP data sets. The 9 Mb high F_{ST} plateau on chromosome 9 has elevated LD compared with the genomic background (Supplementary Figure S11, Figure 4C), and SNPs in this region comprise a single outlier cluster of high LD (SOC-1, Figure 4C). SOC-1 SNPs are in very high LD (median intracluster LD = 0.88) and exhibit genotypes strongly associated with the lake and river lineages (Figure 4B). Indeed, when lake and river chromosome 9 genotypes are examined separately, the 9 Mb block of high LD is no longer detected (Supplementary Figures S12 and S13). The per-SNP LD metric shows a clear drop in linkage just beyond the boundaries of the F_{ST} plateau (Figure 4C). This signature is also observed on chromosome 8 (Supplementary Figure S11), which also contains the only other outlier clusters of high LD (Supplementary Figure S14). However, unlike on chromosome 9, the high LD region on chromosome 8 persists when examining only river genotype individuals (Supplementary Figure S13) and is not strongly associated with divergent lake and river genotypes (Supplementary Figure S14).

We aligned the *E. perlongum* genome to a de novo, chromosome-level assembly of a close relative, *E. nigrum*. Compared with *E. nigrum* scaffold 6, there is a clear inversion at the end of *E. perlongum* chromosome 9 (Figure 4D). This large inversion corresponds precisely with the 9 Mb high F_{ST} plateau. The inversion breakpoint on *E. perlongum* chromosome 9 occurs at approximately 23.84 Mb and is spanned by a single *E. nigrum* contig (Figure 4E). Unlike the chromosome 9 inversion, the region of high LD on *E. perlongum* chromosome 8 is

not inverted relative to the *E. nigrum* genome (Supplementary Figure S1).

The *E. perlongum* chromosome 9 inversion contains 219 SNPs, many of which are in strong LD. To test whether linkage biased the outlier detection results (see previous section), we repeated the pcadapt and BayeScan analyses for each inversion SNP while excluding all other inversion SNPs. In all subsampled analyses, pcadapt identified 109/112 of the same noninversion outlier SNPs as the full analysis. Only one new noninversion SNP was also classified as an outlier. Furthermore, 80/81 of the inversion outliers identified by the full analysis were also identified as outliers in their respective subsampled analysis. In all subsampled analyses, BayeScan identified the same two noninversion outlier SNPs as the full analysis. In 96% of the subsampled BayeScan analyses, one additional noninversion outlier SNP was detected (chromosome 1:14026376). All 3 inversion outliers identified by the full analysis were also identified as outliers in their respective subsampled analysis.

Annotation of the *E. perlongum* genome assembly identifies key genes linked to vision and olfaction within the chromosome 9 inversion, many of which are concentrated within the two D_{xy} peaks (Figure 4F, Supplementary Table S6). Several of these genes, such as *pde6d*, *prdx6*, and *prdm5*, are related to severe retinal disease phenotypes in vertebrates (Burkitt Wright et al., 2011; Fatma et al., 2008; Thomas et al., 2014; Supplementary Table S6). GO enrichment analyses also identify many biological processes that may be overrepresented within the *E. perlongum* chromosome 9 inversion (Supplementary Table S7), though controlling for FDR results in no significant enrichment. Strikingly, the *E. perlongum* chromosome 9 inversion contains many genes, including key vision genes, that are also present in a well-documented inversion supergene on Atlantic Cod linkage group 12 and a massive double inversion supergene on Rainbow Trout chromosome 5 (Figure 4G, Supplementary Table S6) (Matschiner et al., 2022; Pearse et al., 2019). The darter, cod, and trout genomes exhibit conserved macrosynteny, although there are more rearrangements in the trout genome due to the salmonid-specific whole-genome duplication (Supplementary Figure S15).

Discussion

Rapid lacustrine speciation with gene flow

Shallow genetic differentiation is unusual among closely related darter species, which are typically millions of years divergent and exhibit reciprocal monophyly (Near et al., 2011). However, unlike the geologically ancient rivers of southeastern North America home to most darter species, Lake Waccamaw was completely inundated during high sea levels in mid-Pliocene (2.9–3.3 Ma; Rovere et al., 2015). The present configuration of Lake Waccamaw is relatively young; studies of sediment cores and pollen records estimate the lake originated in the late Pleistocene, 15,000–32,000 years ago (Frey, 1955; Stager & Cahoon, 1987). Under the best-fit demographic model with a generation time of 1–3 years (Raney & Lachner, 1943; Shute et al., 1982), speciation of *E. perlongum* initiated in the late Pleistocene or early Holocene (11,000–33,000 years ago), remarkably consistent with the young geologic age of Lake Waccamaw.

Although there is a narrow, contemporary hybrid zone between *E. perlongum* and *E. maculiceps* (Figure 1),

demographic analyses support an initial period of divergence with no gene flow (Supplementary Figure S4). Drought records show that Lake Waccamaw was occasionally isolated from the Waccamaw River (Shute, 1984). Historically, the outlet of Lake Waccamaw may have been smaller and intermittent like the outlets of other Carolina Bay lakes (Frey, 1949, 1951). We hypothesize that historic fluctuations in connectivity between the lake and river created periodic microgeographic isolation, limiting gene flow and initiating speciation. Thus, *E. perlongum* may fit the classic model of ecological speciation where an early allopatric phase is followed by secondary contact and reinforcement (Rundle & Nosil, 2005; Schluter, 2001). However, the best-fit demographic model contains substantial migration between the lake and river lineages (Figure 1E), suggesting that strong reproductive isolation did not evolve during the initial period of microallopatry.

