

# Complex patterns of genetic population structure in the mouthbrooding marine catfish, *Bagre marinus*, in the Gulf of Mexico and U.S. Atlantic

David S. Portnoy<sup>1</sup>  | Shannon J. O'Leary<sup>2</sup>  | Andrew T. Fields<sup>1</sup>  | Christopher M. Hollenbeck<sup>1</sup> | R. Dean Grubbs<sup>3</sup> | Cheston T. Peterson<sup>3</sup> | Jayne M. Gardiner<sup>4</sup> | Douglas H. Adams<sup>5</sup> | Brett Falterman<sup>6</sup> | J. Marcus Drymon<sup>7,8</sup>  | Jeremy M. Higgs<sup>9</sup> | Erin L. Pulster<sup>10,11</sup> | Tonya R. Wiley<sup>12</sup> | Steven A. Murawski<sup>11</sup>

<sup>1</sup>Marine Genomics Laboratory, Department of Life Sciences, Texas A&M University – Corpus Christi, Corpus Christi, Texas, USA

<sup>2</sup>Department of Biological Sciences, Saint Anselm College, Manchester, New Hampshire, USA

<sup>3</sup>Florida State University Coastal and Marine Laboratory, St. Teresa, Florida, USA

<sup>4</sup>New College of Florida, Sarasota, Florida, USA

<sup>5</sup>Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, Indian River Field Lab, Melbourne, Florida, USA

<sup>6</sup>Fisheries Research Support, Mandeville, Louisiana, USA

<sup>7</sup>Mississippi State University Coastal Research and Extension Center, Biloxi, Mississippi, USA

<sup>8</sup>Mississippi-Alabama Sea Grant Consortium, Ocean Springs, Mississippi, USA

<sup>9</sup>Center for Fisheries Research and Development, The University of Southern Mississippi, Ocean Springs, Mississippi, USA

<sup>10</sup>U.S. Geological Survey, Columbia Environmental Research Center, Columbia, Missouri, USA

<sup>11</sup>College of Marine Science, University of South Florida, St. Petersburg, Florida, USA

<sup>12</sup>Havenworth Coastal Conservation, Palmetto, Florida, USA

## Correspondence

David S. Portnoy, Department of Life Sciences, Texas A&M University – Corpus Christi, 6300 Ocean Drive, Corpus Christi, TX 78412, USA.  
 Email: [david.portnoy@tamucc.edu](mailto:david.portnoy@tamucc.edu)

## Funding information

Gulf of Mexico Research Initiative, Grant/Award Number: SA 18-16

## Abstract

Patterns of genetic variation reflect interactions among microevolutionary forces that vary in strength with changing demography. Here, patterns of variation within and among samples of the mouthbrooding gafftopsail catfish (*Bagre marinus*, Family Ariidae) captured in the U.S. Atlantic and throughout the Gulf of Mexico were analyzed using genomics to generate neutral and non-neutral SNP data sets. Because genomic resources are lacking for ariids, linkage disequilibrium network analysis was used to examine patterns of putatively adaptive variation. Finally, historical demographic parameters were estimated from site frequency spectra. The results show four differentiated groups, corresponding to the (1) U.S. Atlantic, and the (2) northeastern, (3) northwestern, and (4) southern Gulf of Mexico. The non-neutral data presented two contrasting signals of structure, one due to increases in diversity moving west to east and north to south, and another to increased heterozygosity in the Atlantic. Demographic analysis suggested that recently reduced long-term effective population size in the Atlantic is likely an important driver of patterns of genetic variation and

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Ecology and Evolution* published by John Wiley & Sons Ltd.

is consistent with a known reduction in population size potentially due to an epizootic. Overall, patterns of genetic variation resemble that of other fishes that use the same estuarine habitats as nurseries, regardless of the presence/absence of a larval phase, supporting the idea that adult/juvenile behavior and habitat are important predictors of contemporary patterns of genetic structure.

#### KEY WORDS

ddRAD sequencing, directional selection, heterozygote advantage, larval dispersal, mouthbrooding

#### TAXONOMY CLASSIFICATION

Population genetics

## 1 | INTRODUCTION

Contemporary patterns of genetic structure are the result of interactions among mutation, genetic drift, gene flow, and selection across many generations (Kidner et al., 2021). As such, the relative importance of these microevolutionary forces in shaping genetic diversity is temporally dynamic, changing with the demography of populations (Marko & Hart, 2011). Furthermore, for species distributed across heterogeneous environments, the relative importance of each force and the nature of their interactions may vary across space (Garant et al., 2007). Ultimately, it is allele frequency changes caused by these forces that underlie all evolutionary processes (Mayr & Provine, 1980), and those changes can be manifested at genome-wide scales (e.g., drift interacting with demographic change; Hollenbeck et al., 2019) or as locus-specific phenomena (e.g., selection operating at local scales in the genome; Gagnaire & Gaggiotti, 2016; Hoey & Pinsky, 2018, O'Leary et al., 2021), making genomic perspectives vital for assessments of genetic population structure (Bernatchez et al., 2017; Luikart et al., 2003).

For marine bony fishes, the proper interpretation of patterns of genetic structure is critical because this information is used by fisheries managers to delineate stock (i.e., population) boundaries and provide insight into both the long- and short-term sustainability of management units (Hilborn et al., 2003). Marine systems tend to be open (i.e., lacking obvious physical boundaries) and most marine bony fishes have large population sizes and are highly dispersive either as adults or especially during prolonged larval phases, and as a result, the traditional paradigm was that species should show genetic, and hence, demographic homogeneity over large spatial expanses (Waples, 1998). More recent empirical and theoretical studies have demonstrated that genetic structuring in marine fishes can occur at much smaller scales than originally predicted (Hauser & Carvalho, 2008), but the processes that lead to genetic discontinuities in the absence of hard barriers remain an important topic of research.

Gafftopsail catfish are distributed in coastal shelf habitats and saline estuaries in the western North Atlantic off New England and southward along the East Coast of the United States (hereafter U.S.) and throughout the Gulf of Mexico (Marceniuk et al., 2022). The species exhibits a seasonal migratory pattern related to reproduction (Yáñez-Arancibia & Lara-Dominguez, 1988). In U.S. waters along the South Atlantic coastline (Atlantic) and the Gulf of Mexico (Gulf), spawning is thought to occur along the nearshore shelf in the late spring into summer, with females producing small clutches of large eggs that are fertilized and taken into the male's mouth (Muncy & Wingo, 1983). Fertilized eggs and developing fry are subsequently carried by the male for a period that may exceed 60 days and released as small, fully formed juveniles in estuarine habitat (Gudger, 1918). While gaff-topsail catfish lack a dispersive larval phase, their use of estuaries as important habitat during the juvenile phase parallels what is seen in several co-distributed marine fishes that do have dispersive larval phases (Able, 2005).

While studies on the biology of gafftopsail catfish and other ariids are limited, there appear to be several adaptations associated with mouthbrooding. For females, these include the production of some of the largest eggs of any bony fish, which are attached to nonfunctional hyaline eggs, and seasonal modification of pelvic fins thought to facilitate the transfer of eggs to the male's mouth (Eastman et al., 1970; Gunter, 1947; Merriman, 1940). For males, adaptations include a period of fasting during incubation and seasonal changes to the volume of the oral cavity (Gudger, 1918; Lee, 1937). In addition, limited tagging data suggest that adult gafftopsail catfish show interannual fidelity to estuarine habitat: 11 of 14 gafftopsail catfish tagged with acoustic transmitters in Apalachicola Bay, Florida, emigrated from the system then returned the next year, with most individuals returning for a third year (Peterson & Grubbs, Florida State University, oral communication, July 2023). The combination of extreme parental investment in offspring and fidelity to spatially heterogeneous habitats during a critical life stage suggests the strong potential for localized adaptation.

Despite the importance of gaff-topsail catfish in regional fisheries and as a dominant midwater predator in nearshore ecosystems (Mendoza-Carranza & Hernández-Franyutti, 2005), they are understudied, and there have been only a few assessments of genetic population structure in this species and other ariid catfishes (Arroyo-Zúñiga et al., 2021; Avise, 1992; Avise et al., 1987; Santos & Quilang, 2012). Therefore, to understand patterns of population structure, this study assessed locus-specific and genome-wide patterns of variation in samples collected across the species' range using nuclear-encoded single nucleotide polymorphism (SNP)-containing loci (microhaplotypes) to decouple drift process from natural selection. Because gaff-topsail catfish are currently distributed in areas impacted by recent glacial cycles (Atlantic and Gulf, Portnoy et al., 2014), modeling based on site frequency spectra was used to understand the impact of historical demography on patterns of contemporary variation.

## 2 | METHODS

### 2.1 | Sampling and library prep

Fin clips were obtained from 382 mixed-age samples of gafftopsail catfish collected from nine geographic sampling locations (hereafter locations; Figure 1) from 2015 to 2018: one in the Atlantic in the Indian River Lagoon, Florida, and adjacent coastal waters (ATL) and eight in the Gulf. Locations in the Gulf were near Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN), near Mobile Bay, Alabama, (MB), in Mississippi Sound, Mississippi (MISS), in Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), in Corpus Christi Bay, Texas (CC) and the Bay of Campeche, Mexico (CAMP). All locations were selected because they represent inshore habitats used by mouthbrooding males for parturition and by juveniles as nursery habitat, except CAMP which was opportunistically sampled further offshore. Sampling took place as part of surveys routinely conducted by state or academic entities, the latter following approved animal care protocols. All fin clips were preserved in 20% DMSO-0.25M EDTA-saturated NaCl buffer (Seutin et al., 1991) and stored at room temperature until the time of extraction.

