

An Analysis of the Spatial Genetic Structure of a Hybrid Zone Between Two Species of Killifish, *Fundulus heteroclitus* and *F. grandis*

By

Andrew Lee Hardy

A thesis submitted to the Department of Biology in
partial fulfillment of the requirements for the degree of Master of Science in Biology

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

December 2022

An analysis of the spatial genetic structure of a hybrid zone between two species of killifish, *Fundulus heteroclitus* and *F. grandis*

Drew Hardy

Abstract

Hybridization offers unique insight into the evolutionary process of speciation and the forces that maintain barriers to reproduction. In northeast Florida, a hybrid zone between two species of killifish, *Fundulus heteroclitus* and *F. grandis*, has been identified near Flagler Beach, FL, although the exact boundaries of the hybrid zone remain unknown (Gonzalez et al. 2009). This study aims to determine the fine-scale spatial genetic patterns of the hybrid zone and test the hypothesis that species ranges are influenced by changes in dominant vegetation. Results from an ADMIXTURE analysis and F_{ST} data indicated the boundary of the area of overlap between the two parental species to exist at Marineland, FL in the north and Tomoka Basin, FL in the south, while the boundaries of hybridization (i.e. the hybrid zone) existed in a smaller region between Bings Landing, FL and Tomoka Basin, FL. The area of overlap showed a highly mosaic pattern of allele frequencies, suggesting the spatial genetic structure of the hybrid zone is influenced by exogenous selection. Reproductive barriers were found to be the strongest in sites where *F. heteroclitus* was the more abundant species, but weaker in areas with predominantly *F. grandis*. No single environmental factor correlated significantly with the genetic distribution throughout the zone of overlap, and it is likely a combination of multiple factors that influence genetic variation, suggesting that habitat differences were likely important in the diversification of these two species.

Introduction

A hybrid zone is a geographic region in which two distinct species overlap in range and produce hybrid offspring of mixed descent (Arnold 1997). These zones present the opportunity to study the formation of barriers to reproduction between two closely related species and how these barriers work to maintain species differences (Buggs 2007; Barbas and Gilg 2018; Taylor and Larson 2019). Stable hybrid zones are typically maintained through a balance between dispersal of parental species into the zone and selection against hybrid genotypes within the zone (Bigelow 1965; Barton and Hewitt 1985). Therefore, hybrid zones provide us with opportunities to investigate how natural selection acts to maintain the spatial distribution of hybrids and parental forms across a geographic range.

Selection within a hybrid zone can often be broken down into two modes: exogenous, or environmentally dependent selection, and endogenous, where selection occurs against hybrids irrespective of the environment (Moore 1977). Through exogenous selection, the parental taxa are adapted to different environments and the fitness of hybrid genotypes varies across space due to one or more environmental factors (Moore 1977). Endogenous selection differs in that hybrids are selected against due to poor epistatic interactions (Dobzhansky 1936; Muller 1942; Orr and Turelli 2001) of the parental species' genomes in the hybrid individual. This form of selection is completely independent of the environment and is determined solely on the hybrid's genotype (Moore 1977).

The mode of selection against hybrids (exogenous or endogenous) forms the basis for a suite of hybrid zone models used to describe the spatial structure of hybridization in nature. Those models that feature exogenous selection include the ecotone (environmental gradient) model (Moore 1977), the bounded hybrid superiority model (Moore 1977), and the mosaic hybrid zone model (Harrison 1986). The ecotone model states that the hybrid zone occurs in concordance with a transition between habitats, with hybrid

fitness controlled exogenously corresponding to the environmental gradient found throughout the zone (Moore 1977). The bounded hybrid superiority model is supported whenever hybrids have higher fitness than parental taxa in an environment intermediate to that of the two parental species' native environments, typically in the center of the hybrid zone (Moore 1977). In the mosaic hybrid zone model, an amalgam of resources, habitat availability, or some other environmental factor controls the distribution of hybrid and parental types at each habitat (Harrison 1986). In this situation, certain environmental factors favor one parental genotype over the other in some regions or patches of the zone, while alternative environments favor the other parental genotype (Harrison 1986). Hybrid genotypes typically have lower fitness than each parental taxa in their respective habitats, although this can vary considerably across genotypes and space (Arnold and Hodges 1995; Arnold and Martin 2010). Conversely, under the tension zone model, gene flow from parental taxa and endogenous selection against hybrid offspring work to maintain the zone. Tension zones are expected to show consistent selective pressure against hybrids throughout the hybrid zone independent of the environment (Bigelow 1965). It should be noted, however, that many hybrid zones are not perfectly described by one theoretical model, as oftentimes there are varying patterns of fitness across loci within the hybrid zone (Arntzen and Wallis 1991; Bert and Arnold 1995; Campbell and Waser 2007).

With zones influenced by exogenous selection, landscape genetics can be a valuable tool in the study of hybridization to investigate which environmental factors are responsible for differences in fitness in different locations (Swenson 2008; Manel et al. 2003). Many experiments have made use of landscape genetics techniques to evaluate the genetic structure and evolutionary processes that occur throughout a geographic range of habitat (Piertney et al. 1998; Cassens et al. 2000; Segelbacher et al. 2008). One such technique combines environmental landscape data with genetic data using a geographic information system (GIS) to map out the genetic boundaries of habitat in populations (Swenson 2008; Manel et al. 2003). In addition to mapping, genotype-environment associations (GEAs) are extremely valuable tools to investigate the effects of environmental variability on genotypic distributions over geographic area. Together, these analyses have provided landscape geneticists the tools necessary to investigate key patterns in hybrid zone biology.

In northeast Florida, a hybrid zone exists between two species of closely related teleost fish, *Fundulus grandis* and *F. heteroclitus*, along the Florida Intracoastal Waterway (Gonzalez et al. 2009; Gilg et al. 2022). Previous research in this area by Gonzalez et al. (2009) identified an area of range overlap and hybridization located somewhere between Marineland, FL and Indian River Lagoon, FL. Due to the broad geographic scale of collections and the small number of genetic markers utilized in Gonzalez et al. (2009), however, the geographic range and spatial patterns of hybridization are not well known. Furthermore, whether the spatial structure of the hybrid zone is determined by environmental factors has not been documented. This region of overlap corresponds to an ecotone from salt-marsh ecosystems in the north to mangrove-dominated ecosystems in the south (Kangas and Lugo 1990), making it possible that the spatial structure of this hybrid zone is environmentally determined.

Historically, *F. heteroclitus* has a native range along the Atlantic coast of the United States, from Maine and Newfoundland in the north to northeast Florida in the south (Kneib 1986; Hardy 1978; Gonzalez et al. 2009). This range coincides with *Spartina*-dominated salt marsh habitats. *Fundulus grandis*, on the other hand, is known to live from parts of northeastern Florida throughout the coast of the Gulf of Mexico (Ellis and Bell 2004). *Fundulus grandis* is known to reside in mangrove-dominated marshes throughout a portion of this area (Ellis and Bell 2004), but Rozas and Reed (1993) noted that they often occupy salt marsh habitat similar to *F. heteroclitus* in the northern coastline of the Gulf of Mexico. Because of this disparity, populations of *F. heteroclitus* have been hypothesized to be adapted to salt marsh habitat, while

it is thought that *F. grandis* has adapted to mangrove swamp environments (Gilg et al. 2022). Each species occupies a similar ecological niche as opportunistic predators of many different marsh invertebrates and as prey to larger crabs, fish, and birds (Kneib 1986; Rozas and LaSalle 1990). It is currently unknown, however, how these two parental species and their hybrids utilize habitat in regions of mixed salt marsh and mangrove ecosystems and there is no direct evidence to support the hypothesis that the two killifish species are adapted to different vegetation.

Previous research has shown evidence of both endogenous and exogenous selection occurring with hybrids of these two species (Barbas and Gilg 2018; Gilg et al. 2022). Barbas and Gilg (2018) demonstrated evidence for endogenous selection in the form of lower developmental success of some hybrid zygotes in laboratory crosses, however the presence of endogenous selection in natural populations has not been investigated. Additionally, Gilg et al. (2022) provided some evidence for exogenous selection through a dynamic cohort analysis, which revealed an inconsistent pattern of selection over time at one site within the hybrid zone. It has yet to be determined, however, if these same patterns of selection are consistent throughout the hybrid zone, and whether allele frequencies are influenced by environmental factors.

The present study aims to address four main questions related to the spatial genetic structure of the hybrid zone: 1) What are the exact boundaries and width of the zone of overlap between these two species? The degree of distance between sample sites in Gonzalez et al. (2009) does not allow for specific edges to be known, so additional sampling was required to identify the genetic changes. 2) Is there evidence of exogenous selection occurring within the zone of overlap? With a tension zone model, genetic structure is uninfluenced by environmental factors, while the ecotone, hybrid superiority, and mosaic hybrid zone models inherently dictate that some genetic variation is due to changes in the environment. 3) Is the degree of hybridization consistent across space and how strong are reproductive barriers throughout the hybrid zone? Previous research (Barbas and Gilg 2018) has shown evidence for the presence of reproductive barriers in some laboratory crosses between species, yet the degree to which the barriers in nature correspond to laboratory patterns has not yet been evaluated. 4) Are there any environmental factors that correlate with the observed spatial genetic and fitness patterns? Hybrid zone models involving exogenous selection predict a direct correlation between one or more environmental variables and the distribution of parental and hybrid genotypes in populations throughout the zone. Conversely, in the case of solely endogenous selection against hybrids, no significant environmental correlations would be expected. Addressing this question will allow us to test the hypothesis that the spatial structure of this *Fundulus* hybrid zone is determined by differential adaptation of the two species to mangrove or marsh grass dominated habitats.