Speciation frequently does not lead to complete reproductive isolation (Mallet, 2005). In darters, hybridization and introgression are common (Near et al., 2011); approximately 25% of darter species are involved with natural hybrid crosses (Keck & Near, 2009). However, hybridization at an early stage of speciation is unusual, as 93% of darter sister species pairs are allopatric, precluding opportunities for gene flow (Near et al., 2011). Even recently diverged darter species such as *P. freemanorum* and *P. kusha* (diverged approximately 300,000 years ago) are allopatric and do not actively hybridize (Near et al., 2021). Geographic isolation is the main driver of darter diversification (Mayden, 1985; Near et al., 2011) and most darter hybridization follows long periods of divergence (Moran et al., 2018). For example, hybridization between *E. caeruleum* and *E. spectabile* is well characterized on a phenotypic and genomic level (Moran et al., 2020), yet these species share most recent common ancestry 22 million years ago (Near et al., 2011). In contrast, *E. perlongum* and *E. maculiceps* diverged within the past 11,000–33,000 years (Figure 1E) but are actively hybridizing. Thus, this young species pair is the only known example among darters where gene flow can occur early in the speciation process. However, demographic analyses and geologic evidence indicate that microgeographic isolation still played an important role during the onset of speciation.

Ecology and morphology reflect genetic divergence

Morphological and ecological divergence provides additional evidence that *E. perlongum* is a distinct species. *Etheostoma perlongum* differs from *E. maculiceps* along multiple ecomorphological axes, including an elongated and shallower body (Figure 2). In fishes, slender body shapes are often associated with sustained swimming speeds, whereas deeper bodies can allow for more maneuverability in complex habitats (Gosline, 1997). Lake Waccamaw is an open lacustrine environment where sustained swimming performance may facilitate predator avoidance (Lindquist et al., 1981). Morphologically and genetically intermediate hybrid individuals in the lake outlet mark the center of a steep, narrow geographic cline (Supplementary Figure S6), which suggests divergent selection is acting on these phenotypes in lentic versus lotic habitats.

In addition to an elongated body shape, differences in head and jaw shape between *E. perlongum* and *E. maculiceps* indicate divergence in foraging ecologies. Lakes typically have an abundance of limnetic prey, while rivers offer mostly

benthic prey; these differences in prey availability drive lake–river body shape divergence in other fish species (Berner et al., 2009). Analysis of gut contents reveals that the two species occupy distinct trophic niches (Figure 2E), with *E. perlongum* associated more strongly with microcrustaceans from the water column such as amphipods, ostracods, and copepods (Thorp & Covich, 2010). These prey taxa appear infrequently in the diets of *E. maculaticeps*, which has a stronger association with strictly benthic prey taxa (Thorp & Covich, 2010) that rarely (Ephemeroptera) or never (Trichoptera and Isopoda) occur in the diets of *E. perlongum* (Supplementary Figure S9).

The combination of a more terminal mouth, an elongated body, and stronger association with limnetic prey (Figure 2) suggests that *E. perlongum* has partially escaped the benthic lifestyle typical of most darter species and utilizes more prey items in water column. Moreover, *E. perlongum* has a smaller lever ratio for lower jaw opening than *E. maculaticeps* (Figure 2D). This indicates enhanced mobility of the lower jaw associated with suction feeding compared with high values which are force-enhanced and associated with biting (Westneat, 2004). Other darter species associated with sandy microhabitats like Lake Waccamaw also have smaller jaw opening-lever ratios, which may aid in suctioning prey from loose substrate or feeding on more evasive zooplankton (Carlson & Wainwright, 2010). In addition to ecological and morphological divergence, *E. perlongum* exhibits an annual life cycle and faster growth rates compared to a generation time of 2–3 years in typical riverine populations of *E. maculaticeps* (Raney & Lachner, 1943; Shute et al., 1982). We hypothesize that the annual life history of *E. perlongum* may be related to greater seasonality of prey or differential predation pressures in lacustrine versus riverine ecosystems (Day & Rowe, 2002).

Divergence concentrated in a large chromosomal inversion

Divergence across the genome is often heterogeneous during speciation (Nosil et al., 2009) and many models of ecological speciation invoke genomic “islands” or “continents” of differentiation (Noor & Feder, 2006). We observe a single 9 Mb “continent” of genomic divergence between *E. perlongum* in Lake Waccamaw and *E. maculaticeps* in the Waccamaw River (Figures 3 and 4) that appears resistant to otherwise high levels of lake–river gene flow (Figure 1). This region contains a plateau of SNPs with high lake–river F_{ST} (Figures 3D and 4A) and extremely steep and narrow geographic allele frequency clines centered at the lake–river boundary (Figure 3A–C). Theory predicts that genomic regions resistant to gene flow should exhibit high levels of relative and absolute divergence (Cruickshank & Hahn, 2014). We find that the only genomic windows with significantly elevated relative (F_{ST}) and absolute (D_{XY}) divergence occur inside the 9 Mb F_{ST} plateau on chromosome 9 (Figure 3D). Within chromosome 9 and across the genome, we observe more F_{ST} outliers than D_{XY} outliers (Figure 3D), matching theoretical predictions that relative divergence metrics are more responsive to reduced gene flow than absolute divergence metrics (Cruickshank & Hahn, 2014). Importantly, the F_{ST} plateau of chromosome 9 does not exhibit reduced genetic diversity (Figures 3D and 4A), which can inflate estimates of F_{ST} (Cruickshank & Hahn, 2014).

Why are collinear SNPs on chromosome 9 resistant to gene flow between *E. perlongum* and *E. maculaticeps*? One

probable explanation is that a chromosomal inversion suppresses recombination in this region, maintaining tight linkage of *E. perlongum* and *E. maculaticeps* genotypes under divergent lake–river selection despite homogenizing gene flow. Models of inversion speciation have broad theoretical and empirical support (Fuller et al., 2019; Noor et al., 2001). We demonstrate that the F_{ST} plateau on chromosome 9 contains an outlier cluster of high LD SNPs (Figure 4C), with genotypes segregating by lake or river habitat (Figure 4B, Supplementary Figure S3). Outlier clusters of high LD are linked to known chromosomal inversions in *Anopheles* mosquitoes (Kempainen et al., 2015), sticklebacks (Roesti et al., 2015), and *Littorina* snails (Faria et al., 2019). The 9 Mb region of high LD on *E. perlongum* chromosome 9 spans several contigs and does not coincide with a scaffold breakpoint (Supplementary Figure S16), indicating that this signature of high LD is not the result of genome assembly errors (Miller et al., 2019).