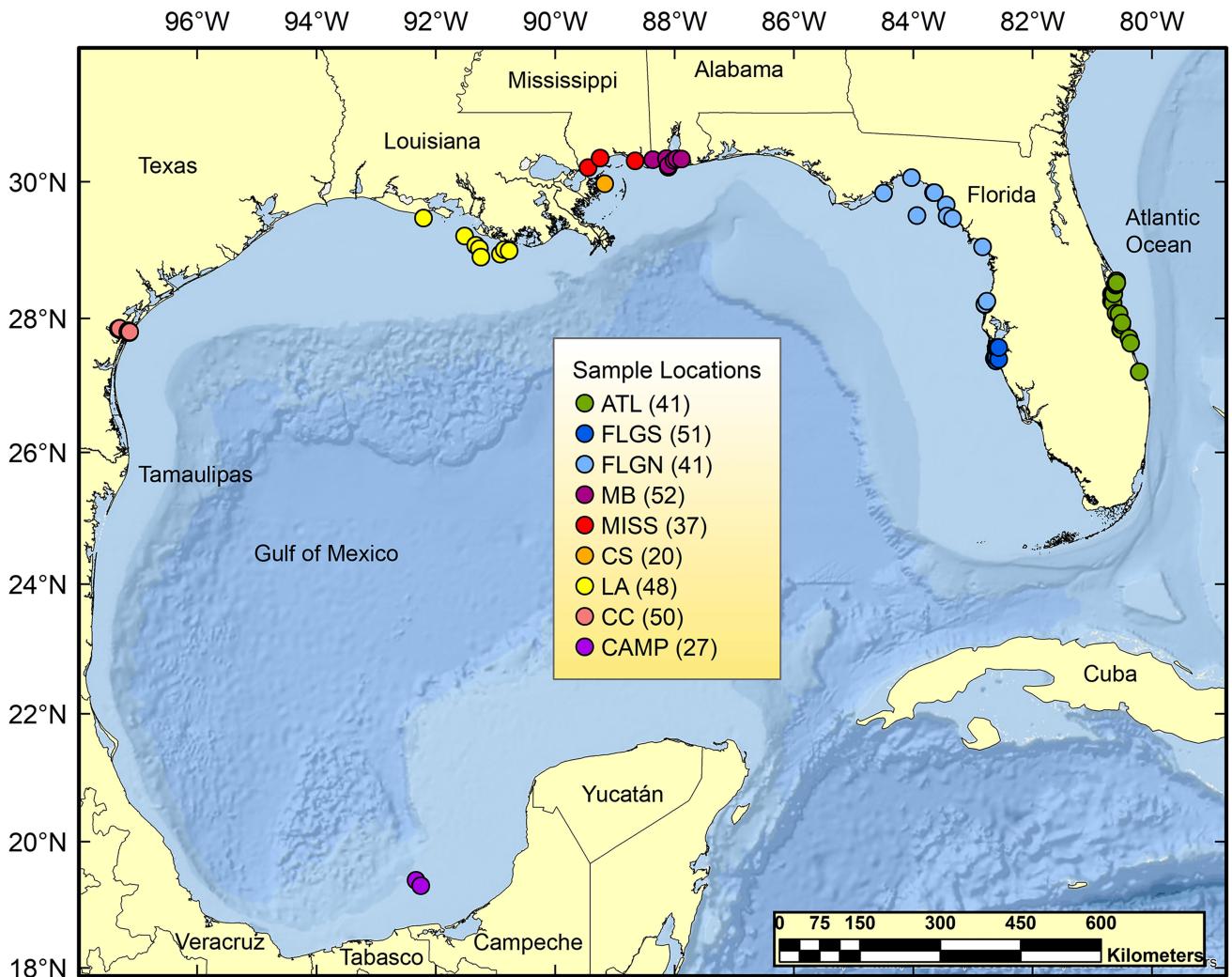
DNA was extracted using Mag-Bind Tissue DNA kits (Omega Bio-Tek, Norcross, GA) and 500–1000 ng of high-quality genomic DNA was used in a modified version of the ddRAD genomic library preparation method (Peterson et al., 2012). Briefly, genomic DNA was digested with two restriction endonucleases (EcoRI, Mspl), and a barcoded adapter was ligated to EcoRI sites while a common adapter was ligated to Mspl sites. Following adapter ligation, individuals were pooled by index and size-selected using a Pippin Prep size-selection system (Sage Science, Beverly, MA) to a standard size range (338–412 base pairs). Polymerase chain reaction (PCR) amplification of fragments was performed to incorporate adaptors necessary for annealing to an Illumina flow cell and index-specific identifiers. Index pools were then combined into libraries of approximately 150 individuals spread across the geographic range of sampling and duplicate

individuals (technical replicates), and three libraries were sequenced (150 bp paired-end) each on a lane of an Illumina HiSeq 4000 DNA sequencer at GeneWiz®, New Jersey, USA.

### 2.2 | Genotyping

RAD sequences retrieved from each run were demultiplexed using *process\_radtags* (Catchen et al., 2011), and quality trimming, reduced-representation reference assembly, read mapping, and SNP calling were performed using the *dDocent* pipeline (Puritz et al., 2014). The ten individuals with the highest number of reads were selected from each lane for de novo reduced-representation reference assembly using the overlapping read (OL) assembly option in *dDocent*. Similarity threshold for clustering ( $c=0.8$ ), minimum within individual coverage ( $K1=5$ ), and minimum number of individuals a read must occur in to be included ( $K2=2$ ) were chosen after comparing mapping statistics for ten individuals randomly chosen from each library and mapped to references generated for  $c=0.8$ ,  $K1=2-10$ , and  $K2=1-10$  using *BWA* (Li & Durbin, 2009) to maximize the number of reads mapped as a proper pair and minimize reads where forward and reverse reads mapped to different contigs. The constructed reduced-representation reference encompassed a total of 10,874,990 base pairs across 37,872 fragments (mean 287 bp; mode 307 bp).

Reads were mapped to the reduced-representation reference using *BWA* (Match = 1, mismatch penalty = 3 and gap penalty = 5; Li, 2013) and SNPs called using *freebayes* (Garrison & Marth, 2012). The resulting data set was filtered to remove low-quality and artefactual SNPs, paralogs, and low-quality individuals using *vcftools* (Danecek et al., 2011) and custom scripts following O'Leary et al. (2018), allowing for the retention of SNPs with more than two alleles. Genotypes with quality <20 and <5 reads were coded as missing, retaining loci with quality >20, genotype call rate >90%, and mean depth 15–300. Loci were also filtered based on allelic balance (remove SNPs <0.25 and >0.75), mapping quality ratios (remove SNPs <0.25 and >1.75), strand balance (remove SNPs with >100x more forward alternate reads than reverse alternate reads and >100x more forward reverse reads than reverse alternate reads), paired status, depth/quality ratio (<0.2), and excess heterozygosity (remove SNPs >0.5 and that deviate significantly from the expectations of Hardy–Weinberg Equilibrium). Individuals with >25% missing data were removed. Finally, *rad\_haplotype* (Willis et al., 2017) was used to merge SNPs on the same fragments into SNP-containing loci (hereafter microhaplotypes), by using a random sample of 20 reads per locus and recording all possible haplotypes, and then discarding haplotypes that are not possible given the SNPs present in the final dataset. Loci are flagged as paralogs if too many haplotypes are called given SNP genotypes. Genotyping error is flagged if an individual has too few haplotypes given SNP genotypes. The resulting haplotyped data set was further filtered to remove loci haplotyped in <90% of individuals, flagged as potential paralogs in >4



**FIGURE 1** Map of the sampling distribution of gafftopsail catfish, *Bagre marinus*, in the U.S. South Atlantic and Gulf of Mexico. The nine geographic locations are Indian River Lagoon, Florida (ATL) in the Atlantic, Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN), Mobile Bay, Alabama (MB), Mississippi Sound, Mississippi (MISS), Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), Corpus Christi Bay, Texas (CC), and the Bay of Campeche, Mexico (CAMP). Numbers next to abbreviated names indicate sample sizes.

individuals, or as affected by genotyping error in >10 individuals. Technical replicates were compared to assess genotyping error, and loci systematically affected by genotyping error or flagged as deviating significantly from the expectations of Hardy-Weinberg Equilibrium (HWE) in >5 locations were removed.

### 2.3 | Test for homogeneity and patterns of variance

Homogeneity in allele and genotype distributions among locations was tested using a single-level, locus-by-locus analysis of molecular variance (AMOVA) to account for uneven levels of missing data among loci (Weir & Cockerham, 1984). Analysis was implemented in ARLEQUIN v. 3.5 (Excoffier & Lischer, 2010) and significance determined at an  $\alpha$ -level of .05 by permuting individuals among locations 20,000 times and bootstrapping the data 20,000 times to generate 95% confidence

intervals. To visualize differences among locations, principal components analysis (PCA) was performed using the R (R Core Team, 2022) package adegenet v.1.7-18 (Jombart, 2008; Jombart & Ahmed, 2011).

### 2.4 | Identification of neutral and non-neutral loci

Two methods were explored to identify loci potentially under directional selection (outlier loci). First, the Bayesian modeling approach for outlier detection, implemented in BAYESCAN (Foll & Gaggiotti, 2008), was run with 25 pilot runs of 5000 iterations, followed by a burn-in of 50,000 iterations and 500,000 iterations sampled 10,000 times, and a  $q$ -value of 0.05. Second, the FDist method (Antao et al., 2008) for outlier detection implemented in ARLEQUIN was used with 20,000 coalescent simulations and a strict island model. Correction for multiple testing employed the Benjamini and Hochberg (1995) False Discovery Rate procedure (BH-FDR). Because

marine fishes typically have low background  $F_{ST}$  (Waples, 1998), assessments for loci under balancing selection were not conducted in either outlier analysis.

Because not all selection leads to patterns of differentiation consistent with assumptions of  $F_{ST}$  outlier analyses (Forester et al., 2018), redundancy analysis (RDA), implemented in *vegan* v. 2.5-6 (Oksanen et al., 2022), was used to explore associations between components of genetic and environmental variation and identify environmentally associated loci (env loci). Two constraining matrices were created, one describing spatial relationships between individuals and the other describing aspects of environmental heterogeneity between discrete sampling locations. Spatial relationships were described using Moran's Eigenvector Maps (MEMs; Dray et al., 2006), with geographic distance calculated using the R-package *codep* (Guenard et al., 2018), as the shortest distance between points on a sphere (great circle distance) using latitude and longitude data associated with the capture location of each individual. A final set of MEMs was then selected using stepwise forward selection, performed using the *ordiR2step* function in *vegan* with 999 permutations and  $\alpha$ -level of .05. An initial set of 68 environmental variables (Table S5) were obtained from the Bio-Oracle (Assis et al., 2018; Tyberghein et al., 2012) and MARSPEC (Sbrocco & Barber, 2013) data sets, with the final set selected as above.

Significant factors (MEMs and environmental variables) were included in a full model, with significance determined at an  $\alpha$  level of .01 using 1000 permutations. Then, variance partitioning was used to compare the relative contribution of geographic distance, environmental difference, and shared effects to observed patterns of genetic variation with the significance of each component determined at an  $\alpha$ -level of .01 using 1000 permutations. Finally, the environmental data were run separately to identify loci strongly associated with environmental variables by flagging alleles with a Mahalanobis distance  $\geq 20.51$  for  $p < .001$  ( $df = 5$ ).

The full dataset was then subdivided into "neutral" and "non-neutral" datasets. The non-neutral dataset was composed of outlier loci putatively under directional selection identified by at least one of the two detection methods and env loci. The neutral dataset contained all remaining loci. Each dataset was analyzed separately in all downstream analyses unless noted.

## 2.5 | Genetic variation within and among locations

To identify the number of genetically distinct groups in the data and assess for patterns of hierarchical structure, a Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010) was used as implemented in *adegenet*, using K-means clustering ( $K=1-40$ ) with the optimal number of clusters identified by comparing Akaike information criterion (AIC) values. To ensure sufficient variance was retained to discriminate among groups but not overfit the data, the optimum number of principal components to retain was determined using stratified cross-validation. Missing data was imputed using mean allele frequencies. Based on the results, individuals were clustered for  $K=2-5$  (20–50 PCs retained) and membership probabilities

of individuals to each inferred cluster were calculated. For selected  $K$  values, the percentages of variance explained by differences between groups and by differences between locations within groups were calculated in a locus-by-locus analysis of molecular variance (AMOVA) framework using ARLEQUIN. Pairwise  $F_{ST}$  was then estimated between each location using ARLEQUIN, with significance assessed at an  $\alpha$ -level of .05, by permutating individuals among locations 10,000 times. Correction for multiple testing employed BH-FDR, as above.