Materials and Methods

Sample Collection

Gonzalez et al. (2009) loosely described the boundaries of the hybrid zone to exist between Marineland, FL and Indian River Lagoon, FL along the Atlantic Intracoastal Waterway of Florida. Much of the land surrounding this region has been highly developed, thus appropriate marsh habitat for *Fundulus* populations is both scarce and scattered throughout the region. Roughly 30 individuals were collected during the summers of 2008 (from Gonzalez et al., 2009) and 2020 from reference populations existing outside the area of range overlap of *F. heteroclitus* and *F. grandis*. These were sampled to include individuals of each species that were unlikely to have been introgressed, providing us with species-

specific SNPs to analyze the spatial structure of the hybrid zone. The first reference population (*F. grandis*, referred hereafter as ‘G’) consisted of 27 individuals from Cedar Key, FL, collected in 2020, and 10 individuals from Port Saint Joe, FL, collected in 2008 by Gonzalez et al. (2009). The second reference population (*F. heteroclitus*, referred hereafter as ‘H’) consisted of 28 individuals collected from Saint Mary’s River, FL in 2020, and 1 individual collected near Georgetown, South Carolina by the University of South Carolina Baruch Marine Field Laboratory. The remaining sampling locations for the present study were chosen to include as many locations as possible between Crescent Beach, FL and Titusville, FL to increase the resolution of the current borders of the hybrid zone (**Figure 1**). Twelve non-reference sites were chosen for the present study listed North-South as follows: Matanzas Inlet (MI), Summer Haven (SH), Marineland (ML), Marineland South (MS), Bing’s Landing (BL), Herschel King (HK), Beverly Beach (BB), Flagler North (FN), Flagler Beach (FB), Highbridge (HB), Tomoka Basin (TB) and the Indian River Lagoon (IR) area (coordinates provided in **Supplemental Table 1**). With the exception of IR, at least 30 individuals from each site were collected using minnow traps and cast nets (exact numbers caught can be found in **Supplemental Table 1**). Fish were euthanized according to IACUC #19-007 and were then stored in 100% ethanol for subsequent genetic analysis. Individuals at each non-reference site were collected during the 2018 summer season (late-May to early-October) whenever possible, except from the sites HK, IR, and SH, from which individuals were collected in 2019 (HK), 2020 (IR), and 2021 (SH).

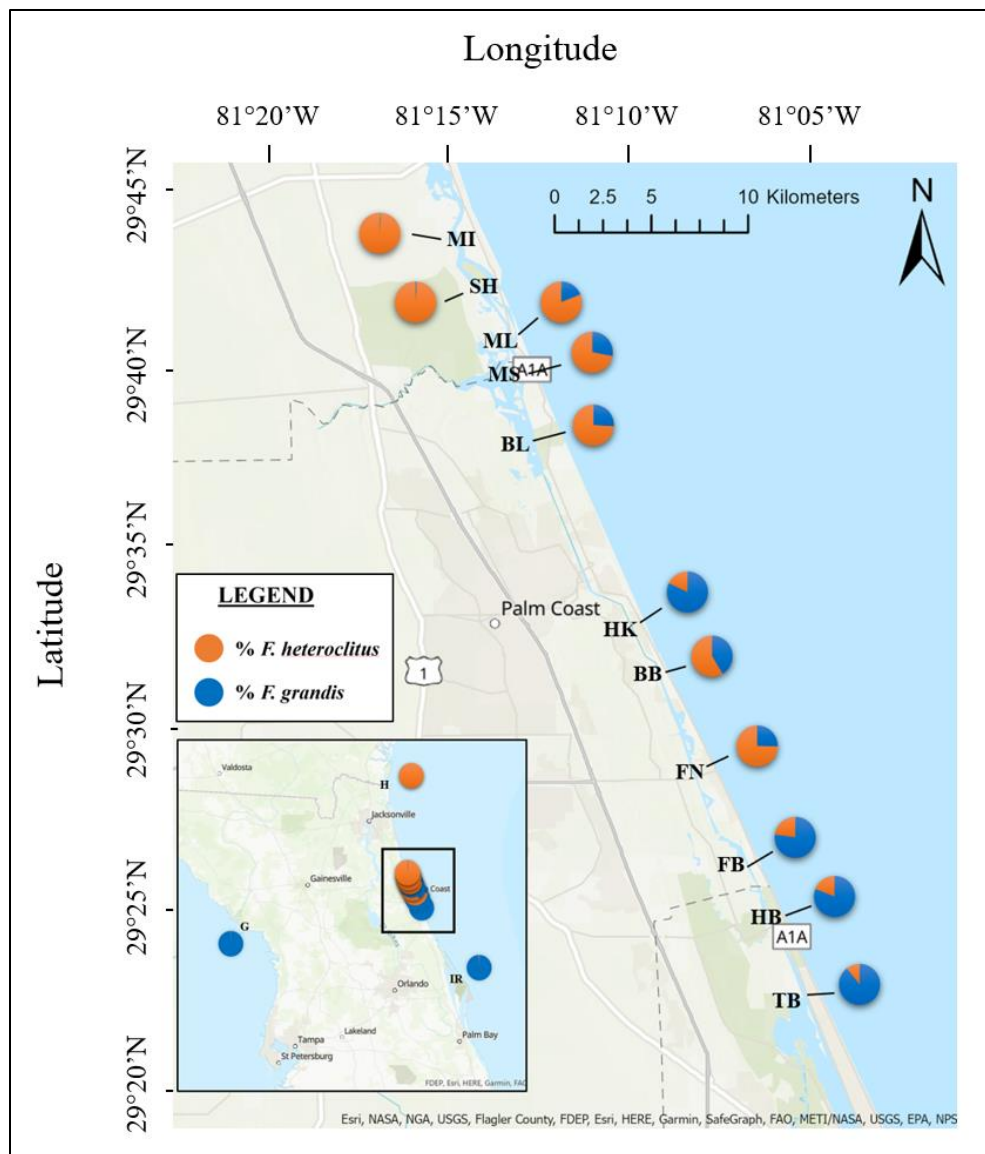


Figure 1: Map of the sampling region, showing the region of overlap (black box) between *Fundulus heteroclitus* and *F. grandis* in northeastern Florida. The location of each sampling site is marked with a two-letter site code (see **Supplemental Table 1**). Pie charts represent the mean admixture proportion of each site. The two reference populations, G and H, along with IR, can be seen outside the region of overlap between the two parental species. Orange area on the pie chart for each site represents the mean estimated percentage of *F. heteroclitus* ancestry, while blue area represents the mean estimated percentage of *F. grandis* ancestry.

Genetic Analysis

DNA was extracted from fin tissue of each individual using the phenol/chloroform method of DNA extraction (Sambrook and Russell 2001). Double digest restriction-site associated DNA (ddRAD) libraries were prepared according to the protocol in Peterson et al. (2012) with minor modifications. SphI-HF (NEB #R3182L) and MluCI (NEB #R3182L) were the two restriction endonucleases utilized, and a

fragment size selection window of 375-450 bps was used prior to index PCR. Unique index and barcode DNA oligos were added to fragments via ligation and PCR reactions (**Supplemental Table 2**). Four hundred thirty-seven total individuals over 14 sites were sequenced through the University of Florida Interdisciplinary Center for Biotechnology Research on the Illumina NovaSeq 6000 Next Generation DNA S4 sequencing platform using a double-end read of 100 base pairs each (2x100).

Data Processing and Population Genetics

Raw sequence reads were transformed using STACKS v2.41 (Rochette et al. 2019). The STACKS data processing pipeline was run *de novo* with the raw sequence data according to Catchen et al. (2013). First, PCR duplicates were removed and raw sequence reads were demultiplexed. Loci-building parameters were chosen to maximize the number of polymorphic loci found in 80% of individuals (the r80 method) via Paris et al. (2017). These included a value of 3 for the minimum number of raw reads required to form a stack (m), a value of 4 for the number of mismatches allowed between stacks to be merged into a putative locus (M), and a value of 4 for the number of mismatches allowed between stacks during construction of the catalog (n) (Paris et al. 2017). Individuals with less than 1,000,000 reads or less than 8x mean coverage were discarded from the catalog prior to data analysis. A random subset of 200 diagnostic loci (SNPs fixed for alternative alleles in the reference populations) were chosen for all downstream analyses.

A combination of ADMIXTURE v 1.3.0 (Alexander et al. 2009) plots and average pairwise F_{ST} comparisons, generated using the R package ‘hierfstat’ (v. 10; Goudet 2005), were used to investigate overall population structure and the genetic boundaries of the hybrid zone. To determine whether the genetic transition between parental types is smooth or mosaic, ADMIXTURE was used with a $K = 2$. This value of ‘K’ was chosen by running a cross validation test and analyzing which value of ‘K’ had the smallest cross validation error. Under endogenous selection, it is expected that environmental factors do not play a role in the distribution of parental genotypes in hybrid individuals, thus a smooth transition between parental types within the hybrid zone would be expected. Conversely, if the zone is influenced by exogenous selection, this transition would depend on the environmental characteristics throughout the zone. If the environmental characteristics responsible for genetic variation in these taxa experience a smooth transition through the hybrid zone, a similarly smooth transition between parental genotypes would be expected over this same region. Conversely, if the environmental characteristics form a mosaic pattern in the transition from one species range to the other, a similarly mosaic pattern of genotypes would be expected to match.

To investigate the degree of hybridization and strength of reproductive barriers throughout the zone, the ‘populations’ program within STACKS was used to calculate deviations from Hardy-Weinberg equilibrium within each population. A false discovery rate (FDR) correction was applied to correct for the effects of multiple comparisons. In addition, population F_{IS} was calculated using the R package ‘hierfstat’, while gametic disequilibrium between loci at each population was determined using the index of association (I_A/\bar{r}_d) calculated using the R package ‘poppr’ (v. 2.9.3; Kamvar, Tabima, and Grünwald 2014). Lastly, a hybrid index was calculated for each individual using a Bayesian MCMC approach with both a Gelman-Rubin diagnostic and an WAIC model correction using the R package ‘gghybrid’ (Bailey 2022). Populations with strong reproductive barriers between the two parental species will produce few hybrids, resulting in a lack of heterozygous and recombinant genotypes which leads to high values of F_{IS} and high associations among loci. Average hybrid indexes were calculated for each population and Pearson’s correlation coefficient was run pairwise between average hybrid index and F_{IS} , and average hybrid index and \bar{r}_d to further assess the relationship of reproductive barrier strength to multi-locus allele frequency patterns at each location.