Most convincingly, alignment of the *E. perlongum* genome to a close relative, *E. nigrum*, reveals a chromosomal inversion that precisely matches the high F_{ST} plateau (Figure 4D and E). *Etheostoma nigrum*, *E. maculaticeps*, and *E. perlongum* share most recent common ancestry approximately 6 million years ago (MacGuigan et al., 2023). We hypothesize that *E. nigrum* represents the ancestral arrangement of chromosome 9, and the 9 Mb inversion arose recently in *E. perlongum*. However, as chromosomal inversions are often polymorphic within species (Matschiner et al., 2022; Pearse et al., 2019), future genome sequencing of *E. maculaticeps* is needed to confirm this hypothesis.

We propose that the large inversion on chromosome 9 suppresses recombination between Lake Waccamaw *E. perlongum* and Waccamaw River *E. maculaticeps* genotypes, providing a potential genomic mechanism to explain divergence and speciation despite gene flow. However, it is challenging to demonstrate that the chromosome 9 inversion played a direct role in speciation between *E. perlongum* and *E. maculaticeps* (Villoutreix et al., 2021). Different mechanisms such as ancestral polymorphism can mislead inference about the importance of inversions in speciation (Fuller et al., 2018). Alternatively, proximity to the centromere could also produce high LD on chromosome 9 even without an inversion. Centromeric regions can play a role in speciation (Stump et al., 2005), though sometimes incidental due to the slow segregation of ancestral alleles (Turner & Hahn, 2010). However, if the 9 Mb region of chromosome 9 exhibits elevated LD simply because it is a centromeric region, we would expect to observe more regions of the *E. perlongum* genome with similarly high LD. Additionally, the 9 Mb region of chromosome 9 does not exhibit other hallmarks of centromeric regions, such as reduced gene density or increased density of repetitive sequences.

Loci outside of the chromosome 9 inversion may also play an important role in lake–river adaptation and speciation. Population structure analyses excluding the inversion SNPs still infer a substantial genetic break at the lake–river boundary (Supplementary Figure S3), and we identify some outlier regions of high lake–river differentiation outside of the inversion (Figure 3). However, the chromosome 9 inversion is the only region with both elevated relative and absolute genetic divergence. Therefore, the inversion exhibits the strongest, most consistent signals of selection across all analyses. Given the substantial inferred migration rates between the lake and

river lineages through the hybrid zone (Figure 1E), a large inversion suppressing recombination and limiting gene flow is likely to play an outsized role in speciation. Future studies utilizing whole-genome resequencing and examining patterns of gene expression will help solidify the relative importance of the chromosome 9 inversion versus other genomic regions.

Targets of selection and inversion supergene convergent evolution

Genetic and phenotypic divergence along narrow geographic clines is consistent with a scenario of divergent selection between *E. perlongum* in Lake Waccamaw and *E. maculiceps* in the Waccamaw River (Barton & Hewitt, 1985). In addition to predator avoidance and prey availability discussed earlier, researchers have long hypothesized that phenotypic divergence in the Waccamaw system is due to differences in light availability between the clearwater Lake Waccamaw and tannic or blackwater Waccamaw River (Hubbs & Raney, 1946; Krabbenhoft et al., 2009). Water clarity and chemistry ecotones are a major driver of speciation in other fishes, especially in the Amazon basin (Beheregaray et al., 2015). A clearwater lacustrine environment will favor greater use of vision for foraging, while riverine blackwater favors olfactory development, particularly since flowing water can transmit olfactory signals over longer distances than in standing water (Cox, 2008; Webster & Weissburg, 2008). The generality of lake-stream differentiation with respect to vision and olfaction is unclear, but there is some previous evidence of this tradeoff across similar environmental gradients (Santacà et al., 2021). Additionally, recent work demonstrates that the number of optic nerve fibers, rods, and cones differs between two other Lake Waccamaw endemic fish species, *F. waccamensis* and *M. extensa*, and their riverine relatives (Alsudani, 2020).

The *E. perlongum* chromosome 9 inversion contains numerous potential targets of divergent lake–river selection. Inversions suppress recombination, which can create supergenes of functionally linked loci that facilitate local adaptation in the face of gene flow (Gutiérrez-Valencia et al., 2021). Chromosomal inversions associated with local adaptation and ecological selection have been recently identified in several fish species (Matschiner et al., 2022; Pearse et al., 2019; Pettersson et al., 2019; Tigano et al., 2021). We hypothesize that the *E. perlongum* chromosome 9 inversion is a supergene containing numerous loci involved in retinal formation and olfaction that are likely targets of selection in lentic clearwater versus lotic blackwater habitats (Figure 4F, Supplementary Table S6). Additionally, the inversion contains genes involved in regulation of circadian rhythms such as *clocka*, which has been implicated in fish reproductive seasonality (Krabbenhoft & Turner, 2014) and may be linked to life-history differences between the annual, semelparous *E. perlongum* and the perennial, iteroparous *E. maculiceps*. Together, these linked genes are protected from homogenizing gene flow by an inversion and may have driven phenotypic and ecological divergence, facilitating speciation between *E. perlongum* and *E. maculiceps*.