Expected heterozygosity ( $H_e$ ; Nei, 1973) and rarefied allelic richness ( $A_R$ ; El Mousadik & Petit, 1996) were estimated for each location using *hierfstat* v0.5-10 (Goudet, 2005). Friedman's tests in R were used to test homogeneity among locations for each measure of diversity. Post hoc Wilcoxon signed-rank tests in the R package *coin* v1.4-3 (Hothorn et al., 2023) were used to assess for pairwise differences between locations. Correction for multiple testing employed BH-FDR.

Chromosomal architecture has been implicated as an important driver in local adaptation because it can lead to co-adapted loci that sort together and suppression of recombination that can reshuffle linked co-adapted variation (Schwander et al., 2014). While some genomic resources are available for the order Siluriformes, none are appropriate for use in this study because the order arose at least 73 million years ago and contains more than 3000 named species and many undescribed species, spread across more than 30 families (Ferraris, 2007; Kappas et al., 2016; Sullivan et al., 2006), and available genomes are from taxa distant from the family Ariidae. Therefore, linkage disequilibrium network analysis (LDna; Kemppainen et al., 2015) was employed using the R package *LDna* v2.0 to identify single outlier clusters (SOCs), which are exclusive clusters of loci that show elevated linkage disequilibrium (LD) with one another relative to genome-wide LD. The rationale behind this analysis is that SOCs represent sets of loci either responding to the same microevolutionary force (in this case directional selection), segregating together due to chromosomal architecture, or a combination of both. Linkage disequilibrium was estimated for all pairs of loci across the entire data set and across all locations as  $r^2$ , following the Burrows method (Weir, 1996) as implemented in *NESTIMATOR* v2.01 (Do et al., 2014) with a minimum allele frequency of 0.01. Subsequently, the analysis was used to generate a series of networks where the nodes were loci and threshold LD values were used to constrain cluster formation. The threshold was lowered iteratively allowing further clustering. By measuring the change in median LD as clusters became more inclusive ( $\lambda$ ), sets of clusters with high  $\lambda$  were flagged and loci within each SOC were identified as neutral, outlier, or environmentally associated (env) loci. Finally, PCA was run separately for each set of loci from a SOC using *adegenet* to visualize the contribution of those loci to previously observed genetic structuring.

## 2.6 | Genetic demographic history

To better understand the demographic history of gafftopsail catfish, a differential approximation of the site frequency spectrum

(SFS) was performed using *MOMENTS* (Jouganous et al., 2017) and the neutral data set. One SNP was randomly selected from each locus and the site frequency spectrum (SFS) down projected using *easySFS* (Gutenkunst et al., 2009) to reduce the effect of missing data and to increase the speed of modeling. Six models were considered using symmetrical migration and population divergence between geographically adjacent populations. Multiple rounds of optimization are required to find the optimal estimates of demographic parameters for each model (Noskova et al., 2020), therefore, the four-step optimization procedure outlined in Portik et al. (2017) was implemented, with the optimal parameters from the previous step and perturbed at least 40 times. The optimization procedure was then rerun multiple times, using the parameter estimates from the previous run as a starting point, until the AIC values no longer changed. Finally, the model with the lowest AIC value was selected and confidence intervals around point estimates of demographic parameters were generated using the Godambe information matrix (Coffman et al., 2016) implemented in *MOMENTS*. Final estimates of demographic parameters used a generation time of 10 years and mutation rates of  $2.5 \times 10^{-8}$  to produce an upper bound,  $3.5 \times 10^{-9}$  to produce a lower bound, and the average ( $1.43 \times 10^{-8}$ ) to produce point estimates. Because the genome-wide mutation rate for ariid catfishes is not known, mutation rates were taken from the literature (Kavembe et al., 2016; Malinsky et al., 2018).

All figures were generated using *ggplot2* (Wickham, 2016) and *UpsetR* (Conway et al., 2017). An Rmarkdown and corresponding rendered html-document containing reproducible code for the complete analysis and functions as a standalone extended presentation of methods and results can be accessed at [https://github.com/marinenomicslab/gafftop\\_popgen\\_2024](https://github.com/marinenomicslab/gafftop_popgen_2024).

## 3 | RESULTS

### 3.1 | Genotyping

Initially, 1311, 068 SNPs were identified among 406 individuals. After filtering, the final data set consisted of 367 gafftopsail catfish individuals and 14,682 SNPs. Read depths per SNP were between 15 and 265, with a mean read depth of 78, and read depths per individual were between 10 and 269, with a mean read depth of 78. After haplotyping there were 5554 microhaplotype loci, with 2–25 alleles and 2–18 SNPs per locus. The final data set had less than 15% missing data per population, less than 10% missing data per locus, and less than 23% missing data per individual.

### 3.2 | Test for homogeneity and patterns of variance

The single-level AMOVA for the full data set revealed significant heterogeneity among locations ( $F_{ST}=0.0123$ ,  $p<.00005$ , 95% CI 0.0118–0.127). The top three principal components explained

2.53% of the total variance with ATL and CAMP separating from locations in the Gulf on PC1 and PC2 and the Gulf locations separating into the northeastern Gulf (FLGS and FLGN) and northwestern Gulf (MB, MISS, CS, LA and CC) along PC3 (Figure S1).

### 3.3 | Identification of neutral and non-neutral loci

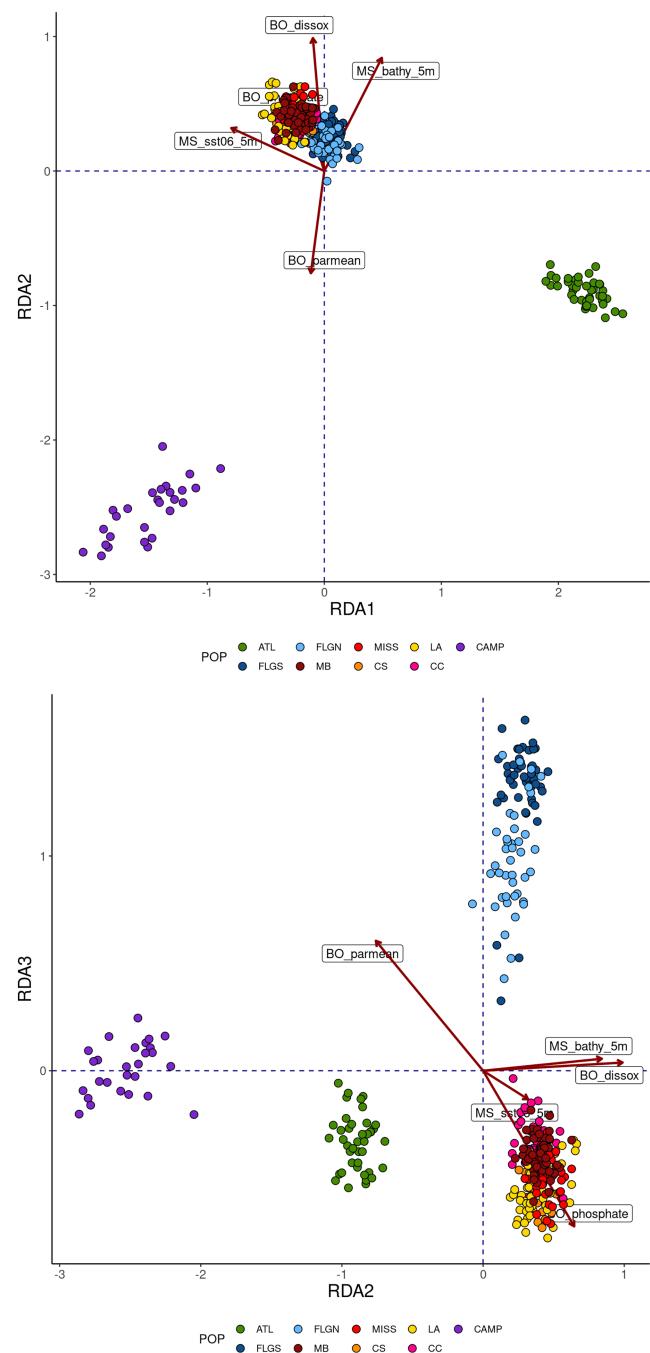
The two outlier detection methods identified 50 outliers (0.9% of all loci) putatively under directional selection. Analysis in ARLEQUIN identified 48 outlier loci, while BAYESCAN identified 15, 13 of which were also identified in ARLEQUIN. The first four MEMs were selected, as were five environmental variables: mean sea surface temperature in June (MS\_sst06\_5m), depth (MS\_bathy\_5m), dissolved oxygen (BO\_dissox), amount of photosynthetically available radiation at the sea surface (BO\_parmean), and concentration of ortho-phosphate (BO\_phosphate). Both geographic position and environment were significant alone and in combination in the full model, as were shared effects (Table 1), with environment explaining a greater portion of the variance than geographic position. Inspection of the RDA biplot revealed clustering similar to that of the initial PCA, with ATL and CAMP separating from the remaining locations in the Gulf along RDA1 and RDA2 and the northeastern and northwestern Gulf separating along RDA3 (Figure 2). A total of 118 loci had an allele with a Mahalanobis distance  $\geq 20.51$  ( $p<.001$ ), with 35 loci matching those previously identified as outliers. For downstream analyses, the full data set was divided into a neutral data set (5421 loci) and a non-neutral data set (133 loci).

### 3.4 | Genetic variation within and among locations

For the neutral data set, the minimum AIC value was obtained for  $K=4$  (Figure S2a) and cross-validation was run for  $K=2$ – $5$ . Assignment plots (Figure 3a) revealed hierarchical structure as  $K=2$  divided ATL from all other locations,  $K=3$  split CAMP from the remaining locations; and  $K=4$  further divided the northeastern Gulf (FLGS, FLGN) and northwestern Gulf (MB, MISS, CS, LA, CC). At  $K=5$  assignment success decreased, with no obvious further geographic clustering in the assignment plot. The component of variance attributable to differences

TABLE 1 Variance partitioning for geographic position (xy), environmental variables (env), and shared effects (shared) due to interaction of geographic position and environment.

Partition	Variance	p-Value
Residuals	0.9800	NA
xy + env + shared	0.0200	<.001
env + shared	0.0189	<.001
env	0.0121	<.001
xy + shared	0.0078	<.001
shared	0.0068	NA
xy	0.0011	<.001



**FIGURE 2** Biplot of redundancy analysis using the environmental model alone. The nine geographic locations are Indian River Lagoon, Florida (ATL) in the Atlantic, Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN) Mobile Bay, Alabama (MB), Mississippi Sound, Mississippi (MISS), Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), Corpus Christi Bay, Texas (CC), and the Bay of Campeche, Mexico (CAMP). Selected environmental variables are mean sea surface temperature in June (MS\_sst06\_5m), depth (MS\_bathy\_5m), dissolved oxygen (BO\_dissox), amount of photosynthetically available radiation at the sea surface (BO\_parmean), and concentration of ortho-phosphate (BO\_phosphate).

between groups was maximized for  $K=3$  (2.34%) but was only slightly larger than for  $K=2$  or 4 (2.14% and 1.56%, respectively), while the component of variance attributable to differences between locations

within groups was smaller for  $K=4$  (0.05%) than for  $K=2$  or 3 (0.62% and 0.26%, respectively). Estimates of pairwise  $F_{ST}$  (Table S1a) were significant after correction for 81% of comparisons (26/36), with small non-significant estimates between FLGS and FLGN and between 9 out of 10 comparisons in the northwestern Gulf, though MB and LA were significantly different.