Environmental Analysis

Environmental association analyses are widely used to investigate associations between genomic SNP data and landscape features that may play a role in the genetic structure of a study system (see Rellstab et al. 2015; Ahrens et al. 2018). Out of the many statistical packages to choose from, the redundancy analysis (RDA) (Bourret et al. 2014) lends the most use to our study due to the nature of the data (linear combinations of both genetic data and multiple environmental predictors). Monthly measurements of salinity, water temperature, and dissolved oxygen from 2017-2018 at 9 water quality stations within the Intracoastal Waterway were obtained from the Saint John's River Water Management District database and the GTM Reserve System-Wide Monitoring Program (SWMP). Interpolation of missing data was performed using the 'spline with barriers' spatial analyst tool in ArcGIS Pro (v. 2.9.0). The minimum and maximum measurements of the year for each water quality factor (salinity, temperature, and dissolved oxygen) were taken and used in the analysis. The area of cordgrass (*Spartina spp.*) and mangrove (*Avicennia germinans*) coverage in each habitat was measured using high-resolution satellite imagery from 2017 and 2018 provided by the Florida DEP Land Boundary Information System (LABINS) archive.

The RDA was used to test for associations between environmental variables and genotypes at multiple loci at each site throughout the hybrid zone using the 'psych' and 'vegan' R packages (Revelle 2021; Oksanen et al. 2020). Two separate analyses were run: the first with all 14 populations, and the second using only sites within the area of overlap (ML to TB). Environmental predictors were tested for correlations prior to running the model, as suggested by Dormann et al. (2012). Pairs of predictors with a correlation coefficient (r) greater than $|0.7|$ had one of the two removed from the dataset, while the other predictor remained in the analysis. In the first analysis, maximum water temperature was significantly correlated with minimum salinity, with an $r = -0.74$. Therefore, maximum water temperature was removed from the dataset before running. In the second analysis, 3 predictors (minimum salinity, maximum salinity, and minimum water temperature) all had multiple significant correlations with other variables in the dataset (1. min. salinity with max. salinity and both measures of water temp. 2. Max. salinity with min. water temp., Min. D.O. 3. Min. water temp. with both measures of salinity and max. D.O.), thus all 3 were removed prior to running the analysis. Variance inflation factors for each predictor were also assessed to ensure minimal issues with multicollinearity. The remaining environmental predictors for each analysis were then scaled by subtracting the mean from each value and dividing by the standard deviation. Significant associations between one or more environmental factors and genotype data in populations of mixed ancestry would be expected from this RDA if exogenous selection plays a role in the maintenance of this hybrid zone.

Results

ddRADsequencing/Processing

The Illumina Novaseq produced an average of 14,607,254 reads per individual, which was reduced to an average of 6,953,283 reads per individual after filtering. Individuals were genotyped at 32,302 loci and a 670-locus subset of diagnostic loci (loci with alleles fixed in both parental reference populations) was created. From this subset, a random sample of 200 loci were pulled for use in all downstream analyses.

Boundaries and Population Structure

The edges of the region of overlap were evaluated using a combination of the ADMIXTURE plot (**Figure 2**) and pairwise population F_{ST} comparisons (**Figure 3**). From the ADMIXTURE plot there is little indication of *F. grandis* north of ML and little evidence of *F. heteroclitus* south of TB. Sites between these two locations, however, tend to show the presence of alleles of both species. Pairwise F_{ST} comparisons further support that ML and TB represent the edges of the region of overlap. The *F. heteroclitus* reference site (H) shows very little genetic distance from the sites MI and SH ($F_{ST} = 0.011$, 0.038 , respectively), but a sharp increase in F_{ST} is observed when compared to ML ($F_{ST} = 0.166$). Likewise, in the south, the *F. grandis* reference site (G) shows a slight difference from IR ($F_{ST} = 0.079$) but becomes more dissimilar when compared to TB ($F_{ST} = 0.112$). Based on these data, the boundaries of the area of overlap appear to be from ML in the north to TB in the south.

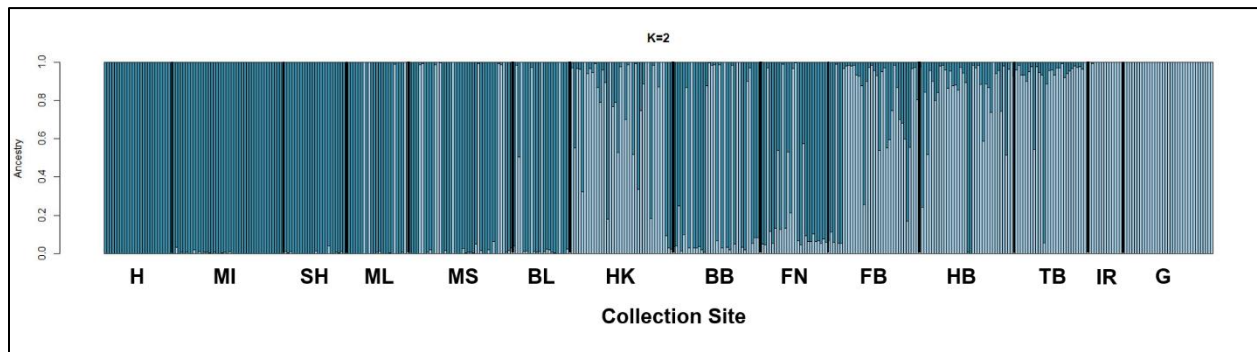


Figure 2: ADMIXTURE v 1.3.0 (Alexander et al. 2009) ancestry plot of region of overlap between *F. grandis* and *F. heteroclitus* in northeastern Florida with a $K=2$ ($n=437$). In this figure, ‘H’ represents the *F. heteroclitus* reference population, while ‘G’ represents the *F. grandis* reference population. Individuals are represented by each vertical blue bar, while populations are segmented by each thick black bar. Darker blue signifies the ancestry from *F. heteroclitus*, while the lighter blue signifies ancestry from *F. grandis*.

The genetic structure of parental types and their hybrids throughout the zone of overlap forms a mosaic pattern, with some sites showing higher proportions of individuals with predominantly *F. grandis* ancestry and others showing higher proportions of individuals with predominantly *F. heteroclitus* ancestry, independent of latitude (**Figures 2 and 3**). For example, near the center of the area of overlap, the population structure switches from predominantly *F. heteroclitus* sites (at BL) to one dominated by *F. grandis* and hybrids with high proportions of *F. grandis* ancestry (HK) (Figure 2). The pattern seen at HK is quickly reversed at the next two sites, BB and FN, until *F. grandis* genotypes become very common in individuals at the rest of the southern sites (FB, HB, TB, and IR). These same patterns are also reflected in the pairwise F_{ST} comparisons, where jumps in both directions are observed between sites when moving from one reference population to the other. This closely matches the predictions of the mosaic hybrid zone model, where fluctuations in allele frequency across space are expected.

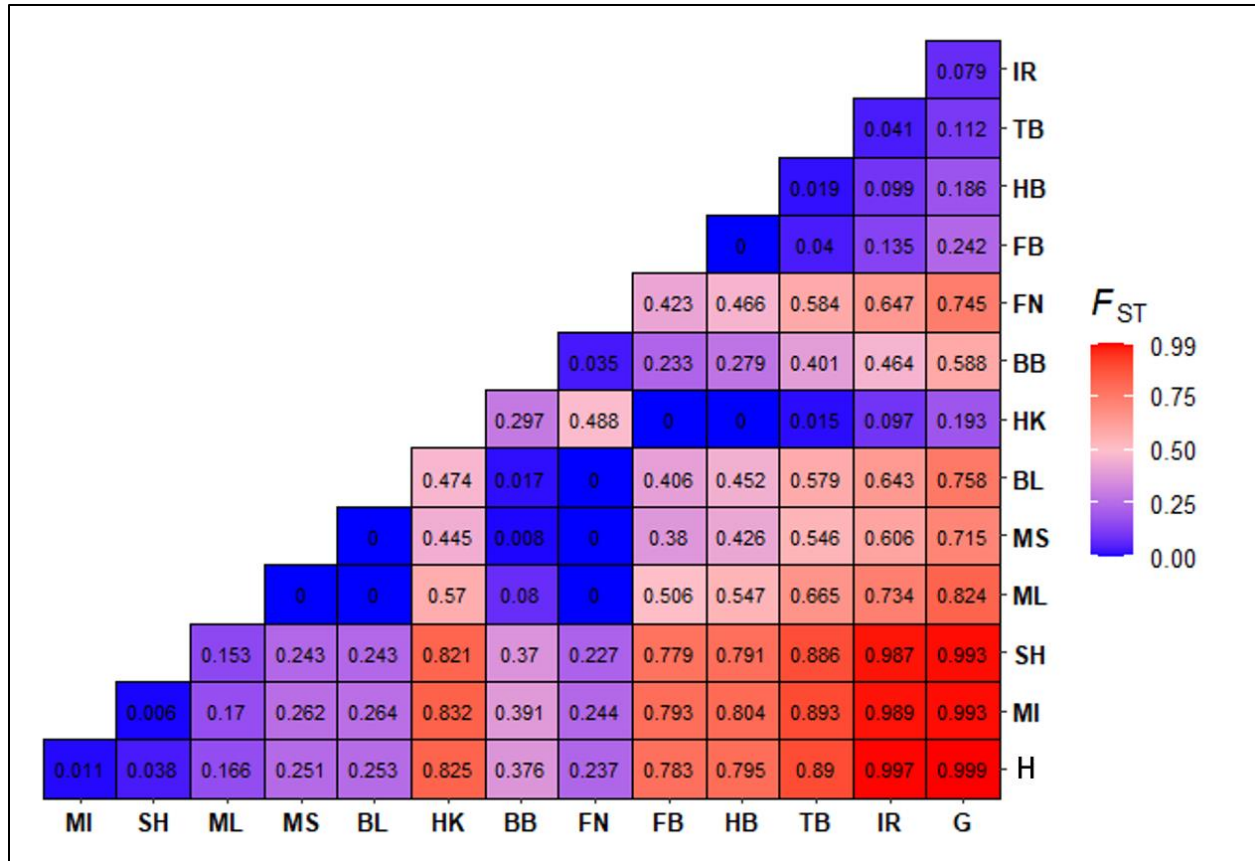


Figure 3: Pairwise F_{ST} comparisons between all sampled populations. In this figure, “H” represents the *F. heteroclitus* reference population, while “G” represents the *F. grandis* reference population. Confidence intervals (95%) for each comparison are shown in **Supplemental Table 4**.