Surprisingly, the *E. perlongum* chromosome 9 inversion exhibits gene synteny with two other recently documented inversion supergenes in Atlantic Cod (*Gadus morhua*) and Rainbow Trout (*Oncorhynchus mykiss*; Matschiner et al., 2022; Pearse et al., 2019; Figure 4G). Several genes, including *barlh2*, *prdx6*, *nrl*, and *kita*, are present in the inversions of all three species. Indeed, these four genes are localized to

the D_{xy} peaks within the darter inversion (Figure 4F), suggesting that they may be under stronger positive selection than the rest of the inversion. The similarity of gene content in these inversion supergenes is remarkable considering cod, darters, and trout share a most recent common ancestor in the late Triassic, approximately 225 million years ago (Near et al., 2012). Shared inversion genes are not simply a coincidental byproduct of genome rearrangement, as the darter, cod, and trout genomes have broadly conserved macrosynteny (Supplementary Figure S15). Rather, the inversions appear to have occurred independently in homologous regions of these genomes. We hypothesize that these inversions may share homologous breakpoints containing features such as highly repetitive sequence (Porubsky et al., 2020) or fragile z-DNA (Wang et al., 2006; Xie et al., 2019) that are prone to breakage and could provide a molecular mechanism for convergent genomic rearrangement. Alternatively, the exact breakpoints may not be homologous, but inversions in these regions are favored because they generate linkage among a set of genes that play a conserved role in local adaptation.

The genomic regions containing these convergent fish inversions also share deep homology with the mammalian X-chromosome, which has many of the same retinal development and circadian rhythm genes. Divergence of ancestral mammalian X- and Y-chromosomes was also potentiated by an inversion, which disrupted recombination and allowed divergence through sex-antagonistic selection (Kirkpatrick, 2010). In fishes, the same molecular mechanism (chromosomal inversion) in a homologous region of the genome (a portion of the ancestral X-chromosome) has repeatedly facilitated divergence through population-antagonistic selection. In Atlantic Cod and Rainbow Trout, these supergenes are polymorphic across hundreds of kilometers and associated with local adaptation (cod) or divergent migratory behavior (trout; Matschiner et al., 2022; Pearse et al., 2019). In contrast, genotype frequencies within the *E. perlongum* chromosome 9 inversion shift drastically over the span of just a few kilometers (Figures 3A–C and 4B). Thus, inversions in the same genomic regions containing similar sets of genes underlie both local adaptation across broad spatial scales and ecological speciation at fine spatial scales.

Darters, trout, and cod have substantially different life histories and ecologies, but there are some broad similarities regarding the homologous inversions. The Omy05 inversion in Rainbow Trout contains a QTL linked to temperature-dependent developmental rates; the derived, inverted genotype is associated with faster developmental rates in non-migratory individuals, which mature in colder environments (Nichols et al., 2008; Pearse et al., 2019). In Atlantic Cod, the LG12 inversion is associated with coastal versus offshore environments with differences in temperature, salinity, and dissolved oxygen (Berg et al., 2015, 2016; Bradbury et al., 2010; Sodeland et al., 2016). Ecomorphs in these environments differ in many traits, including age at maturity and growth rate (Hemmer-Hansen et al., 2013; Nordeide et al., 2011). Intriguingly, *E. perlongum* exhibits a higher growth rate, faster development, and shorter life span than *E. maculiceps* and other close relatives (Paine, 1990; Shute et al., 1982). While the specifics in each case differ, convergent evolution of inversions in these homologous genomic regions may allow repeated formation of similar, tightly linked gene complexes that influence major life-history traits.

Conclusions

Together, our results uncover an atypical example of lacustrine, ecological speciation in darters. The *E. perlongum* genome assembly, in combination with ddRAD sequencing, reveals that *E. perlongum* is a young species that originated following the formation of Lake Waccamaw in the late Pleistocene. *Etheostoma perlongum* in Lake Waccamaw has low levels of genetic differentiation from its closest relative, *E. maculiceps*, in the Waccamaw River and exhibits subtle but significant differences in functional traits associated with body shape, feeding system, and diet. Morphological and genetic intermediacy indicate an active hybrid zone between *E. perlongum* and *E. maculiceps* in a short stretch of the Waccamaw River just downstream of Lake Waccamaw. While there is strong evidence of historical and contemporary gene flow, a 9 Mb genomic region on chromosome 9 appears resistant to introgression between *E. perlongum* and *E. maculiceps*. This region contains an abundance of SNPs with steep, narrow allele frequency clines and elevated lake–river divergence relative to the genomic background. Elevated linkage disequilibrium in this region suggests that a chromosomal inversion is suppressing recombination, allowing alleles important for local adaptation or reproductive isolation to persist in the face of gene flow. This inversion may constitute a supergene involved in retinal development, olfaction, and circadian regulation, all potential targets of divergent lake–river selection. Inversion supergenes in two distantly related fish species have gene content that is strikingly similar to the darter inversion, suggesting deep evolutionary convergence of structural genomic variation may repeatedly facilitate local adaptation or speciation. *Etheostoma perlongum* demonstrates how, even in clades dominated by allopatric diversification, structural genomic variation can overcome the effects of homogenizing gene flow and enable rapid ecological speciation.

Supplementary material

Supplementary material is available online at *Evolution* (<https://academic.oup.com/evolut/qpad067>).

Data availability

The *Etheostoma perlongum* genome assembly is available through NCBI (PRJNA682188) and all associated sequence data (Nanopore reads, 10x Chromium Illumina reads, Hi-C Illumina reads, and RNA-seq reads) are deposited at the NCBI SRA (PRJNA682188). The *E. perlongum* genome annotation GFF3 file is available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.5hqbkzkhb5>. The *Etheostoma nigrum* genome assembly is also available through NCBI (PRJNA895006), and all associated sequence data (Nanopore reads, Hi-C Illumina reads) are deposited at the NCBI SRA (SRR22084753). Demultiplexed ddRAD sequence data are available on NCBI (PRJNA835500). Meristic data, morphometric data, osteological measurements, diet data, images used for 2D body shape morphometrics, and TPS landmark data are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.5hqbkzkhb5>. Raw images generated from μ CT scans are accessible through MorphoSource (<https://www.morphosource.org/projects/000512599>). Code for genome assembly and analyses available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.5hqbkzkhb5>.