For the non-neutral data set, AIC values continued to decrease with an increase in the number of clusters, but the magnitude of change became relatively small after  $K=4$  (Figure S2b), and cross-validation was run for  $K=2$ –5. Assignment plots (Figure 3b) revealed a hierarchical structure that differed from the neutral data. At  $K=2$ , ATL separated from the rest of the locations, while  $K=3$  split FLGS, FLGN, and CAMP from the northwestern Gulf locations, and  $K=4$  split CAMP from the northeastern Gulf. At  $K=5$ , assignment success decreased, with no obvious further geographic clustering in the assignment plot. The component of variance attributable to differences between groups was maximized for  $K=2$  (8.63%) and the component of variance attributable to differences between locations within groups was minimized at  $K=4$  (<0.01%). Estimates of pairwise  $F_{ST}$  (Table S1b) were significant after correction for 81% of comparisons (26/36), with non-significant estimates at or near zero between all ten comparisons in the northwestern Gulf; however, FLGN and FLGS were significantly different. Divergence estimates that were significant for both data sets were approximately three to ten times larger in the non-neutral data set relative to the neutral data set (Table S1c).

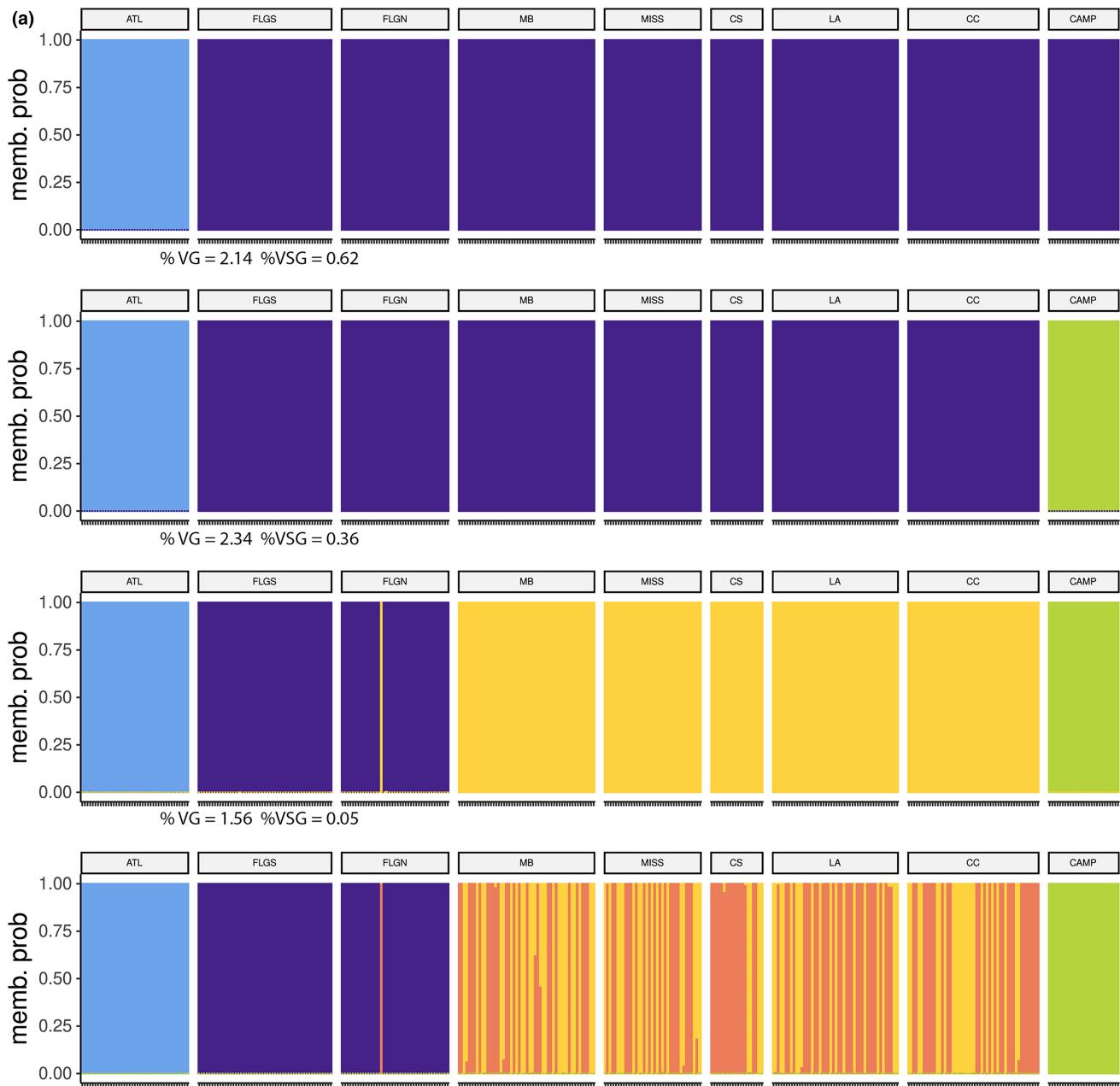
Expected heterozygosity differed significantly among the nine locations for both data sets ( $p < .0001$ ). For the neutral data set, estimated  $H_e$  was lowest in ATL (0.269) and greatest in CAMP (0.284; Table 2). After correction, nine pairwise comparisons were significant, with eight comparisons involving ATL (Table 2, Table S2). For the non-neutral data set, estimated  $H_e$  was lowest in MISS (0.198) and greatest in ATL (0.349, Table 2). After correction, ten comparisons were significant, with seven involving ATL and four involving MISS (Table 2, Table S3). Allelic richness also differed significantly among the nine locations for both data sets ( $p < .0001$ ). For the neutral data set, the estimated  $A_r$  was lowest in ATL (2.274) and highest in CAMP (2.586), and estimates for northwestern Gulf locations were uniformly higher than for northeastern Gulf locations (Table 2). After correction, 30 pairwise comparisons were significant (Table S3), with ATL less diverse and CAMP more diverse than all other locations. For the non-neutral data set, the estimated  $A_r$  was lowest in MISS (1.94) and greatest in ATL (2.161, Table 2). After correction, five pairwise comparisons were significant (Table S3), with ATL and CS more diverse than LA and MISS, and CAMP more diverse than LA.

For LDna analysis, the plot of  $\lambda$  values appeared to be asymptotic; therefore, a threshold value near the inflection point was selected (Figure S3). Four SOCs were identified (Figure 4a; Figure S3), two of which contained only neutral loci, with nine loci in Cluster 1 and eight loci in Cluster 2. The other two clusters contained a mix of neutral, outlier, and env loci. Cluster 3 contained two neutral loci, four outlier, and five env loci (four loci overlapped between outlier and env loci). Cluster 4 contained 15 neutral loci, 14 outlier, and 11 env loci (ten loci overlapped between outlier and env loci). The PCA plots using loci from Cluster 3 and Cluster 4 were in stark contrast,

with the plot resulting from Cluster 3 showing the northeastern Gulf and CAMP pulling away from the other regions (Figure 4b), while the plot resulting from Cluster 4 showed ATL separating from all other regions (Figure 4c). Allele frequencies at non-neutral loci from Cluster 3 displayed a latitudinal pattern in the Gulf, with more alleles present at lower latitudes (Figure 4d). By contrast, non-neutral loci from Cluster 4 had relatively similar numbers of alleles across geographic samples but displayed elevated  $H_e$  in ATL relative to all other locations (Figure 4e).

### 3.5 | Genetic demographic history

Model three (Figure 5), which had the lowest AIC (Table 3), identified CAMP as separating from all other populations first (T1), followed by the northwestern Gulf (T2), with a more recent split (T3) between the northeastern Gulf and the Atlantic. All other models are shown in Figure S4. Estimates of split times for T1 (159,226 ybp; 87,314–677,432) and T2 (158,586 ybp; 87,005–674,415) were similar, while the estimate of T3 (17 ybp; 10–73) was much more recent. Estimated



**FIGURE 3** Assignment plots for K-means clustering ( $K=2-5$ ) for the neutral (a) and outlier (b) data sets. The nine geographic locations are Indian River Lagoon, Florida, in the Atlantic (ATL), Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN) Mobile Bay, Alabama (MB), Mississippi Sound, Mississippi (MISS), Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), Corpus Christi Bay, Texas (CC), and the Bay of Campeche, Mexico (CAMP). The component of variance (calculated in an AMOVA framework) attributable to differences between groups (%VG) and the component of variance attributable to differences between samples within groups (%VSG) are reported for  $K=2-4$ .

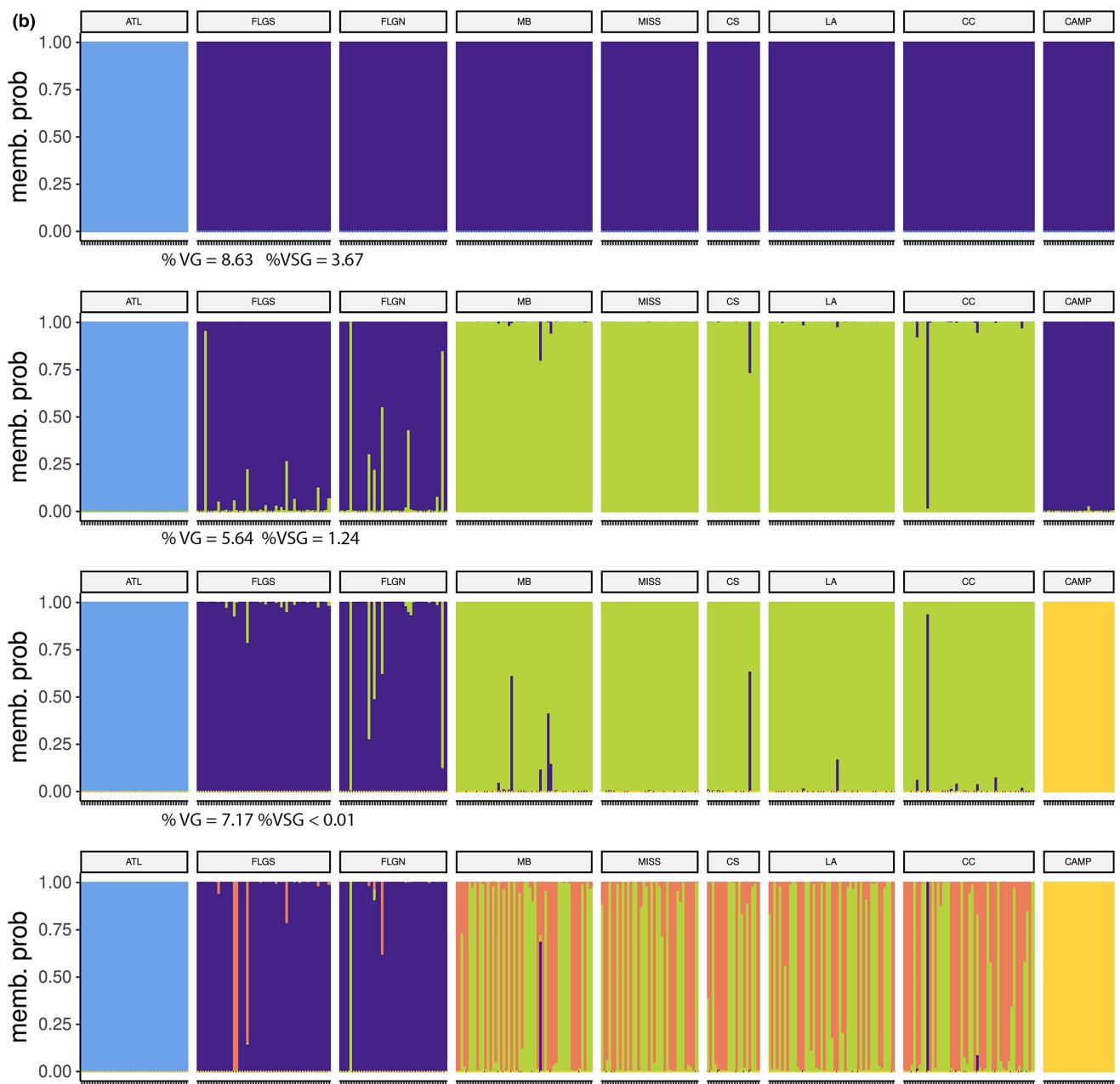


FIGURE 3 (Continued)