Degree of Hybridization and Reproductive Barriers

The degree of hybridization and strength of reproductive barriers were evaluated using deviations from Hardy-Weinberg equilibrium (HWE), measures of population F_{IS} , and gametic disequilibrium. The sampled populations could be placed into three different categories: 1) those populations fixed or nearly fixed for one parental allele at all loci in most individuals, 2) populations with both parental species present but little evidence of hybridization, and 3) populations with both parental species present that feature extensive evidence of hybridization. Sites fixed for one parental allele at all loci in most individuals included MI and SH (*F. heteroclitus* alleles), and IR (*F. grandis* alleles). These sites experienced very few loci deviating from HWE (0-1 out of 200; **Figure 4**), low F_{IS} values (0 - 0.00416; **Table 5**), and index of association values (\bar{r}_d) close to 0 (0.006 – 0.052; **Table 5**). Sites with both parental species present but little to no evidence of hybridization included ML, MS, BL, and BB. The number of loci deviating from HWE in this group was very high (199-200 out of 200) due to a large deficit of heterozygotes, ($F_{IS} = 0.839 - 0.964$). Estimates of \bar{r}_d for these sites were also among the highest observed in the area of overlap, with a range of 0.606 – 0.877 suggesting a lack of recombinant genotypes. The combination of high values for all three of these statistics signifies strong barriers to reproduction and therefore limited hybridization in these populations. Lastly, the group of sites with both species present and evidence of extensive hybridization include HK, FN, FB, HB, and TB. These sites showed

intermediate numbers of loci deviating from HWE (34-161 out of 200), and strong, but substantially lower values of both F_{IS} (0.289 – 0.535) and \bar{r}_d (0.266 – 0.495), than observed in the previous group of sites. These values indicate substantially weaker interspecific barriers to reproduction at HK, FN, FB, HB, and TB, and more hybridization.

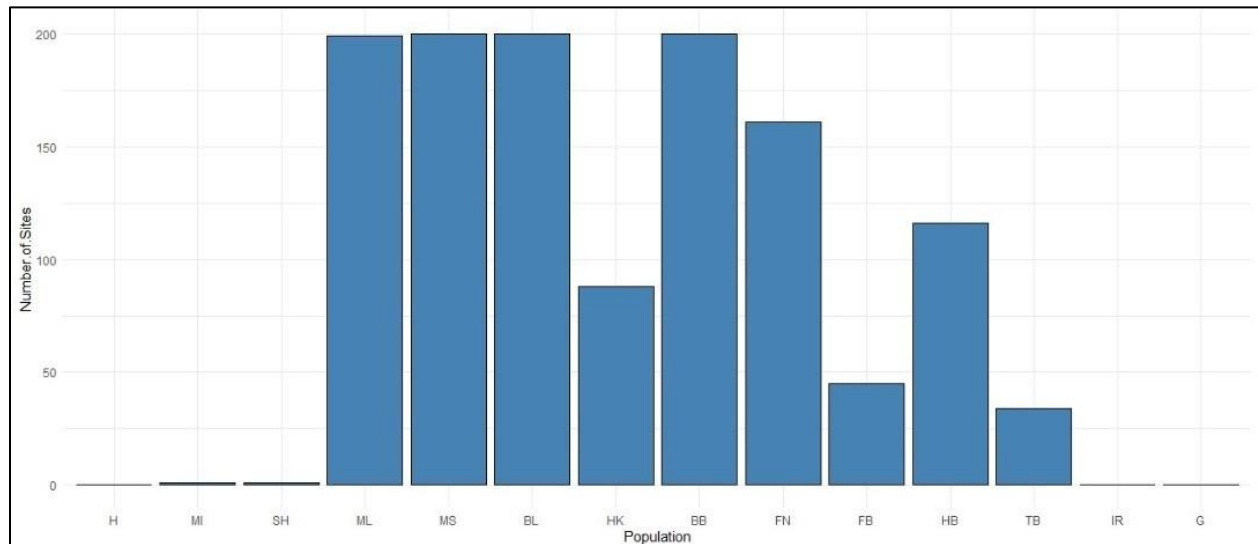


Figure 4: Number of variable sites (out of 200) found to be significantly deviating from HWE ($p < 0.05$) in each population, corrected using the false discovery rate (FDR). Sites ‘H’ and ‘G’ represent the *F. heteroclitus* and *F. grandis* reference populations, respectively.

Table 5. Wright’s F_{IS} (inbreeding coefficient) and index of association standardized for sample size (\bar{r}_d) for collection sites MI-IR. 95% confidence intervals for F_{IS} measurements and p-values ($\alpha=0.05$) created from a distribution of 999 permutations for index of association measurements are provided. Bold p-values indicate statistical significance.

Site	F_{IS}	95% Confidence Interval	Index of Association (\bar{r}_d)	p-value ($\alpha=0.05$)
MI	0.0029	[-0.023, 0.105]	0.006	0.068
SH	0.00416	[-0.018, 0.091]	0.052	0.001
ML	0.96374	[0.948, 0.972]	0.877	0.001
MS	0.94989	[0.940, 0.957]	0.791	0.001
BL	0.85758	[0.845, 0.870]	0.798	0.001
HK	0.38874	[0.373, 0.444]	0.266	0.001
BB	0.83939	[0.828, 0.856]	0.606	0.001
FN	0.53511	[0.511, 0.555]	0.495	0.001
FB	0.28879	[0.249, 0.316]	0.349	0.001
HB	0.43099	[0.396, 0.442]	0.426	0.001
TB	0.32685	[0.286, 0.346]	0.461	0.001
IR	-0.00042	[-0.023, 0]	0.018	0.762

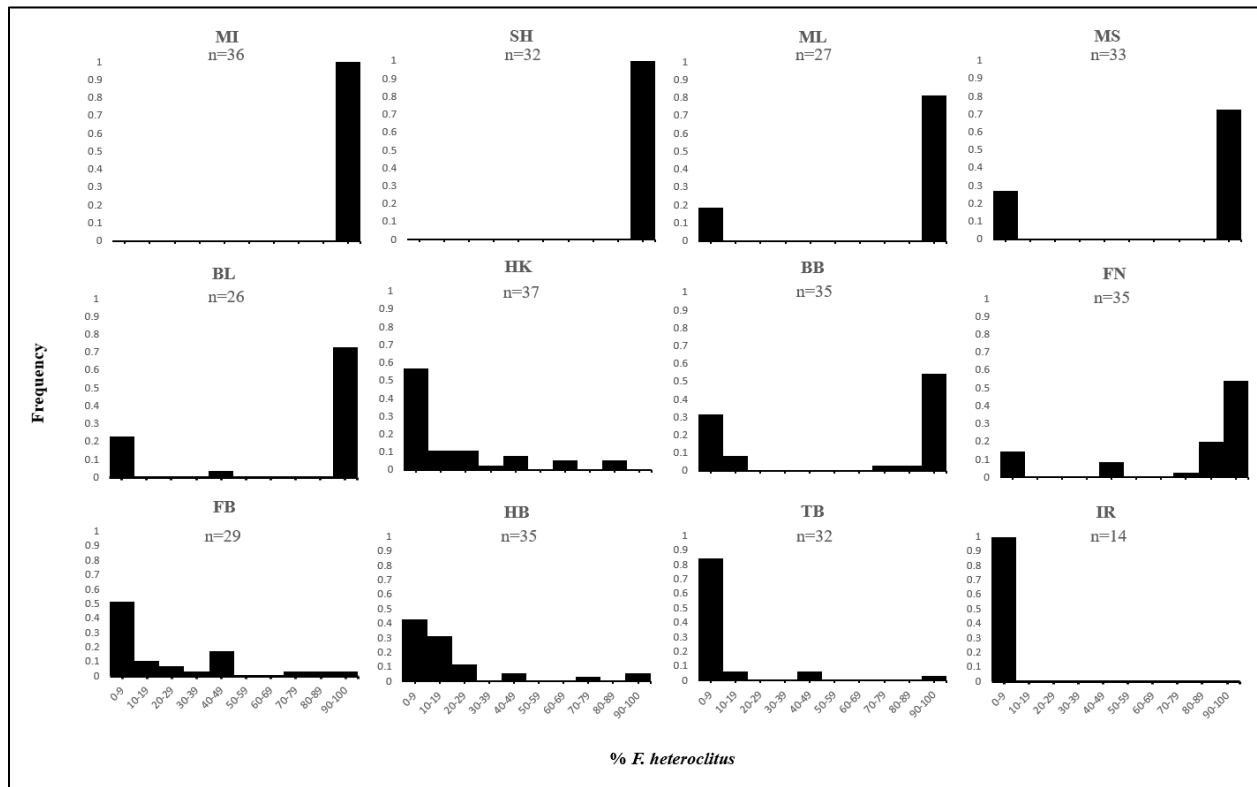


Figure 5: The frequency of individual hybrid index scores, measured as the percentage of *F. heteroclitus* alleles present at 200 diagnostic loci at 12 sampling sites in the area of overlap. Individual hybrid indexes were calculated using a Bayesian MCMC approach with a WAIC model correction.