Conflict of interest: The authors declare no conflict of interest.

Author contributions

D.J.M., T.J.K., and T.J.N. designed research; D.J.M., T.J.K., R.C.H., D.K.W., N.J.C.B., and T.J.N. performed research; D.J.M., T.J.K., R.C.H., D.K.W., N.J.C.B., and T.J.N. contributed new reagents/analytic tools; D.J.M., T.J.K., R.C.H., and D.K.W. analyzed data; D.J.M., T.J.K., R.C.H., D.K.W., and T.J.N. wrote the paper.

Funding

This material is based upon work supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2109761, the Yale Peabody Museum Bingham Oceanographic Fund, and the Yale Institute for Biospheric Studies Small Grants Program. D.J.M. was supported in part by the Yale Training Program in Genetics (National Institutes of Health grant number T32 GM007499).

Acknowledgements

A. Dornburg, L. L. Bowman, A. Ghezelayagh, D. Kim, and M. Correa assisted in field collections. G. Watkins-Colwell secured collecting permits and provided museum collections support. B. Tracy (NC DEQ) and G. Hogue (NCSM) provided tissue samples for this work. The Yale Center for Genome Analysis and T. Lan assisted sequencing efforts for the *Etheostoma perlongum* genome. Lastly, the authors thank V. A. Albert, O. Gokcumen, and members of the Near, Donoghue, Muñoz, and Krabbenhoft labs for their feedback on earlier versions of the manuscript.

References

- Alsudani, H. M. H. (2020). *Morphological changes in the visual sensory system of congeneric fishes inhabiting clear and turbid aquatic environments*. University of South Carolina.
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16(1), 113–148. <https://doi.org/10.1146/annurev.es.16.110185.000553>
- Beheregaray, L. B., Cooke, G. M., Chao, N. L., & Landguth, E. L. (2015). Ecological speciation in the tropics: Insights from comparative genetic studies in Amazonia. *Frontiers in Genetics*, 5, 477.
- Beissinger, T. M., Rosa, G. J. M., Kaepler, S. M., Gianola, D., & Leon, N. de (2015). Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics, Selection, Evolution*, 47, 30.
- Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S., & André, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution*, 7(6), 1644–1663. <https://doi.org/10.1093/gbe/evv093>
- Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., Jakobsen, K. S., & Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, 6, 23246.
- Berner, D., Grandchamp, A.-C., & Hendry, A. P. (2009). Variable progress toward ecological speciation in parapatry: Stickleback across eight lake-stream transitions. *Evolution*, 63, 1740–1753.
- Beugin, M. P., Gayet, T., Pontier, D., Devillard, S., & Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem.