current long-term  $N_e$  was largest for the northwestern Gulf (129,683; 62,701–611,829), followed by the CAMP (7369; 15,558–74,494) and the northeastern Gulf (926; 500–3993), with a relatively small estimate for the Atlantic (10; 6–43). Point estimates of current long-term  $N_e$  for the Atlantic and northeastern Gulf were also relatively smaller than estimates of long-term  $N_e$  post-split (Atlantic,  $N_e=22$ ; northeastern Gulf;  $N_e=13,960$ ). Parameter estimates are shown in Table S4.

## 4 | DISCUSSION

Patterns of genetic variation among gafftopsail catfish sampled at eight locations across the Gulf of Mexico and one in the U.S.

Atlantic were assessed using over 5000 microhaplotypes. Analyses using datasets of neutral and non-neutral loci recovered four independent genetic units that corresponded to the samples collected in the Atlantic (ATL), the northeastern Gulf of Mexico (FLN+FLS), the northwestern Gulf (MB, MISS, CS, LA, and CC) and the Bay of Campeche (CAMP); however, hierarchical patterns of structuring differed among the data sets. Expected heterozygosity was lower for ATL than all other locations for neutral loci, but higher for non-neutral loci. Allelic richness was lowest in ATL and highest in CAMP for neutral loci and generally showed higher values moving from west to east and north to south in the Gulf for non-neutral loci. Linkage disequilibrium network analysis detected two clusters involving multiple outlier/environmental loci. The first cluster

**TABLE 2** Estimates of expected heterozygosity ( $H_e$ ) and rarefied allelic richness ( $A_r$ ) for the nine geographic samples (locations) using neutral (Neu) and non-neutral (NN) data sets. The nine geographic locations are Indian River Lagoon, Florida (ATL) in the Atlantic, Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN) Mobile Bay, Alabama, (MB), Mississippi Sound, Mississippi (MISS), Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), Corpus Christi Bay, Texas (CC), and the Bay of Campeche, Mexico (CAMP). The number of significant pairwise comparisons in which a location was higher (+) or lower (-) are indicated.

$H_e$	Neu	+	-	NN	+	-
ATL	0.269	0	8	0.349	7	0
FLGS	0.281	1	1	0.252	1	1
FLGN	0.282	1	0	0.255	1	1
MB	0.282	2	0	0.213	0	1
MISS	0.282	1	0	0.198	0	4
CS	0.283	1	0	0.218	0	1
LA	0.282	1	0	0.211	0	1
CC	0.282	1	0	0.218	0	1
CAMP	0.284	1	0	0.274	1	0
$A_r$	Neu	+	-	NN	+	-
ATL	2.274	0	8	2.161	2	0
FLGS	2.461	1	7	2.074	0	0
FLGN	2.468	2	6	2.123	0	0
MB	2.492	4	2	2.003	0	0
MISS	2.489	4	1	1.938	0	2
CS	2.469	3	4	2.053	2	0
LA	2.487	4	1	1.965	0	3
CC	2.496	4	1	1.992	0	0
CAMP	2.586	8	0	2.139	1	0

had five loci, with allelic gradients changing from west to east and north to south, consistent with patterns of within-group diversity. By contrast, the second cluster featured loci that showed elevated heterozygosity, but not allelic richness, in ATL relative to all other locations. Furthermore, demographic analyses suggest that patterns of variation in the northeastern Gulf and ATL may have been impacted by relatively recent declines in population size.

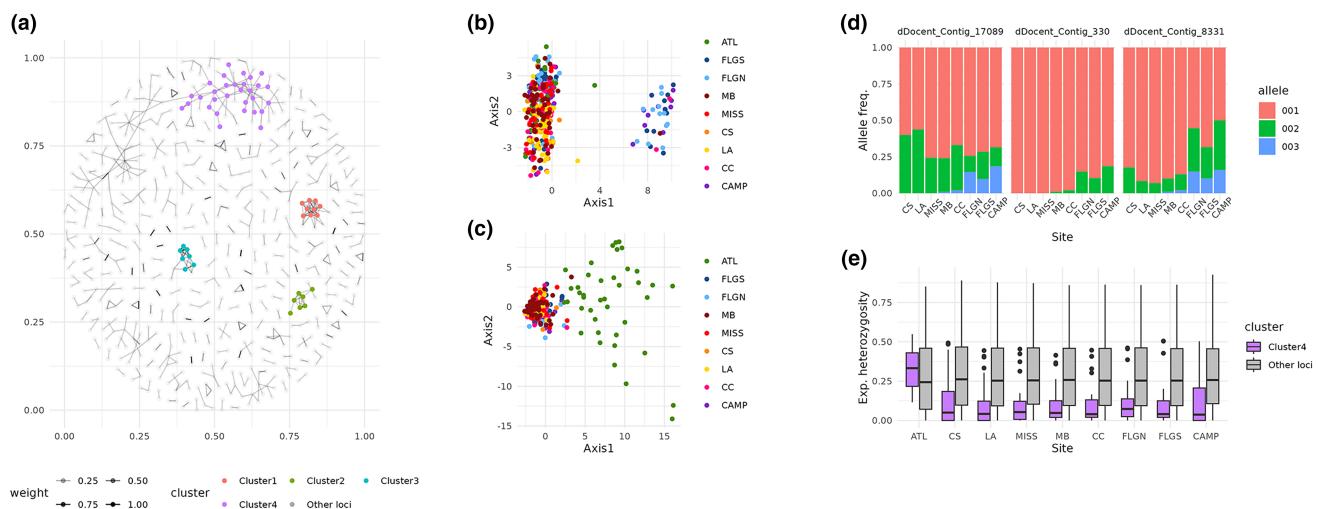
The neutral population structure exhibited in this study is congruent with what has been observed in a variety of nearshore/estuarine fishes along the U.S. Atlantic and Gulf Coast, in terms of the regional groupings recovered and the inferred location of geographic boundaries for those groups (sheepshead, *Archosargus probatocephalus*, Seyoum et al., 2017; spotted seatrout, *Cynoscion nebulosus*, Seyoum et al., 2018; red drum, *Sciaenops ocellatus*, Hollenbeck et al., 2019). The clear divergence of Atlantic gafftopsail catfish from all Gulf samples (Figures 2 and 3) is consistent with distributional data; gafftopsail catfish are found in high numbers along the Gulf Coast of Florida and historically on the Atlantic Coast of Florida from Indian River Lagoon north, but more rarely encountered in coastal southeastern Florida and the Keys (Armstrong et al., 1996; Serafy

et al., 2003). This break in distribution coincides with the Florida Vicariance Zone (Neigel, 2009), an area defined by a narrowing of the continental shelf and subsequent reduction/absence of estuarine habitat south of West Palm Beach, Florida, which is influenced by the swift-moving surface currents in the Florida Straits (Lynch-Stieglitz & Slowey, 1999). While it has been suggested that the east-flowing Florida Current may facilitate larval connectivity from west to east for discontinuously distributed species (e.g., Karnauskas et al., 2022), for species like gafftopsail catfish that lack a dispersive larval phase the area may represent a hard barrier (Portnoy et al., 2014). Consistent with this idea, pairwise differences between the ATL and northeastern Gulf locations were of similar magnitude to differences between the ATL and northwestern Gulf locations and largely decoupled from distance (Table S1).

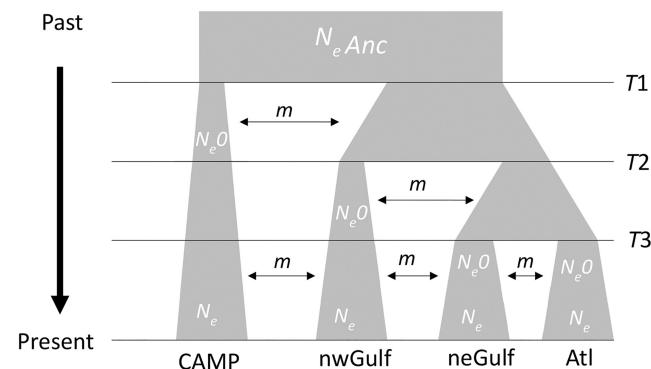
Divergence between CAMP and all other locations also is consistent with distributional data. The species is commonly encountered in the Bay of Campeche and along the Texas Coast, but less frequently encountered on the Yucatan platform and along portions of the Tamaulipas and Veracruz coastlines in Mexico, where the shelf narrows and estuarine habitat is reduced (DOF, 2018). In addition, while the Yucatan Channel and open waters of the Gulf can be traversed by marine fishes during the larval stage, resulting in connectivity between Mexican waters and the West Florida Shelf (Johnston & Bernard, 2017), this area likely proves a more effective barrier for gafftopsail catfish.

The divide between the northeastern and the northwestern Gulf falls within a well characterized biogeographic break centered on Mobile Bay, Alabama (McClure & McEachran, 1992), which has been attributed to a variety of current and historical processes (Portnoy & Gold, 2012) and is consistent with genetic breaks observed in a number of estuarine dependent species, including live bearing sharks and broadcast spawning bony fish with a larval phase (e.g., Hollenbeck et al., 2019; Portnoy et al., 2016; Seyoum et al., 2018). For gafftopsail catfish, estimated levels of divergence between northeastern and northwestern Gulf samples were approximately five times smaller than what was seen between northern Gulf samples and CAMP or ATL suggesting that the divergence is more recent and/or levels of recurrent gene flow persist across the northern Gulf, as the species appears continuously distributed in the region. Though there was no strong evidence of admixture in the data, the latter possibility is consistent with the idea that the central northern Gulf represents a marine suture zone, as has been suggested for red drum (Hollenbeck et al., 2019).