While the strength of reproductive barriers varied considerably among sites, hybridization was relatively rare at all sites. The distribution of hybrid indexes was shown to be largely bimodal throughout the region of overlap (**Figure 5**), where most individuals at all sites were either a pure *F. heteroclitus* or *F. grandis* at the loci investigated. Out of the 115 hybrids throughout the area of overlap, eight (6.9%) were putatively F1 hybrids, with hybrid indexes of 0.46 – 0.54. The rest (93.1%) of the hybrids were thought to be multi-generation or backcrossed individuals, based on hybrid index scores (0.05–0.46; 0.54–0.95). The putative F1 hybrids were found at 6 sites within the area of overlap: BL (1) HK (2), FB (1), FN (2), HB (1), and TB (1). Overall, hybridization was much rarer in sites where *F. heteroclitus* was the common parental species, and hybridization was more common in sites where *F. grandis* was the more common species. When observing the mean hybrid index (mean percentage of *F. grandis* alleles) for each site plotted with F_{IS} and \bar{r}_d values (**Figure 6**), it becomes clear a correlation exists between the three. In fact, pairwise combinations of mean hybrid index with F_{IS} and \bar{r}_d produced significant positive correlations (HI x F_{IS} : $r = 0.874$, $p = 0.002$; HI x \bar{r}_d : $r = 0.836$, $p = 0.005$). Given this, it can be inferred that reproductive barriers to hybridization are strongest in areas with more *F. heteroclitus* and weaker in areas with more *F. grandis*.

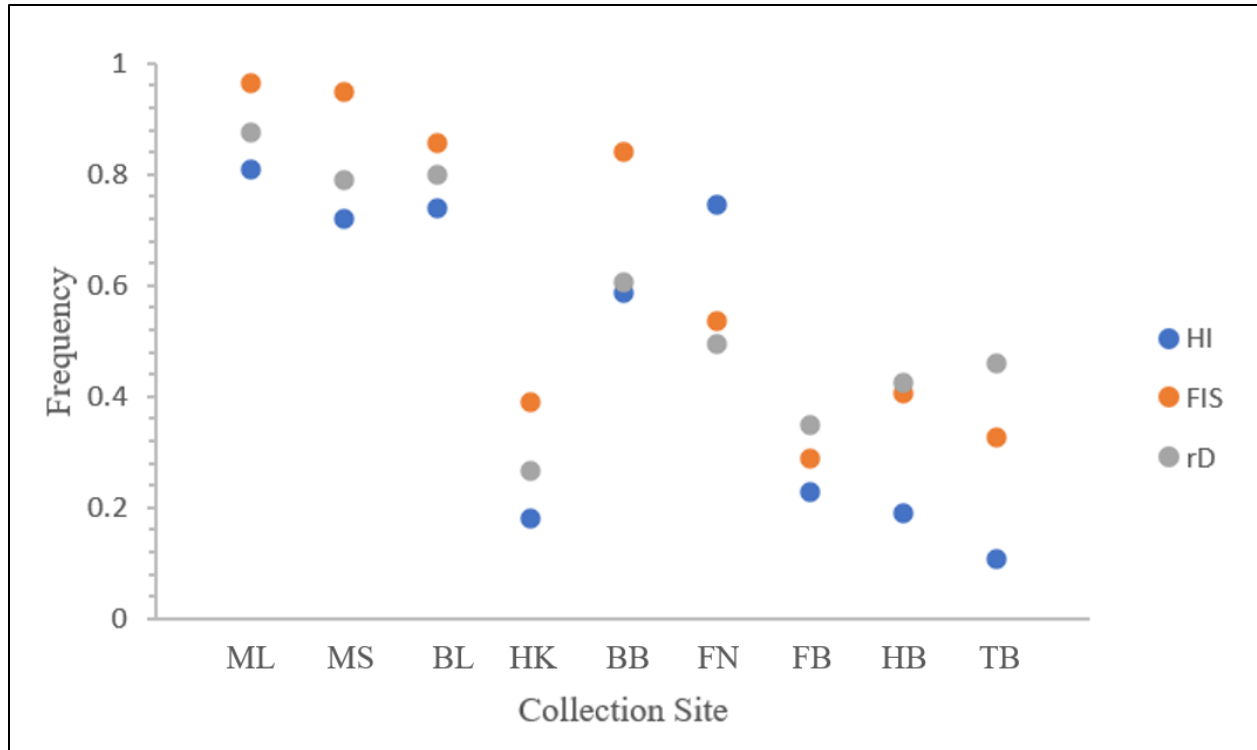


Figure 6: Comparison of the F_{IS} and \bar{r}_d of each population within the area of overlap (ML to TB) to the corresponding mean hybrid index (the average proportion of *F. heteroclitus* alleles in an individual at that site).

Environmental Analysis

The redundancy analysis was first run with all 14 populations, the results of which can be seen plotted in **Figure 7**. Following Montgomery and Peck (1992), a variance inflation factor (VIF) threshold of 10 or less was used, which we deemed was an acceptable trade-off between pruning factors out of the analysis and minimizing collinearity. VIF scores for all remaining variables (minimum salinity, maximum salinity, minimum water temperature, minimum dissolved oxygen, maximum dissolved oxygen, and mangrove area) after pruning were all below 7, with a low of 1.44 (minimum salinity) and a high of 6.53 (minimum water temperature). After performing the analysis, the first loading axis, RDA1, was found to explain 98% of the variance in the genetic data, with RDA axes 2-6 explaining the other 2%. Therefore, all analyses were focused on the first RDA loading axis.

Plotting the two RDA axes most responsible for the variation, five main clusters of individuals were identified, grouped by hybrid index (the proportion of *F. grandis* alleles in an individual): *F. heteroclitus* (HE), *F. heteroclitus*-like (HL), mixed (MX), *F. grandis*-like (GL), and *F. grandis* (GR) (**Figure 7**). The first cluster, HE, consists of individuals from H, MI, SH, ML, MS, BL, BB, and HB that had a hybrid index < 0.05 . The second group, HL, contains individuals primarily from BB and FN, and included all individuals with a hybrid index between 0.05 and 0.30. The mixed group, MX, had hybrid indexes of 0.3 – 0.7 and consisted of individuals from BL, FB, FN, HB, HK, and TB. The last large set of individuals were from GL and GR, which had hybrid index ranges of 0.70 – 0.95 and 0.95 – 1.0, respectively. All sites except for H, MI, and SH had at least one representative in the GL or GR clusters.