- Methods in Ecology and Evolution*, 9(4), 1006–1016. <https://doi.org/10.1111/2041-210X.12968>
- Bonhomme, V., Picq, S., Gaucherel, C., & Claude, J. (2014). Momocs : Outline analysis using R. *Journal of Statistical Software*, 56(13), 24.
- Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., & Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, 277, 3725–3734.
- Burkitt Wright, E. M. M., Spencer, H. L., Daly, S. B., Manson, F. D. C., Zeef, L. A. H., Urquhart, J., Zoppi, N., Bonshek, R., Tosounidis, I., Mohan, M., Madden, C., Dodds, A., Chandler, K. E., Banks, S., Au, L., Clayton-Smith, J., Khan, N., Biesecker, L. G., Wilson, M., Rohrbach, M., Colombi, M., Giunta, C., & Black, G. C. M. (2011). Mutations in PRDM5 in brittle cornea syndrome identify a pathway regulating extracellular matrix development and maintenance. *American Journal of Human Genetics*, 88(6), 767–777. <https://doi.org/10.1016/j.ajhg.2011.05.007>
- Carlson, R. L., & Wainwright, P. C. (2010). The ecological morphology of darter fishes (Percidae: Etheostominae). *Biological Journal of the Linnean Society*, 100(1), 30–45. <https://doi.org/10.1111/j.1095-8312.2010.01417.x>
- Carroll, A. M., Wainwright, P. C., Huskey, S. H., Collar, D. C., & Turingan, R. G. (2004). Morphology predicts suction feeding performance in centrarchid fishes. *Journal of Experimental Biology*, 207(22), 3873–3881. <https://doi.org/10.1242/jeb.01227>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), s13742–015. <https://doi.org/10.1186/s13742-015-0047-8>
- Cox, J. P. L. (2008). Hydrodynamic aspects of fish olfaction. *Journal of the Royal Society Interface*, 5, 575.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates, Inc.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13), 3133–3157. <https://doi.org/10.1111/mec.12796>
- Daane, J. M., Dornburg, A., Smits, P., MacGuigan, D. J., Brent Hawkins, M., Near, T. J., William Detrich, H. III, & Harris, M. P. (2019). Historical contingency shapes adaptive radiation in Antarctic fishes. *Nature Ecology and Evolution*, 3(7), 1102–1109.
- 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. 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Day, T., & Rowe, L. (2002). Developmental thresholds and the evolution of reaction norms for age and size at life-history transitions. *American Naturalist*, 159(4), 338–350. <https://doi.org/10.1086/338989>
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). Hzar: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14(3), 652–663. <https://doi.org/10.1111/1755-0998.12209>
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9(10), e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A. M., & Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology*, 28(6), 1375–1393. <https://doi.org/10.1111/mec.14972>
- Fatma, N., Kubo, E., Sen, M., Agarwal, N., Thoreson, W. B., Camras, C. B., & Singh, D. P. (2008). Peroxiredoxin 6 delivery attenuates TNF- α and glutamate-induced retinal ganglion cell death by limiting ROS levels and maintaining Ca²⁺ homeostasis. *Brain Research*, 1233, 63–78. <https://doi.org/10.1016/j.brainres.2008.07.076>
- Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Eché, C., Valière, S., Donnadiou, C., Haffray, P., Bestin, A., Morvezen, R., Acloque, H., Euclide, P. T., Wen, M., Jouano, E., Scharlt, M., Postlethwait, J. H., Schraidt, C., ... Guiguen, Y. (2020). Characterization of a Y-specific duplication/insertion of the anti-Müllerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Molecular Ecology Resources*, 20(2), 531–543. <https://doi.org/10.1111/1755-0998.13133>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Frey, D. G. (1949). Morphometry and hydrography of some natural lakes of the North Carolina Coastal Plain: The Bay Lake as a morphometric type. *Journal of the Elisha Mitchell Scientific Society*, 65, 1–37.
- Frey, D. G. (1951). The fishes Of North Carolina's Bay Lakes and their intraspecific variation. *Journal of the Elisha Mitchell Scientific Society*, 1, 1–44.
- Frey, D. G. (1955). A time revision of the Pleistocene pollen chronology of Southeastern North Carolina. *Ecology*, 36(4), 762–763. <https://doi.org/10.2307/1931316>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. <https://doi.org/10.1111/2041-210x.12382>
- Fuller, Z. L., Koury, S. A., Phadnis, N., & Schaeffer, S. W. (2019). How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. *Molecular Ecology*, 28(6), 1283–1301. <https://doi.org/10.1111/mec.14923>
- Fuller, Z. L., Leonard, C. J., Young, R. E., Schaeffer, S. W., & Phadnis, N. (2018). Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLoS Genetics*, 14(7), e1007526. <https://doi.org/10.1371/journal.pgen.1007526>
- Gosline, W. A. (1997). Functional morphology of the caudal skeleton in teleostean fishes. *Ichthyological Research*, 44, 137–141. <https://doi.org/10.1007/BF02678693>
- Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L., & Slotte, T. (2021). The genomic architecture and evolutionary fates of supergenes. *Genome Biology and Evolution*, 13, 1–19.
- Hemmer-Hansen, J., Nielsen, E. E., Therkildsen, N. O., Taylor, M. I., Ogden, R., Geffen, A. J., Bekkevold, D., Helyar, S., Pampoulie, C., Johansen, T., & Carvalho, G. R.; FishPopTrace Consortium. (2013). A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, 22(10), 2653–2667. <https://doi.org/10.1111/mec.12284>
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12(1), 1–14. <https://doi.org/10.1186/1471-2105-12-491>
- Hubbs, C. L., & Raney, E. C. (1946). *Endemic fish fauna of Lake Waccamaw, North Carolina* (Vol. 65). University of Michigan, Museum of Zoology.
- Johnson, P. D., Bogan, A. E., Brown, K. M., Burkhead, N. M., Cordeiro, J. R., Garner, J. T., Hartfield, P. D., Lepitzki, D. A., Mackie, G. L., Pip, E., Tarpley, T. A., Tiemann, J. S., Whelan, N. V., & Strong, E. E. (2013). Conservation status of freshwater gastropods of Canada and the United States. *Fisheries*, 38, 247–282.
- Jombart, T. (2008). Adegnet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>

- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Jonsson, B., & Jonsson, N. (2001). Polymorphism and speciation in Arctic charr. *Journal of Fish Biology*, 58(3), 605–638.
- Kakioka, R., Kokita, T., Kumada, H., Watanabe, K., & Okuda, N. (2015). Genomic architecture of habitat-related divergence and signature of directional selection in the body shapes of *Gnathopogon* fishes. *Molecular Ecology*, 24(16), 4159–4174. <https://doi.org/10.1111/mec.13309>
- Keck, B. P., & Near, T. J. (2009). 2009. Patterns of natural hybridization in darters (Percidae: Etheostominae). *Copeia*, 758, 773.
- Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., Somboon, P., Mahanta, J., & Walton, C. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Molecular Ecology Resources*, 15(5), 1031–1045. <https://doi.org/10.1111/1755-0998.12369>
- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501. <https://doi.org/10.1371/journal.pbio.1000501>
- Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztröcy, A. W., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GO-TOOLS: A Python library for gene ontology analyses. *Scientific Reports*, 8(1), 1–17. <https://doi.org/10.1038/s41598-018-28948-z>
- 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. *Biological Journal of the Linnean Society*, 98(3), 636–645. <https://doi.org/10.1111/j.1095-8312.2009.01305.x>
- Krabbenhoft, T. J., & Turner, T. F. (2014). Clock gene evolution: Seasonal timing, phylogenetic signal, or functional constraint. *Journal of Heredity*, 105(3), 407–415. <https://doi.org/10.1093/jhered/esu008>
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3), 639–647. <https://doi.org/10.1111/1755-0998.12995>
- Lindquist, D. G., Shute, J. R., & Shute, P. W. (1981). Spawning and nesting behavior of the Waccamaw Darter, *Etheostoma perlongum*. *Environmental Biology of Fishes*, 6, 191.
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592>
- MacGuigan, D. J., Orr, O. D., & Near, T. J. (2023). Phylogeography, hybridization, and species discovery in the *Etheostoma nigrum* complex (Percidae: Etheostoma: Boleosoma). *Molecular Phylogenetics and Evolution*, 178, 107645. <https://doi.org/10.1016/j.ympev.2022.107645>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20(5), 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Briec, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology and Evolution*, 6(4), 469–481.
- Mayden, R. L. (1985). Biogeography of Ouachita Highland fishes. *Southwestern Naturalist*, 30(2), 195–211. <https://doi.org/10.2307/3670734>
- Mayr, E. (1942). *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press.
- McCartney, M. A., & Barreto, F. S. (2010). A mitochondrial DNA analysis of the species status of the endemic Waccamaw Darter, *Etheostoma perlongum*. *Copeia*, 103, 113.
- McCartney, M. A., Bogan, A. E., Sommer, K. M., & Wilbur, A. E. (2016). Phylogenetic analysis of Lake Waccamaw endemic freshwater mussel species. *American Malacological Bulletin*, 34(2), 109–120. <https://doi.org/10.4003/006.034.0207>
- Miller, J. T., Reid, N. M., Nacci, D. E., & Whitehead, A. (2019). Developing a high-quality linkage map for the Atlantic Killifish *Fundulus heteroclitus*. *G3 Genes/Genomes/Genetics*, 9, 2851–2862. <https://doi.org/10.1534/g3.119.400262>
- Mokodongan, D. F., Montenegro, J., Mochida, K., Fujimoto, S., Ishikawa, A., Kakioka, R., Yong, L., Mulis, Hadiaty, R. K., Mandagi, I. F., Masengi, K. W. A., Wachi, N., Hashiguchi, Y., Kitano, J., & Yamahira, K. (2018). Phylogenomics reveals habitat-associated body shape divergence in *Oryzias woworae* species group (Teleostei: Adrianichthyidae). *Molecular Phylogenetics and Evolution*, 118, 194–203. <https://doi.org/10.1016/j.ympev.2017.10.005>
- Moran, R. L., Catchen, J. M., & Fuller, R. C. (2020). Genomic resources for darters (Percidae: Etheostominae) provide insight into postzygotic barriers implicated in speciation. *Molecular Biology and Evolution*, 37(3), 711–729. <https://doi.org/10.1093/molbev/msz260>
- Moran, R. L., Zhou, M., Catchen, J. M., & Fuller, R. C. (2018). Hybridization and postzygotic isolation promote reinforcement of male mating preferences in a diverse group of fishes with traditional sex roles. *Ecology and Evolution*, 8(18), 9282–9294. <https://doi.org/10.1002/ece3.4434>
- 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: Etheostominae). *Systematic Biology*, 60(5), 565–595. <https://doi.org/10.1093/sysbio/syr052>
- Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., Wainwright, P. C., Friedman, M., & Smith, W. L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34), 13698–13703. <https://doi.org/10.1073/pnas.1206625109>
- Near, T. J., MacGuigan, D. J., Boring, E. L., Simmons, J. W., Albanese, B., Keck, B. P., Harrington, R. C., & Dinkins, G. R. (2021). A new species of bridled darter endemic to the Etowah River System in Georgia (Percidae: Etheostominae: Percina). *Bulletin of the Peabody Museum of Natural History*, 62(1), 15–42. <https://doi.org/10.3374/014.062.0102>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nichols, K. M., Edo, A. F., Wheeler, P. A., & Thorgaard, G. H. (2008). The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. *Genetics*, 179(3), 1559–1575. <https://doi.org/10.1534/genetics.107.084251>
- Noor, M. A. F., & Feder, J. L. (2006). Speciation genetics: Evolving approaches. *Nature Reviews Genetics*, 7(11), 851–861. <https://doi.org/10.1038/nrg1968>
- Noor, M. A. F., Gratos, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21), 12084–12088. <https://doi.org/10.1073/pnas.221274498>
- Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O., & Moum, T. (2011). Population connectivity among migratory and stationary cod *Gadus morhua* in the North-East Atlantic—A review of 80 years of study. *Marine Ecology Progress Series*, 435, 269–283. <https://doi.org/10.3354/meps09232>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.
- Oliveira, D. R., Reid, B. N., & Fitzpatrick, S. W. (2021). Genome-wide diversity and habitat underlie fine-scale phenotypic differentiation in the rainbow darter (*Etheostoma caeruleum*). *Evolutionary Applications*, 14(2), 498–512. <https://doi.org/10.1111/eva.13135>
- Orr, H. A. (1996). Dobzhansky, Bateson, and the genetics of speciation. *Genetics*, 144(4), 1331–1335. <https://doi.org/10.1093/genetics/144.4.1331>
- Ozerov, M. Y., Ahmad, F., Gross, R., Pukk, L., Kahar, S., Kisand, V., & Vasemägi, A. (2018). Highly continuous genome assembly of