Site frequency spectra analysis provided insight into how historical demography has contributed to patterns of contemporary variation. Estimated split times for CAMP and the northwestern Gulf were very similar, falling largely within the Pleistocene (2.6 my – 11.8 kya), during which there were repeated glaciation events (see Ehlers & Gibbard, 2008 for a review), with point estimates in the glacial period that ended ~130 kya (Hughes & Gibbard, 2018). A combination of changes in salinity, reduced sea surface temperature, and decreased availability of nearshore habitat due to sea level decline (Bruner, 1982; Simms et al., 2007) may



**FIGURE 4** Results of linkage disequilibrium network analysis showing the four distinct clusters (a), with PCAs for loci in Cluster 3 (b) and Cluster 4 (c), as well as representative allele frequency plots for 3 of the 5 outlier/env loci in Cluster 3 (d) and a comparison of expected heterozygosity for 15 outlier/env loci in Cluster 4 with expected heterozygosity at all other loci for each location (e). The nine geographic locations are Indian River Lagoon, Florida, in the Atlantic (ATL), Tampa Bay, Florida (FLGS), North of Tampa Bay, Florida (FLGN) Mobile Bay, Alabama (MB), Mississippi Sound, Mississippi (MISS), Chandeleur Sound, LA (CS), off Louisiana west of the Mississippi River (LA), Corpus Christi Bay, Texas (CC), and the Bay of Campeche, Mexico (CAMP).



**FIGURE 5** Model 3 selected using site frequency spectrum analysis. Parameters of the model include ancestral long-term effective size ( $N_e \text{Anc}$ ), long-term effective size post-split ( $N_e O$ ), current long-term effective size ( $N_e$ ), split times ( $T$ ), and symmetrical migration rates ( $m$ ). The four populations are Campeche in the southern Gulf of Mexico (CAMP), the northwestern Gulf of Mexico (nwGulf), the northeastern Gulf of Mexico (neGulf), and the U.S. Atlantic (Atl).

periodically have made portions of the northern Gulf inhospitable for estuarine-dependent species like gafftopsail catfish. This may have caused periodic retreats into glacial refugia (Provan & Bennett, 2008) that lead to population subdivision via drift and selection, a scenario that has already been hypothesized to be an important driver of contemporary patterns of variation across taxa in the Gulf (Hollenbeck et al., 2019; Portnoy et al., 2014). By contrast, the estimated split time between ATL and the northeastern Gulf was relatively recent (<73 ya) and the estimate of long-term effective size of ATL very small (<43), suggesting that accelerated drift processes in ATL may have a large impact on current patterns

**TABLE 3** Results of site frequency spectrum model analysis showing the number of optimization rounds required (Opt), the log likelihood values (LnL), as well as the Akaike Information criteria values (AIC), and their change ( $\Delta$ AIC).

Model	Opt	LnL	AIC	$\Delta$ AIC
model_3	9	-2787.2	5614.3	0
model_2	11	-6376.8	12,795.7	-7181
model_1	10	-6416.9	12,875.8	-7261
model_4	5	-6631.7	13,299.4	-7685
model_5	6	-6822.1	13,684.2	-8070
Null	2	-8455.3	16,920.6	-11,306

of variation. Furthermore, the results indicate declines in long term effective size for both the northeastern Gulf (~15X based on point estimates) and ATL (~2X based on point estimates). The latter result is consistent with the documented declines in abundance of gafftopsail catfish along the U.S. Atlantic coast beginning in the late 1990's that are potentially attributable to "hardhead catfish virus" which caused documented massive mortality events for the hardhead catfish, *Ariopsis felis*, in the Atlantic around the same time period (Overstreet & Hawkins, 2017; Webster, 2015).

A comparison of the neutral and non-neutral datasets suggests that localized adaptation driven by regional environmental differences may play an important role in maintaining patterns of neutral structure (Nosil et al., 2008). While both data sets recovered the same four genetic units (ATL, northeastern Gulf, northwestern Gulf, and CAMP), patterns of hierarchical structuring, explored through K-means clustering, were not congruent. In addition, estimates of  $F_{ST}$  were larger for the non-neutral dataset, which is to be expected, but post hoc Mantel testing indicated that patterns between the

datasets were not significantly correlated ( $p > .05$ ), with the magnitude of increase greatest for comparisons between northeastern Gulf and northwestern Gulf samples (Table S1c). This reinforces the notion that the non-neutral loci are not simply drift outliers. This idea is further supported by estimates of diversity within groups. Estimates of  $H_e$  and  $A_r$  were high in FLGS, FLGN, and CAMP relative to northwestern Gulf samples for non-neutral loci, suggesting strong directional selection in the western Gulf or diversifying selection in the eastern Gulf. Given that the western Gulf population extends to Mobile Bay, well past the Mississippi River, and the lack of another obvious physical barrier, the results suggest isolation by adaptation dynamic (Orsini et al., 2013), which has been suggested for a variety of species where divergence is associated with differences in habitat that occur on scales well within species dispersal ranges (Bond et al., 2014; Hollenbeck et al., 2022; Jiang et al., 2019). Similar patterns of divergence and spatial structuring in neutral and non-neutral loci are seen in the co-distributed red drum (Hollenbeck et al., 2019), suggesting that selection may be operating on both species in a similar way. In this study, two sets of non-neutral loci were observed, those that distinguished the southern and northeastern Gulf from the northwestern Gulf, corresponding to shifts in allele frequencies, and those that separated the Atlantic from all other samples, characterized by elevated expected heterozygosity in the Atlantic (Figure 4c,e).

Directional selection associated with environmental heterogeneity can be an important driver of contemporary patterns of genetic structure (Nosil et al., 2009). In this study, five environmental factors were identified as significantly correlated with components of genetic variation and three (mean sea surface temperature in June, amount of photosynthetically available radiation at the sea surface, and concentration of ortho-phosphate) are associated with environmental parameters, like temperature and rainfall, that vary with latitude in the northeastern Gulf of Mexico (Hollenbeck et al., 2019). For species in the northern hemisphere, lower latitudes were less impacted by recent glacial cycles and thus represent relatively stable environments over evolutionary time scales; therefore, populations at low latitudes may display greater overall levels of genetic variation at genome-wide scales (Adams & Hadly, 2013; Hasselman et al., 2013). In the case of gafftopsail catfish, however, increased diversity with latitude is present at a subset of loci (though the southern Gulf features relatively high neutral diversity as well) consistent with latitude-associated selection, something that has been detected in other marine taxa (e.g., fishes: Bradbury et al., 2010; corals: Thomas et al., 2017; bivalves: Vendrami et al., 2019), including in the Gulf of Mexico (sharks: Portnoy et al., 2015).

Heterozygote advantage, which can result from numerous causes including increased enzymatic capacity and increased immune function, has not been frequently documented in wild systems (Hedrick, 2012), but in such cases is almost always associated with immune function (Gemmell & Slate, 2006). Increased immune function is an intriguing possibility for gafftopsail catfish in the Atlantic because the species, once very common, experienced a rapid decline

in abundance (>90%) beginning in the mid-1990s, a phenomenon potentially caused by a pathogen (Webster, 2015). In this situation, perhaps strong selection for immune diversity, associated with increased functionality for pathogen resistance (Evans & Neff, 2009; Osborne et al., 2015) may have created the pattern observed in the data. There are no appropriate genomic resources currently available for arid catfishes that can be used to support or refute either idea. Furthermore, because of the species continued rarity along the Atlantic Coast of the U.S., this study was only able to acquire a sufficient number of individuals from one U.S. Atlantic location in Florida. Thus, it is not known whether the patterns seen in the genomic data are coast-wide or locality-specific and further research would be needed to explore this possibility.

## 5 | CONCLUSIONS

A common theme in the marine population genetics literature is understanding the effect that larval dispersal has on patterns of spatial genetic structure (Faurby & Barber, 2012). Because gafftopsail catfish are a directly developing species in which males brood the young and deliver them into estuarine habitat, they are an interesting comparison to most bony fishes with dispersive larval phases. Patterns of structure found here in gafftopsail catfish were very similar to what is seen for several bony fishes in the Gulf of Mexico and U.S. South Atlantic that have larval phases but, like gafftopsail catfish, are estuarine obligates for part of their life cycle (see Hollenbeck et al., 2019). Furthermore, similar patterns of structure have been documented in several shark species that use estuaries as nursery habitat (Portnoy et al., 2014, 2016; Swift et al., 2023). By contrast, bony fish species that use less consolidated shelf habitat (e.g., red snapper, *Lutjanus campechanus*) show less structure over the same geographic area (see Portnoy et al., 2022). Taken together, this adds to the argument that aspects of adult/juvenile behavior (including dispersal) as well as patterns of habitat use are likely important but neglected predictors of the scale and pattern of population structure in marine species (Portnoy et al., 2022).

While elucidating the underlying mechanisms will require further work, this study demonstrated that patterns of diversity at non-neutral loci were attributable to different evolutionary processes. Structure associated with non-neutral loci should be complex in wild systems because interactions between genes and the environment differ in time, space, and across genomes (White & Butlin, 2021), yet many studies treat selection as unidimensional. Access to genomic resources, such as whole genomes or linkage maps, will allow researchers to characterize the distribution of differences across genomes within and across populations (the genomic landscape of divergence, Quilodrán et al., 2020), which is important for understanding more complicated evolutionary scenarios and investigating forces such as recombination that have been historically underexplored. The development of such resources is a clear next step toward understanding the adaptive variation presented in this study.

## AUTHOR CONTRIBUTIONS

**David S. Portnoy:** Conceptualization (lead); formal analysis (equal); funding acquisition (equal); project administration (lead); visualization (equal); writing – original draft (lead); writing – review and editing (equal). **Shannon J. O'Leary:** Conceptualization (equal); data curation (equal); formal analysis (equal); visualization (equal); writing – original draft (supporting); writing – review and editing (equal). **Andrew T. Fields:** Data curation (equal); formal analysis (equal); visualization (equal); writing – original draft (supporting); writing – review and editing (equal). **Christopher M. Hollenbeck:** Data curation (equal); formal analysis (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **R. Dean Grubbs:** Conceptualization (supporting); resources (equal); writing – review and editing (equal). **Cheston T. Peterson:** Resources (supporting); writing – review and editing (supporting). **Jayne M. Gardiner:** Resources (equal); writing – review and editing (supporting). **Douglas H. Adams:** Resources (equal); writing – review and editing (equal). **Brett Falterman:** Resources (equal); writing – review and editing (supporting). **J. Marcus Drymon:** Resources (equal); writing – review and editing (supporting). **Jeremy M. Higgs:** Resources (equal); writing – review and editing (supporting). **Erin L. Pulster:** Resources (supporting); writing – review and editing (supporting). **Tonya R. Wiley:** Resources (equal); writing – review and editing (supporting). **Steven A. Murawski:** Conceptualization (equal); funding acquisition (lead); resources (equal); writing – review and editing (equal).