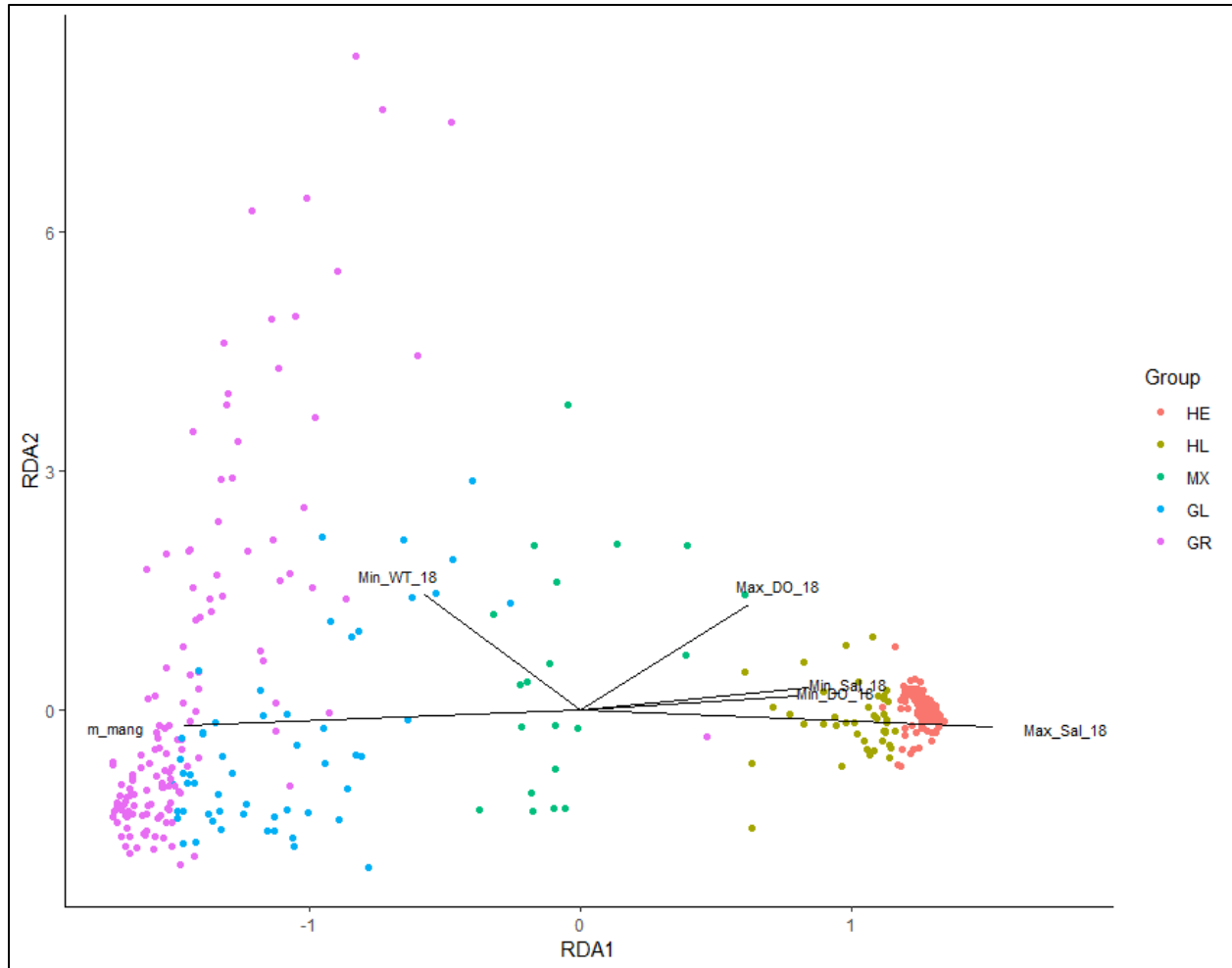


Figure 7: RDA loading axis 1 plotted against RDA loading axis 2 (all 14 populations). Individuals were split into 5 groupings based on hybrid index: *F. heteroclitus* (HE), *F. heteroclitus*-like (HL), mixed (MX), *F. grandis*-like (GL), and *F. grandis* (GR) ($0 < GR < 0.05 < GL < 0.30 < MX < 0.70 < HL < 0.95 < HE < 1$). Vectors indicate environmental predictor variables, with the length of the vector used to indicate the strength of that predictor. The direction of the vector indicates the individuals it correlated positively with, while negative correlations are indicated opposite the direction of each vector.

Two environmental predictors correlated strongly with RDA loading axis 1: maximum salinity ($r = 0.76$) and the area of mangroves found at each site ($r = -0.72$). The next closest drivers of RDA1 were minimum salinity ($r = 0.42$) and minimum dissolved oxygen ($r = 0.40$), while the last two remaining factors (min. water temperature and max. dissolved oxygen) presented smaller values (-0.29 and 0.31 , respectively). No one environmental factor was the clear driver of genetic variation, but a combination of mangrove area, max. salinity, min. salinity, and min. dissolved oxygen appear to combine to drive the variation in the sampled individuals within these clusters. Individuals with predominantly *F. grandis* genotypes correlated positively with mangrove area, but negatively with measurements of salinity and minimum dissolved oxygen. Individuals with predominantly *F. heteroclitus* genotypes experienced largely the opposite patterns, correlating negatively with mangrove area and positively with minimum and maximum salinity

and minimum dissolved oxygen. For the mixed (MX) group, minimum water temperature and maximum dissolved oxygen were significant drivers of genetic variation.

To determine if the same environmental factors are important for determining genetic structure within the area of overlap, a second RDA was performed using only sites between ML and TB. Variance inflation factors for variables after correlation filtering (max. water temperature, min. dissolved oxygen, max. dissolved oxygen, and mangrove area) were below 5, with a high of 4.22 (min. dissolved oxygen) and a low of 2.41 (max. dissolved oxygen), indicating minimal effects from multicollinearity. Similar to the previous analysis, RDA loading axis 1 (RDA1) explained 96% of the genetic variance, with axes 2-4 explaining the remaining 4%. Focusing on the RDA1 axis again, we were able to identify similar patterns as seen in the first RDA (**Figure 8**). Individuals in the GR and GL groups clustered together, while individuals from the HE and HL groups clustered on the opposite side of the RDA 1 axis, with individuals of mixed ancestry in between.

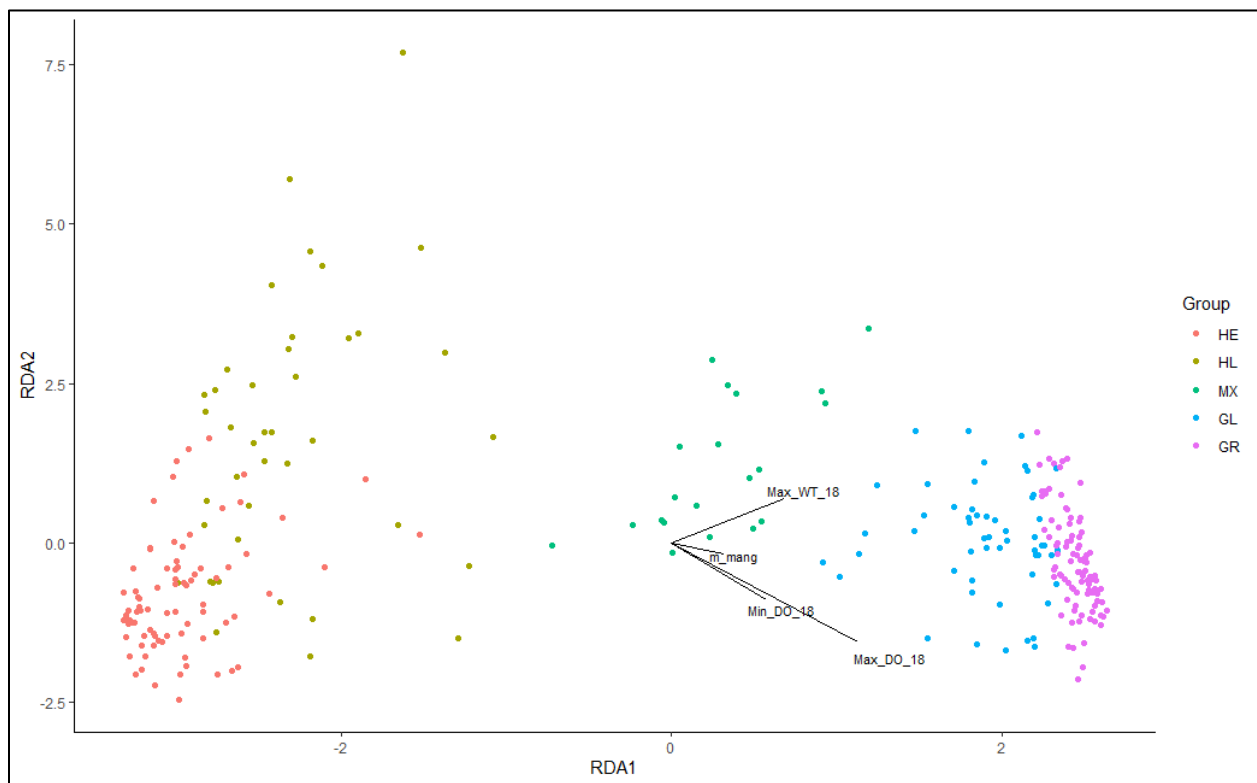


Figure 8: RDA loading axis 1 plotted against RDA loading axis 2 (9 populations: ML to TB). Individuals were split into 5 groupings based on hybrid index: *F. heteroclitus* (HE), *F. heteroclitus*-like (HL), mixed (MX), *F. grandis*-like (GL), and *F. grandis* (GR) ($0 < GR < 0.05 < GL < 0.30 < MX < 0.70 < HL < 0.95 < HE < 1$). Vectors indicate environmental predictor variables, with the length of the vector used to indicate the strength of that predictor. The direction of the vector indicates the individuals it correlated positively with, while negative correlations are indicated opposite the direction of each vector.

In the second RDA, only one environmental predictor, max. dissolved oxygen, correlated somewhat highly with RDA1 ($r = 0.56$). The other three variables all showed weak correlations with RDA1 (max. water temperature $r = 0.34$, min. dissolved oxygen $r = 0.29$, area of mangroves $r = 0.16$). As seen in the first analysis, a combination of all four environmental predictors are responsible for the variation present

in the genetic dataset, with no single predictor carrying a significant amount of weight over the others. Individuals with majority *F. grandis* ancestry correlated positively with all four predictors, while negative correlations were found for all four predictors in the clusters of individuals with majority *F. heteroclitus* ancestry. The mixed ancestry group (MX) was slightly positively correlated with maximum water temperature and did not show strong correlations to the other three environmental predictors. In general, all factors correlated much weaker with all individuals in the second RDA compared to the first RDA run with all 14 populations.

Discussion

Boundaries, Hybridization, and Reproductive Barriers

Data from the present study suggests the boundaries of the zone of overlap between *F. heteroclitus* and *F. grandis*, are at Marineland, FL (ML) in the north, and Tomoka Basin, FL (TB) in the south, a width of roughly 37 km. Previous estimates of the range of overlap from Gonzalez et al. (2009) found no significant evidence of *F. grandis* at any site from ML northward, which suggests recent movement of the area of overlap northwards into ML. A clear division exists between the absence and presence of *F. grandis* between sites SH and ML in the north, a distance of only 4 km. In the south, a less distinct transition exists between the presence and absence of *F. heteroclitus* from sites TB to IR. The distance between TB and IR is roughly 112 km, while the population pairwise F_{ST} data suggests these two populations are more genetically similar than SH is to ML. There was evidence of one individual very close to being considered ‘pure’ *F. heteroclitus* (hybrid index = 0.06) that was sampled at TB, while all fish sampled at IR were shown with high confidence to have hybrid indexes greater than 0.98, suggesting these are all individuals of ‘relatively pure’ *F. grandis* with some evidence of introgression. It is intriguing, however, that some alleles of *F. heteroclitus* were able to introgress into the IR population, given the vast distance between sites TB and IR (roughly 115 km), and the low presence of *F. heteroclitus* individuals at TB (1 individual with hybrid index = 0.06). In addition, the large metropolitan area of Daytona sits in between sites, likely removing any typical *Fundulus* habitat that would be naturally present. Therefore, it is unlikely for individuals of *F. heteroclitus* to have recently migrated to IR, but rather this could be a signature of past introgression and potentially a range shift for *F. heteroclitus*. Our low sample size of only 14 individuals at IR, however, should not be ignored and this data should therefore be observed with caution.