- Eurasian perch (*Perca fluviatilis*) using linked-read sequencing. *G3 Genes|Genomes|Genetics*, 8, 3737–3743. <https://doi.org/10.1534/g3.118.200768>
- Paine, M. D. (1990). Life history tactics of darters (Percidae: Etheostomatiini) and their relationship with body size, reproductive behaviour, latitude and rarity. *Journal of Fish Biology*, 37(3), 473–488. <https://doi.org/10.1111/j.1095-8649.1990.tb05877.x>
- Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., ... Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology and Evolution*, 3(12), 1731–1742.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Pettersson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., Fan, G., Hong, X., Xu, Q., Zhang, H., Liu, S., Liu, X., Haggerty, L., Hunt, T., Martin, F. J., Flicek, P., Bunikis, I., Folkvord, A., & Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research*, 29(11), 1919–1928. <https://doi.org/10.1101/gr.253435.119>
- Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, 7(2), e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Porubsky, D., Sanders, A. D., Höps, W., Hsieh, P. H., Sulovari, A., Li, R., Mercuri, L., Sorensen, M., Murali, S. C., Gordon, D., Cantsilieris, S., Pollen, A. A., Ventura, M., Antonacci, F., Marschall, T., Korb, J. O., & Eichler, E. E. (2020). Recurrent inversion toggling and great ape genome evolution. *Nature Genetics*, 52, 849–858.
- Raney, E. C., & Lachner, E. A. (1943). Age and growth of Johnny darters, *Boleosoma nigrum olmstedii* (Storer) and *Boleosoma longimanum* (Jordan). *American Midland Naturalist*, 29(1), 229–238. <https://doi.org/10.2307/2420995>
- Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, 6(1), 8767. <https://doi.org/10.1038/ncomms9767>
- Rohlf, F. J. (2006). *tpsDig, version 2.10*. Department of Ecology and Evolution, State University of New York at Stony Brook.
- 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-Science Reviews*, 145, 117–131.
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology*, 8, 336–352.
- Santacà, M., Dadda, M., & Bisazza, A. (2021). The role of visual and olfactory cues in social decisions of guppies and zebrafish. *Animal Behaviour*, 180, 209–217. <https://doi.org/10.1016/j.anbehav.2021.08.017>
- Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, 16(7), 372–380. [https://doi.org/10.1016/s0169-5347\(01\)02198-x](https://doi.org/10.1016/s0169-5347(01)02198-x)
- Schluter, D., & G. L. Conte. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, 106(supplement_1), 9955–9962.
- Schwarzer, J., Herder, F., Misof, B., Hadiaty, R. K., & Schlieven, U. K. (2008). Gene flow at the margin of Lake Matano's adaptive sailfin silverside radiation: Telmatherinidae of River Petea in Sulawesi. *Hydrobiologia*, 615(1), 201–213. <https://doi.org/10.1007/s10750-008-9561-3>
- Seehausen, O., & Wagner, C. E. (2014). Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, 45, 621–651. <https://doi.org/10.1146/annurev-ecolsys-120213-091818>
- Shute, J. R. (1984). A systematic evaluation of the Waccamaw Darter, *Etheostoma perlongum* (Hubbs and Raney), with comments on relationships within the subgenus *Boleosoma* (Percidae: Etheostominae). University of Tennessee, Knoxville.
- Shute, J. R., Shute, P. W., & Lindquist, D. G. (1981). Fishes of the Waccamaw River drainage. *Brimleyana*, 6, 1–24.
- Shute, P. W., Shute, J. R., & Lindquist, D. G. D. A. A. (1982). Age, growth, and early life history of the Waccamaw Darter, *Etheostoma perlongum*. *Copeia*, 1982, 561–567.
- Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., Kent, M. P., Arnyasi, M., Olsen, E. M., & Knutsen, H. (2016). “Islands of Divergence” in the Atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, 8(4), 1012–1022. <https://doi.org/10.1093/gbe/evw057>
- Stager, J. C., & Cahoon, L. B. (1987). The age and trophic history of Lake Waccamaw, North Carolina. *Journal of the Elisha Mitchell Scientific Society*, 103, 1–13.
- Stump, A. D., Fitzpatrick, M. C., Lobo, N. F., Traoré, S., Sagnon, N., Costantini, C., Collins, F. H., & Besansky, N. J. (2005). Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*, 102(44), 15930–15935. <https://doi.org/10.1073/pnas.0508161102>
- Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6(7), e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Thomas, S., Wright, K. J., Le Corre, S., Micalizzi, A., Romani, M., Abhyankar, A., Saada, J., Perrault, I., Amiel, J., Litzler, J., Filhol, E., Elkhartoufi, N., Kwong, M., Casanova, J. L., Boddaert, N., Baehr, W., Lyonnet, S., Munnich, A., Burglen, L., ... Attié-Bitach, T. (2014). A homozygous PDE6D mutation in Joubert syndrome impairs targeting of farnesylated INPP5E protein to the primary cilium. *Human Mutation*, 35(1), 137–146. <https://doi.org/10.1002/humu.22470>
- Thorp, J. H., & Covich, A. P. (2010). *Ecology and classification of North American freshwater invertebrates* (3rd ed.). Academic Press.
- Tigano, A., Jacobs, A., Wilder, A. P., Nand, A., Zhan, Y., Dekker, J., & Therikildsen, N. O. (2021). Chromosome-level assembly of the Atlantic silverside genome reveals extreme levels of sequence diversity and structural genetic variation. *Genome Biology and Evolution*, 13.
- Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, 19(5), 848–850.
- Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, 30(12), 2738–2755. <https://doi.org/10.1111/mec.15907>
- von Hippel, F. A. (2008). Conservation of threespine and ninespine stickleback radiations in the Cook Inlet Basin, Alaska. *Behaviour*, 145, 693–724.
- Wang, G., Christensen, L. A., & Vasquez, K. M. (2006). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), 2677–2682. <https://doi.org/10.1073/pnas.0511084103>
- Webster, D. R., & Weissburg, M. J. (2008). The hydrodynamics of chemical cues among aquatic organisms. *Annual Reviews*, 41, 73–90. <https://doi.org/10.1146/annurev.fluid.010908.165240>
- Westneat, M. W. (2004). Evolution of levers and linkages in the feeding mechanisms of fishes. *Integrative and Comparative Biology*, 44(5), 378–389. <https://doi.org/10.1093/icb/44.5.378>
- Wiley, A. E. O., & Mayden, R. L. (1985). Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. *Annals of the Missouri Botanical Garden*, 72, 596–635.
- Williams, J. E., Johnson, J. E., Hendrickson, D. A., Contreras-Balderas, S., Williams, J. D., Navarro-Mendoza, M., McAllister, D. E., & Deacon, J. E. (1989). Fishes of North America endangered, threatened, or of special concern: 1989. *Fisheries*, 14(6), 2–20. [https://doi.org/10.1577/1548-8446\(1989\)014<0002:fonaet>2.0.co;2](https://doi.org/10.1577/1548-8446(1989)014<0002:fonaet>2.0.co;2)
- Wright, S. (1950). Genetical structure of populations. *Nature*, 166, 247–249. <https://doi.org/10.1038/166247a0>
- Xie, K. T., Wang, G., Thompson, A. C., Wucherpfennig, J. I., Reimchen, T. E., MacColl, A. D. C., Schluter, D., Bell, M. A., Vasquez, K. M., & Kingsley, D. M. (2019). DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, 363(6422), 81–84. <https://doi.org/10.1126/science.aan1425>