## ACKNOWLEDGMENTS

The authors appreciate the effort put into sample acquisition by Joel Beaver (NCF), scientists of the Fisheries-Independent Monitoring program at the Indian River Field Laboratory of the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute, Cape Canaveral Scientific, Inc., Chloe Dean and the LDWF SEAMAP BLL team from the Grand Isle Fisheries Research Lab, the staff of USM's Center for Fisheries Research and Development (under USM IACUC # 11092203 and 15101509) and the crew and research party aboard the R/V Weatherbird II. We also wish to thank Julia Wallen Scott (TAMUCC), who assisted with lab work and Elizabeth Hunt (TAMUCC) for providing logistical support in the lab. We appreciate the USA and Mexican Departments of State, PEMEX, SEMARNAT, and CONAPESCA for permitting our sampling conducted in Mexican waters and Adolfo Gracia of UNAM University in Mexico City for logistical support. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This manuscript is based on work done by JMG while serving at the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This research was made possible by a grant from The Gulf of Mexico Research Initiative/C-IMAGE III, No. SA 18-16. This is publication 38 of the Marine Genomics Laboratory and 130 of Genetic Studies in Fishes.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

## DATA AVAILABILITY STATEMENT

Individual SNP genotypes, custom scripts, and filtering details are available at [https://github.com/marinegenomicslab/gafftop\\_pogen\\_2024](https://github.com/marinegenomicslab/gafftop_pogen_2024). Metadata is publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at <https://data.gulfresearchinitiative.org> (doi: [10.7266/n7-8jya-4m57]). Raw sequence reads under in the NCBI Sequence Read Archive (accession numbers: SRR25523260-SRR25523626). The final microhaplotype dataset, in VCF format, and the neutral and outlier haplotype datasets, in GENEPOP format, are available at Dryad doi: [10.5061/dryad.nvx0k6f0n](https://doi.org/10.5061/dryad.nvx0k6f0n). Benefits from this research revolve around sharing the results in presentations and publications and the data on public databases as described above. The work also involved postdocs, graduate, and undergraduate students who received professional training through their efforts.

## ORCID

- David S. Portnoy  <https://orcid.org/0000-0002-8178-1018>  
 Shannon J. O'Leary  <https://orcid.org/0000-0001-9775-9846>  
 Andrew T. Fields  <https://orcid.org/0000-0002-2817-2141>  
 J. Marcus Drymon  <https://orcid.org/0000-0002-2104-004X>

## REFERENCES

- Able, K. W. (2005). A re-examination of fish estuarine dependence: Evidence for connectivity between estuarine and ocean habitats. *Estuarine, Coastal and Shelf Science*, 64, 5–17. <https://doi.org/10.1016/j.ecss.2005.02.002>
- Adams, R. I., & Hadly, E. A. (2013). Genetic diversity within vertebrate species is greater at lower latitudes. *Evolutionary Ecology*, 27, 133–143. <https://doi.org/10.1007/s10682-012-9587>
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a  $F_{st}$ -outlier method. *BMC Bioinformatics*, 9, 323.
- Armstrong, M. P., Murphy, M. D., Muller, R. G., Harshany, D. P., & Crabtree, R. E. (1996). A stock assessment of hardhead catfish, *Arius felis*, and gafftopsail catfish, *Bagre marinus*, in Florida waters (report to the Florida marine fisheries commission October 22, 1996). Florida Marine Research Institute.
- Arroyo-Zúñiga, K. I., Granados-Amores, J., Palacios-Salgado, D. S., Peraza-Gómez, V., & Valenzuela-Quiñonez, F. (2021). Genetic diversity, population structure, and demographic history of the chihui sea catfish *Bagre panamensis* in northwestern Mexico inferred from mitochondrial DNA. *Mitochondrial DNA Part A DNA Mapping, Sequencing, and Analysis*, 32, 195–201. <https://doi.org/10.1080/24701394>
- Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27, 277–284. <https://doi.org/10.1111/geb.12693>
- Avise, J. C. (1992). Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. *Oikos*, 63, 62–76. <https://doi.org/10.2307/3545516>
- Avise, J. C., Reeb, C. A., & Saunders, N. C. (1987). Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning

- toadfishes (Batrachoididae). *Evolution*, 41, 991–1002. <https://doi.org/10.1111/j.1558-5646.1987.tb05872.x>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289–300. <https://doi.org/10.2307/2346101>
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H. O., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32, 665–680. <https://doi.org/10.1016/j.tree.2017.06.010>
- Bond, M. H., Crane, P. A., Larson, W. A., & Quinn, T. P. (2014). Is isolation by adaptation driving genetic divergence among proximate Dolly Varden char populations? *Ecology and Evolution*, 4, 2515–2532. <https://doi.org/10.1002/ece3.1113>
- Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snellgrove, 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. <https://doi.org/10.1098/rspb.2010.0985>
- Bruner, C. A. (1982). Paleoceanography of surface waters in the Gulf of Mexico during the late quaternary. *Quaternary Research*, 17, 105–119. [https://doi.org/10.1016/0033-5894\(82\)90048-5](https://doi.org/10.1016/0033-5894(82)90048-5)
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182. <https://doi.org/10.1534/g3.111.000240>
- Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33, 591–593. <https://doi.org/10.1093/molbev/msv255>
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCF tools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size ( $N_E$ ) from genetic data. *Molecular Ecology Resources*, 14, 209–214. <https://doi.org/10.1111/1755-0998.12157>
- DOF. (2018). *Carta Nacional Pesquera 2017*. Diario Oficial de la Federación.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modeling: A comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modeling*, 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Eastman, J. T., Moore, R. H., & Phillips, G. L. (1970). *Arius felis*: Pelvic fin modification in female. *Transactions of the American Microscopical Society*, 89, 427–430.
- Ehlers, J., & Gibbard, P. L. (2008). Extent and chronology of quaternary glaciation. *Episodes*, 31, 211–218. <https://doi.org/10.18814/epiuwg/2008/v31i2/004>
- El Mousadik, A., & Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92, 832–839. <https://doi.org/10.1007/BF0022189>
- Evans, M. L., & Neff, B. D. (2009). Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, 18, 4716–4729. <https://doi.org/10.1111/j.1365-294X.2009.04374.x>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Faourby, S., & Barber, P. H. (2012). Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology*, 21, 3419–3432. <https://doi.org/10.1111/j.1365-294X.2012.05609.x>
- Ferraris, C. J. (2007). Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes), and catalogue of siluriform primary types. *Zootaxa*, 1418, 1–628. <https://doi.org/10.11646/zootaxa.1418.1.1>
- 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, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27, 2215–2233. <https://doi.org/10.1111/mec.14584>
- Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, 62, 603–616. <https://doi.org/10.1093/cz/zow088>
- Garant, D., Forde, S. E., & Hendry, A. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, 21, 434–443. <https://doi.org/10.1111/j.1365-2435.2006.01228.x>
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907 [q-bio.GN]*.
- Gemmell, N. J., & Slate, J. (2006). Heterozygote advantage for fecundity. *PLoS One*, 1, e125. <https://doi.org/10.1371/journal.pone.0000125>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Gudger, E. W. (1918). Oral gestation in the gaff-topsail catfish. *Papers from the Department of Marine Biology of the Carnegie Institution of Washington*, 252, 25–542.
- Guenard, G., Legendre, P., & Pages, B. (2018). Package ‘codem’. <https://cran.r-project.org/web/packages/codem/codem.pdf>
- Gunter, G. (1947). Observations on breeding of the marine catfish, *Galeichthys felis*. *Copeia*, 1947, 217–223.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP data. *PLoS Genetics*, 5, e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Hasselman, D. J., Ricard, D., & Bentzen, P. (2013). Genetic diversity and differentiation in a wide-ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology*, 22, 1558–1583. <https://doi.org/10.1111/mec.12197>
- Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9, 333–362. <https://doi.org/10.1111/j.1467-2979.2008.00299.x>
- Hedrick, P. W. (2012). What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution*, 27, 698–704. <https://doi.org/10.1016/j.tree.2012.08.012>
- Hilborn, R., Quinn, T. P., Schindler, D. E., & Rogers, D. E. (2003). Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Science USA*, 100, 6564–6568. <https://doi.org/10.1073/pnas.103727410>
- Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-pandmixia in summer flounder. *Evolutionary Applications*, 11, 1732–1747. <https://doi.org/10.1111/eva.12676>

- Hollenbeck, C. M., Portnoy, D. S., de la Serrana, D. G., Magnesen, T., Matejusova, I., & Johnston, I. A. (2022). Temperature-associated selection linked to putative chromosomal inversion in king scallop (*Pecten maximus*). *Proceeding of the Royal Academy of Sciences B, Biological Science*, 289, 20221573. <https://doi.org/10.1098/rspb.2022.1573>
- Hollenbeck, C. M., Portnoy, D. S., & Gold, J. R. (2019). Evolution of population structure in an estuarine-dependent, marine fish. *Ecology and Evolution*, 9, 3141–3152. <https://doi.org/10.1002/ece3.4936>
- Hothorn, t., Winell, H., Hornik, K., Van de Wiel, M. A., & Zeileis, A. (2023). *Conditional inference procedures in a permutation test framework*. <https://cran.r-project.org/web/packages/coin/coin.pdf>
- Hughes, P., & Gibbard, P. (2018). Global glacier dynamics during 100 ka Pleistocene glacial cycles. *Quaternary Research*, 90, 222–243. <https://doi.org/10.1017/qua.2018.37>
- Jiang, S., Luo, M. X., Gao, R. H., Zhang, W., Yang, Y.-Z., Li, Y.-J., & Liao, P.-C. (2019). Isolation-by-environment as a driver of genetic differentiation among populations of the only broad-leaved evergreen shrub *Ammopiptanthus mongolicus* in Asian temperate deserts. *Scientific Reports*, 9, 12008. <https://doi.org/10.1038/s41598-019-48472-y>
- Johnston, M. W., & Bernard, A. M. (2017). A bank divided: Quantifying a spatial and temporal connectivity break between the Campeche Bank and the northeastern Gulf of Mexico. *Marine Biology*, 164, 12. <https://doi.org/10.1007/s00227-016-3038-0>
- Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, 206, 1549–1567. <https://doi.org/10.1534/genetics.117.200493>
- Kappas, I., Vittas, S., Pantzartzi, C. N., Drosopoulou, E., & Scouras, G. (2016). A time-calibrated mitogenome phylogeny of catfish (Teleostei: Siluriformes). *PlosOne*, 11, e0166988. <https://doi.org/10.1371/journal.pone.0166988>
- Karnauskas, M., Shertzer, K. W., Paris, C. B., Farmer, N. A., Switzer, T. S., Lowerre-Barbieri, S. K., Kellison, G. T., He, R., & Vaz, A. C. (2022). Source-sink recruitment of red snapper: Connectivity between the Gulf of Mexico and Atlantic Ocean. *Fisheries Oceanography*, 2022, 1–16. <https://doi.org/10.1111/fog.12607>
- Kavembe, G. D., Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Eco-morphological differentiation in Lake Magadi tilapia, an extremophile cichlid fish living in hot, alkaline and hypersaline lakes in East Africa. *Molecular Ecology*, 25, 1610–1625. <https://doi.org/10.1111/mec.13461>
- Kempainen, 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, 1031–1045. <https://doi.org/10.1111/1755-0998.12369>
- Kidner, J., Theodorou, P., Engler, J. O., Taubert, M., & Husemann, M. (2021). A brief history and popularity of methods and tools used to estimate micro-evolutionary forces. *Ecology and Evolution*, 11, 13723–13743. <https://doi.org/10.1002/ece3.8076>
- Lee, G. (1937). Oral gestation in the marine catfish, *Galeichthys felis*. *Copeia*, 1937, 49–56.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*: 1202.3997v2.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Luikart, G., England, P., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, 4, 981–994. <https://doi.org/10.1038/nrg1226>
- Lynch-Stieglitz, J., Curry, W. B., & Slowey, N. (1999). Weaker gulf stream in the Florida Straits during the last glacial maximum. *Nature*, 402, 644–648. <https://doi.org/10.1038/45204>
- Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2, 1940–1955. <https://doi.org/10.1038/s41559-018-0717-x>
- Marceniuk, A. P., Oliveria, C., Sales, J. B. L., & Betancur-R, R. (2022). The marine catfishes of the genus *Bagre* (Siluriformes; Ariidae) from the western Atlantic. *Marine Biodiversity*, 52, 12. <https://doi.org/10.1007/s12526-021-01238-0>
- Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, 26, 448–456. <https://doi.org/10.1016/j.tree.2011.05.007>
- Mayr, E., & Provine, W. B. (1980). *The evolutionary synthesis: Perspectives on the unification of biology*. Harvard University Press.
- McClure, M. R., & McEachran, J. D. (1992). Hybridization between *Prionotus alatus* and *P. paralatus* in the northern Gulf of Mexico (Pisces: Triglidae). *Copeia*, 1992, 1039–1046. <https://doi.org/10.2307/1446635>
- Mendoza-Carranza, M., & Hernández-Franyutti, A. (2005). Annual reproductive cycle of gafftopsail catfish, *Bagre marinus* (Ariidae) in a tropical coastal environment in the Gulf of Mexico. *Hydrobiológica*, 15, 275–282.
- Merriman, D. (1940). Morphological and embryological studies on two species of marine catfish, *Bagre marinus* and *Galeichthys felis*. *Zoologica*, 25, 221–248.
- Muncy, R. J., & Wingo, W. M. (1983). *Species profiles: Life histories and environmental requirements of coastal fishes and invertebrates (Gulf of Mexico) – sea catfish and gafftopsail catfish*. U.S. Fish and Wildlife Service, Division of Biological Services, FWS/OBS-82/11.5. U.S. Army Corps of Engineers, TR EL-82-4. 17 pp.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70, 3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Neigel, J. (2009). Population genetics and biogeography of the Gulf of Mexico. In D. L. Felder & D. K. Camp (Eds.), *Gulf of Mexico – Origins, waters, and biota: Biodiversity* (pp. 1353–1369). Texas A&M University Press.
- Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by adaptation' and multiple roles for divergent selection. *Evolution*, 62, 316–336.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402. <https://doi.org/10.1111/j.1558-5646.2007.00299.x>
- Noskova, E., Ulyantsev, V., Koepfli, K., O'Brien, S. J., & Dobrynin, P. (2020). GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience*, 9, gaa005. <https://doi.org/10.1093/gigascience/gaa005>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2022). *Vegan: Community ecology package. R Package Version 2.5-6*. <https://CRAN.R-project.org/package=vegan>
- O'Leary, S. J., Hollenbeck, C. M., Vega, R. R., & Portnoy, D. S. (2021). Disentangling spatial, environmental, and demographic signals in population genomics of an exploited, estuarine-dependent flatfish.