What is also intriguing is that the small amount of *F. grandis* individuals sampled at northern sites (ML, MS, and BL) were certainly outnumbered by *F. heteroclitus* individuals but appear to have close to 100% *F. grandis* ancestry, providing no evidence of hybridization with individuals of *F. heteroclitus*. Therefore, actual evidence of hybridization, i.e. the boundary of the hybrid zone, did not entirely align with the zone of overlap between the two parental species, described above. The first evidence of hybridization was found at BL, six kilometers south of ML. The southern boundary of the hybrid zone, however, coincided with the first substantial evidence of *F. heteroclitus* at TB. There may be several explanations for this discrepancy. First, it is possible that several individuals of *F. grandis* were recently introduced to the area (ML, MS, and BL), whether by natural migration or introduction via fisherman, and have not had sufficient time to hybridize. Both *F. grandis* and *F. heteroclitus* are popular baitfish for fisherman in the area and are sold in many bait shops along the Florida Atlantic Coast (Brown et al. 2011; Mulvey et al. 2003). While it could be the case that some of these introduced fish were sampled in this study, it would seem unlikely that this introduction has only happened recently at some sites but has been occurring for longer periods of time at others throughout the zone. It is also possible these *F. grandis* individuals

recently migrated (within one year of collection) north to ML, MS, and BL, either during a single migration event from a nearby site (like HK) or by ‘jumping’ from one patch of habitat to the next northwards over several generations. This is also unlikely since both species are known to have small home ranges (less than 650 meters) and little to no migratory behaviors (Lotrich 1975, Sweeney et al. 1998, Jensen et al. 2019, Nelson et al. 2014). Further evidence against this idea is found in the F_{ST} data, which suggests several adjacent sites throughout the zone of overlap are soundly differentiated and little gene flow is occurring between them.

Secondly, it is possible that a small group of *F. grandis* have existed at these northern sites (ML, MS, and BL) for several generations. If this is the case, there seems to exist a conspecific mating preference either among individuals of *F. heteroclitus* or among the few individuals of *F. grandis* at these northern sites (ML, MS, and BL) that results in the bimodal genetic pattern seen at these sites in **Figure 2**. This pattern is observed to a much lesser extent near the southern boundary of the area of overlap, where both parental species exist and evidence of hybridization has been observed. A previous study by Barbas and Gilg (2018) quantified the reproductive isolating barriers between *F. heteroclitus* and *F. grandis* in a laboratory setting with choice and no-choice matings. While barriers to reproduction existed, they were found to be incomplete and differed in strength depending on the direction of the cross. Specifically, females of *F. heteroclitus* experienced almost complete reproductive isolation when crossed with a male of *F. grandis* under choice conditions, while reproductive isolation in the reverse cross (*F. grandis* female x *F. heteroclitus* male) was considerably weaker. This may explain the tendency for *F. heteroclitus* females in these northern sites to prefer and mate selectively with males of the same species over the few males of *F. grandis* also present in the area. Lastly, it should not be ignored that our somewhat small sample sizes of roughly 30 individuals per site in combination with the possibility that F1 hybrids are rare in most populations could be insufficient to detect the presence hybrids by chance alone. It could be the case that a small amount of F1 hybrids are present but were not sampled, leading us to believe hybridization does not occur at these sites when in fact it does, albeit rarely.

All sites throughout the area of overlap, even those that showed evidence of hybridization, had a deficiency of hybrid genotypes compared to expectations. This lack of heterozygosity throughout the area of overlap could be explained by either a low incidence of interspecific matings, low survival of hybrids, or some combination of both. If the patterns of reproductive isolation observed in Barbas and Gilg (2018) also exist in nature, higher levels of hybridization would be expected in areas where *F. grandis* is the more common species. The data presented here could suggest this to be the case, as we observed just that. In addition, this isolation could result in the rarity of heterospecific matings, which would reveal itself in the form of a lack of F1 hybrids compared to expectations from random mating. This F1 deficiency is also present in our data at all sites (12 of the 155 hybrids thought to be F1s), but extremely present at sites with a majority *F. heteroclitus* hybrid index (only 3 out of 12 putative F1s; **Figure 5**). Similar assortative mating has been hypothesized to maintain bimodal distributions and a lack of F1 hybrids in a *F. heteroclitus* subspecies hybrid zone (McKenzie et al. 2016) and a *F. notatus*/*F. olivaceus* hybrid zone (Duvernell et al. 2007). Assortative mating is also a common factor in maintaining hybrid deficiencies in many non-Fundulus hybrid zones (see MacCallum et al. 1998, Cruzan and Arnold 1993, and Arntzen et al. 2014), indicating its importance and pervasiveness in hybrid zone biology among many different taxa. Therefore, it seems plausible that assortative mating by females of *F. heteroclitus*, when those females have a plethora of conspecific mates to choose from, could be a valid explanation of the lack of hybridization observed at ML, MS, and BL.

Alternatively, it is possible selection (either endogenous or exogenous) against hybrids is occurring throughout the area of overlap, causing hybrids to be rare. Barbas and Gilg (2018) also showed some

evidence of post-zygotic selection against F1 hybrids in some but not all laboratory crosses in the form of hatching success, although these were weaker in comparison to pre-zygotic barriers under choice matings, which would be expected to occur in nature. In addition to endogenous selection, exogenous selection could play a role in the fitness of hybrids (as was suggested by Gilg et al. 2018 at FB) or could assist in preventing heterospecific matings (see MacCallum et al. 1998). Previous research has also suggested that selection commonly combines with the effects of assortative mating to contribute to hybrid deficiencies (Duvernell et al. 2007; McKenzie et al. 2016; Arntzen et al. 2014).

The deficiency of F1 hybrids compared to multi-generation or backcrossed hybrids is of interest as it suggests two things: either heterospecific matings are rare or stronger selection against F1s exists that is not present for multigeneration or backcrossed hybrids (or combinations of both). In addition, the large proportion of hybrids (147/155) that are thought to be backcrossed or multigenerational hybrids suggests F1 hybrids are fertile to some degree. This brings about another question: do parental types favor matings with hybrid individuals over heterospecific matings with the other parental species? No experimental crosses between F1 and parental types have been performed in a laboratory setting for these species, but evidence of a preference for male hybrids by parental types has been shown to exist in a swordtail hybrid zone, and could potentially be occurring in this system as well (Powell 2019). No matter the cause of hybrid deficiencies, it is clear reproductive barriers tend to be stronger in sites with higher proportions of *F. heteroclitus* within the area of overlap (ML, MS, and BL) and weaker throughout the rest of the region.

Mosaic Nature and Environmental Influences of the Hybrid Zone

The spatial genetic structure observed in the present study showed significant evidence of a mosaic hybrid zone between these two species. The mosaic hybrid zone model states that a varying pattern of allele frequencies occurs throughout the transition between parental species, resulting in sites with higher allele frequencies for one parental type in some sites and lower in others. A classic example of this type of hybrid zone was described by Richard Harrison (1986), where assortative mating was found to be a primary factor that maintained a field cricket hybrid zone. As mosaic hybrid zones also indicate the influence of exogenous selection working in the zone, soil type was suggested by Harrison (1986) to be a potential environmental parameter responsible for variation in allele frequencies in the field cricket system. This hypothesis was later confirmed by Rand and Harrison (1989) when evidence for significant genetic differentiation based on soil type was found in this same system.

Like the example described by Rand and Harrison (1989), many mosaic hybrid zones have been detailed in the literature, however, few have been able to determine the environmental parameter(s) responsible for exogenous selection within the hybrid zone. In 2013 a darter mosaic hybrid zone was described by Bossu and Near. While the group found evidence of a strongly bimodal genotypic distribution, with mate choice a primary contributing factor, work was not performed to examine any environmental influence on the distribution. Likewise, a nase hybrid zone in a freshwater system in France (Sinama et al. 2013) and a reed frog hybrid zone on São Tomé Island (Bell and Irian 2019) have mosaic systems with unexplored causes of exogenous selection. Oftentimes, assortative mating, endogenous selection, and recent contact between parental types are listed as co-factors with exogenous selection in working to maintain the spatial genetic patterns of each zone.

Several hybrid zones between two different species of *Fundulus* are known to exist in nature, and of those, the *F. heteroclitus*/*F. grandis* species complex studied here is the only known example that fits many aspects of a mosaic hybrid zone. Both the *F. heteroclitus heteroclitus*/*F. heteroclitus macrolepidotus* complex (McKenzie et al. 2016) and the *F. olivaceus*/*F. notatus* complex (Duvernell et al. 2007; Schaefer et al. 2016) show a smooth transition from one parental subspecies to the other. While

both suggest an exogenous influence on selection within their respective hybrid zones, and those causes of exogenous selection remain largely unknown, the mosaic patterns of environmental characteristics and genetic distributions of hybrid and parental types are not present.

When Gonzalez et al. (2009) first described the hybrid zone between *F. heteroclitus* and *F. grandis*, they noted it coincided with an ecotone between marshgrass-dominated marshes in the north and mangrove-dominated swamps in the south. This led to the hypothesis that the two species of fish may be adapted to different types of vegetation, which would then determine the distribution of hybrid and parental types throughout the zone. Additional evidence for the presence of exogenous selection in the zone was provided by Gilg et al. (2022), when they showed differential survival among genotypes between the first and second year classes as opposed to early in development, but that study did not identify the factor that led to differential survival. Lastly, data from the current study provides evidence for the hybrid zone adhering to the mosaic model, suggesting a clear environmental influence on selection within the system.

In an attempt to test whether vegetation or other environmental factors are driving the observed genetic patterns in the present study, redundancy analyses were employed. An initial analysis of all 14 sampling locations (including the reference sites) provided support for the hypothesis that the genetic patterns are largely driven by differences in vegetation (**Figure 7**), however, these correlations mostly disappeared when removing the reference sites and sites in which one allele was fixed at all loci (**Figure 8**). Therefore, it would seem to appear both species are not necessarily adapted to some aspect of the vegetation in their environments, but that salt marsh habitat largely dominates the vast majority of the *F. heteroclitus* range while mangrove swamp habitat dominates a majority of the *F. grandis* habitat, thus the presence of alleles of either species in their ‘home habitat’ outside of the hybrid zone are simply correlated and provides little evidence that vegetation type is a contributing factor to genetic differentiation among populations.