- Ecology and Evolution*, 11, 13415–13429. <https://doi.org/10.1002/ece3.8064>
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27, 3193–3206. <https://doi.org/10.1111/mec.14792>
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22, 5983–5999. <https://doi.org/10.1111/mec.12561>
- Osborne, A. J., Pearson, J., Negro, S. S., Chilvers, B. L., Kennedy, M. A., & Gemmell, N. J. (2015). Heterozygote advantage at MHC DRB may influence response to infectious disease epizootics. *Molecular Ecology*, 24, 1419–1432. <https://doi.org/10.1111/mec.13128>
- Overstreet, R. M., & Hawkins, W. E. (2017). Diseases and mortalities of fishes and other animals in the Gulf of Mexico. In H. Ward (Ed.), *Habitats and biota of the Gulf of Mexico: Before the Deepwater horizon oil spill, Volume 2* (pp. 1589–1745). Rice University.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7, e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26, 5245–5263. <https://doi.org/10.1111/mec.14266>
- Portnoy, D. S., Fields, A. T., Purtiz, J. B., Hollenbeck, C. M., & Patterson, W. F., III. (2022). Population genomic analysis of northern red snapper structuring in the U.S. Atlantic and Gulf of Mexico. *ICES Journal of Marine Science*, 79, 12–21. <https://doi.org/10.1093/icesjms/fsab239>
- Portnoy, D. S., & Gold, J. R. (2012). Evidence of multiple vicariance in a marine suture-zone in the Gulf of Mexico. *Journal of Biogeography*, 39, 1499–1507. <https://doi.org/10.1111/j.1365-2699.2012.02699.x>
- Portnoy, D. S., Hollenbeck, C. M., Belcher, C. N., Driggers, W. B., III, Frazier, B. S., Gelsleichter, J., Grubbs, R. D., & Gold, J. R. (2014). Contemporary population structure and post-glacial genetic demography in a migratory marine species, the blacknose shark, *Carcharhinus acronotus*. *Molecular Ecology*, 23, 5480–5495. <https://doi.org/10.1111/mec.12954>
- Portnoy, D. S., Hollenbeck, C. M., Bethea, D. M., Frazier, B. S., Gelsleichter, J., & Gold, J. R. (2016). Population structure, gene flow, and historical demography of a small coastal shark (*Carcharhinus isodon*) in U.S. waters of the western Atlantic Ocean. *ICES Journal of Marine Science*, 73, 2322–2332. <https://doi.org/10.1093/icesjms/fsw098>
- Portnoy, D. S., Puritz, J. B., Hollenbeck, C. M., Gelsleichter, J., Chapman, D., & Gold, J. R. (2015). Selection and sex-biased gene flow: Philopatric sorting of adaptive alleles in a coastal shark. *Molecular Ecology*, 24, 5877–5885. <https://doi.org/10.1111/mec.13441>
- Provan, J., & Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, 23, 564–571. <https://doi.org/10.1016/j.tree.2008.06.010>
- Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431. <https://doi.org/10.7717/peerj.431>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, 11, 6–21. <https://doi.org/10.1111/2041-210X.13324>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Santos, B. S., & Quilang, J. P. (2012). Genetic diversity analysis of *Arius manillensis* (Siluriformes: Ariidae) using the mitochondrial control region. *Mitochondrial DNA*, 23, 45–52. <https://doi.org/10.3109/19401736.2011.653796>
- Sbrocco, E. J., & Barber, P. H. (2013). MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology*, 94, 979. <https://doi.org/10.1890/12-1358.1>
- Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, 24, R288–R294. <https://doi.org/10.1016/j.cub.2014.01.056>
- Serafy, J. E., Faunce, C. H., & Lorenz, J. J. (2003). Mangrove shoreline fishes of Biscayne Bay, Florida. *Bulletin of Marine Science*, 72, 161–180.
- Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, 69, 82–90. <https://doi.org/10.1139/z91-013>
- Seyoum, S., McBride, R. S., Puchutulegui, C., Dutka-Gianelli, J., Alvarez, A. C., & Panzner, K. (2017). Genetic population structure of sheepshead, *Archosargus probatocephalus* (Sparidae), a coastal marine fish off the southeastern United States: Multiple population clusters based on species-specific microsatellite markers. *Bulletin of Marine Science*, 93, 691–713. <https://doi.org/10.5343/bms.2016.1069>
- Seyoum, S., McBride, R. S., Tringali, M. D., Villanova, V. L., Puchutulegui, C., Gray, S., & Bibber, N. V. (2018). Genetic population structure of the spotted seatrout (*Cynoscion nebulosus*): Simultaneous examination of the mtDNA control region and microsatellite results. *Bulletin of Marine Science*, 94, 47–71. <https://doi.org/10.5343/bms.2017.1060>
- Simms, A. R., Lambeck, K., Purcell, A., Anderson, J. B., & Rodriguez, A. B. (2007). Sea-level history of the Gulf of Mexico since the last glacial maximum with implications for the melting history of the Laurentide ice sheet. *Quaternary Science Reviews*, 26, 920–940. <https://doi.org/10.1016/j.quascirev.2007.01.001>
- Sullivan, J. P., Lundberg, J. G., & Hardman, M. (2006). A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 41, 636–662. <https://doi.org/10.1016/j.ympev.2006.05.044>
- Swift, D. G., O'Leary, S. J., Grubbs, R. D., Frazier, B. S., Gardiner, J. M., Drymon, J. M., Bethea, D. M., Wiley, T. R., & Portnoy, D. S. (2023). Philopatry influences the genetic population structure of the blacktip shark (*Carcharhinus limbatus*) at multiple spatial scales. *Molecular Ecology*, 32, 4953–4970.
- Thomas, L., Kennington, W. J., Evans, R. D., Kendrick, G. A., & Stat, M. (2017). Restricted gene flow and local adaptation highlight the vulnerability of high-latitude reefs to rapid environmental change. *Global Change Biology*, 23, 2197–2205. <https://doi.org/10.1111/gcb.13639>
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, 21, 272–281. <https://doi.org/10.1111/j.1466-8238.2011.00656.x>
- Vendrami, D. L. J., De Noia, M., Telesca, L., Handal, W., Charrier, G., Boudry, P., Eberhart-Phillips, L., & Hoffman, J. I. (2019). RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. *Scientific Reports*, 9, 7455. <https://doi.org/10.1038/s41598-019-43939-4>
- Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89, 438–450. <https://doi.org/10.1093/jhered/89.5.438>
- Webster, P. (2015). *Marine catfishes guild*. South Carolin Species of Conservation Concern: Supplemental Volume, 6 pp.
- Weir, B. S. (1996). *Genetic data analysis* (2nd ed.). Sinauer.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370. <https://doi.org/10.2307/2408641>

- White, N. J., & Butlin, R. K. (2021). Multidimensional divergent selection, local adaptation, and speciation. *Evolution*, 75, 2167–2178. <https://doi.org/10.1111/evo.14312>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Gold, J. R., & Portnoy, D. S. (2017). Haplotyping RAD loci: An efficient method to filter paralogs and account for physical linkage. *Molecular Ecology Resources*, 17, 955–965. <https://doi.org/10.1111/1755-0998.12647>
- Yáñez-Arancibia, A., & Lara-Domínguez, A. L. (1988). Ecology of three sea catfishes (Ariidae) in a tropical coastal ecosystem-southern Gulf of Mexico. *Marine Ecology Progress Series*, 49, 215–230.

**How to cite this article:** Portnoy, D. S., O'Leary, S. J., Fields, A. T., Hollenbeck, C. M., Grubbs, R. D., Peterson, C. T., Gardiner, J. M., Adams, D. H., Falterman, B., Drymon, J. M., Higgs, J. M., Pulster, E. L., Wiley, T. R., & Murawski, S. A. (2024). Complex patterns of genetic population structure in the mouthbrooding marine catfish, *Bagre marinus*, in the Gulf of Mexico and U.S. Atlantic. *Ecology and Evolution*, 14, e11514. <https://doi.org/10.1002/ece3.11514>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.