Observationally, more evidence against the hypothesis that vegetation is a major factor in shaping genetic distributions within the area of overlap comes from two sites, BB and HB. BB is a site that features a mean hybrid index of 0.41 (mean slight abundance of *F. heteroclitus* alleles) and contains almost no marshgrass (0.01) in the habitat, while having the highest proportion of mangrove area out of any site studied (~0.57). Comparatively, HB is a site that has a hybrid index of 0.81 (abundance of *F. grandis* alleles) but contains a much smaller proportion of mangroves (0.34) and a similarly small proportion of marshgrass (0.04). These discrepancies make it hard to predict mean hybrid index of a site from vegetation proportion, thus weakening the link between the two. Since both species and their young have been hypothesized to use mangrove and marshgrass vegetation as places to seek refuge, mate, lay eggs, and feed during high tide, it is possible that these fish are opportunistic habitat-seekers and do not prefer one vegetation type over another, as long as either one is present in some capacity.

A significant correlation ($r = 0.76$) between the presence of *F. grandis* genotypes and maximum salinity was also observed in the first RDA (**Figure 7**), but due to significant issues with multicollinearity, this same parameter was unable to be run in the subsequent RDA with only 9 sites (**Figure 8**). The correlation between maximum salinity and genotypes in the dataset was surprising, given that both *F. heteroclitus* and *F. grandis* have similarly high tolerances of salinity ranges, being able to survive up to 110 ppt and salinities as low as 0.4 ppt (Whitehead 2010; Griffith 1974). While they are able to survive these extremes, it is unclear whether the two species have preferences for salinity levels and whether this affects their distribution. Previous research indicates some subspecies of *F. heteroclitus* have different salinity tolerances when moved to hypoosmotic environments (Whitehead et al. 2011; Scott et al. 2004), and McKenzie et al. (2016) identified one SNP in the *Fundulus heteroclitus heteroclitus*/*F. heteroclitus macrolepidotus* hybrid zone which controlled one aspect of osmoregulation signaling that experienced strong underdominance. Therefore, it may be possible to observe tolerance differences within species that

differ from those in published work (Whitehead 2010; Griffith 1974). Since both *F. heteroclitus* and *F. grandis* have large home ranges but small migration distances, it could be that intraspecific differences in salinity tolerance exist. If this is the case, different tolerances or preferences between *F. heteroclitus* and *F. grandis* in the zone of overlap combined with differences in environmental salinity could play a role in the observed distribution of parental and hybrid types in individuals within the area of overlap in northeastern Florida.

It is also unclear whether higher or lower salinities at sites could contribute to fitness incrementally when combined with other environmental and genetic parameters. This has been shown to occur in other hybrid zones, such as the hybrid zone between *Spartina spp.* in California, where increased salinity combined with flooding decreased the fitness of parental types but led to a hybrid advantage for dispersal (Gallego-Tevar et al. 2020). Another study suggested salinity and wave action to be the main extrinsic factors involved in the distribution of a *Mytilus* hybrid zone, although they were not able to test the effects of these factors directly (Riginos and Cunningham 2005). Since water temperature and dissolved oxygen also experienced slight correlations to either parental species (**Figures 7 & 8**), it is possible some combination of these and salinity contribute to genetic variation in the system. Further testing is required to identify potential relationships among factors and how they relate to fitness within this system, and to investigate additional factors that may play a role in the genetic distribution.

We therefore argue that multiple environmental parameters have an influence on the distribution of parental and hybrid genotypes throughout the zone, and no one factor is especially strong. It is likely that each parameter may influence a relatively small number of loci, but additional testing is necessary to identify these regions and their exact effects on the fitness of individuals within the area of overlap. The correlations between mixed (MX) hybrid individuals and higher dissolved oxygen and colder water temperatures in the first RDA should also be investigated as this could suggest a fitness advantage among hybrids compared to parental types. It is important to note that these analyses simply produced correlational hypotheses, and no conclusions about the true cause of exogenous selection in this study system can be made with this analysis. It is possible other factors, either correlated factors that were pruned out of this analysis due to concerns with multicollinearity or other environmental factors that we did not consider here, may be responsible for the observed genetic patterns within the area of overlap. Lastly, we recognize that the power of our redundancy analyses to detect correlations in our environmental and genetic data could be somewhat limited by the environmental dataset collected, which was itself limited and made accurate interpolation of water quality data at each site within the zone of overlap difficult.

Synopsis

The present study sheds light on the mosaic nature of the spatial genetic structure occurring within the *Fundulus spp.* hybrid zone in northeast Florida and suggests potential explanations for the patterns of selection and the presence of reproductive barriers seen in the region. Evidence of a potential mating preference present in some but not all sites suggests the presence of incomplete and circumstantial barriers to reproduction between the two species. In addition, the fact that F1 hybrids in the zone are fertile to some degree suggests that *F. heteroclitus* and *F. grandis* are still closely related species and that their hybrids do not appear to suffer any mating penalties, although more work is needed to properly examine this idea. As with all mosaic hybrid zones, environmental components of selection affecting hybrid fitness are likely present in this system, although the exact factor most responsible for the majority of genetic variation remains unknown. Future work should focus on identifying the loci under selection in hybrids, investigating the root cause of the inbreeding of *F. grandis* individuals at the northern boundary

of the hybrid zone, and examining potential movement of the hybrid zone northwards as the climate changes along the eastern Atlantic coast.

Index

Supplemental Table 1. GPS coordinates of all included sample sites listed North-South, along with sample sizes for fish collected at each site. The reference site for *F. heteroclitus* was near Saint Mary's River, FL, while the reference site for *F. grandis* was near Cedar Key, FL.

Site Name	Sample Size (n)	Latitude	Longitude
Saint Mary's (SM/H) <i>F. het.</i>	29	30.683080	-81.505489
Matanzas Inlet (MI)	36	29.705250	-81.239839
Marineland (ML)	27	29.669563	-81.219225
Marineland South (MS)	33	29.654725	-81.215950
Summer Haven (SH)	32	29.690904	-81.227672
Bing's Landing (BL)	26	29.614792	-81.206289
Herschel King Park (HK)	37	29.525369	-81.154336
Beverly Beach (BB)	35	29.515250	-81.149081
Flagler North (FN)	35	29.470519	-81.132816
Flagler Beach (FB)	29	29.438763	-81.111988
Highbridge (HB)	35	29.401750	-81.096325
Tomoka Basin (TB)	32	29.365849	-81.094725
Indian River (IR)	14	28.641497	-80.740790
Cedar Key (CK/G) <i>F. gra.</i>	37	29.142931	-83.045214

Supplemental Table 2. List of i5 and i7 Illumina indices used in creation of library during ddRADseq preparation.

i5 Index Name	Index Sequence	i7 Index Name	Index Sequence
A1	CGATGT	PCR2_N701_TAAGGCGA	TAAGGCGA
A2	TTAGGC	PCR2_N702_CGTACTAG	CGTACTAG
A3	TGACCA	PCR2_N703_AGGCAGAA	AGGCAGAA
A4	ACAGTG	PCR2_N704_TCCTGAGC	TCCTGAGC
A5	GCCAAT	PCR2_N705_GGACTCCT	GGACTCCT
A6	CAGATC	PCR2_N706_TAGGCATG	TAGGCATG
A7	ACTTGA	PCR2_N707_CTCTCTAC	CTCTCTAC
A8	GATCAG	PCR2_N708_CAGAGAGG	CAGAGAGG
A9	TAGCTT	PCR2_N709_GCTACGCT	GCTACGCT
A10	GGCTAC	PCR2_N710_CGAGGCTG	CGAGGCTG
A11	CTTGTA	PCR2_N711_AAGAGGCA	AAGAGGCA
A12	AGTCAA	PCR2_N712_GTAGAGGA	GTAGAGGA
A13	AGTTCC	PCR2_N714_GTTCATGA	GTTCATGA
A14	ATGTCA	PCR2_N715_ATCTCAGG	ATCTCAGG
A15	CCGTCC	PCR2_N716_ACTCGCTA	ACTCGCTA
A16	GTGAAA	PCR2_N718_GGAGCTAC	GGAGCTAC
		PCR2_N719_GCGTAGTA	GCGTAGTA
		PCR2_N720_CGGAGCCT	CGGAGCCT

Supplemental Table 3. List of all ADMIXTURE ‘K’ values used in the cross-validation test and their respective cross-validation error.

‘K’ value	Cross-Validation (CV) error
2	0.17696
3	0.17855
4	0.17903
5	0.17935
6	0.36759
7	0.18140
8	0.18483
9	0.55620

Supplemental Table 4. The 95% confidence intervals of pairwise F_{ST} comparisons for each population pairing. In the table, ‘G’ represents the *F. grandis* reference population, while the ‘H’ represents the *F. heteroclitus* reference population. Upper limits are shown above the diagonal, while lower limits are shown below the diagonal.

Population	BB	BL	G	FB	FN	H	HB	HK	IR	MI	ML	MS	SH	TB
BB	NA	0.020435	0.594996	0.243749	0.04139	0.381211	0.290007	0.308557	0.472711	0.398377	0.086228	0.010798	0.379961	0.411774
BL	0.011707	NA	0.760953	0.415618	-0.01999	0.259048	0.461063	0.487346	0.648278	0.269456	-0.01572	-0.0319	0.251692	0.587896
G	0.580599	0.753653	NA	0.25128	0.752396	0.999648	0.194669	0.205274	0.09909	0.995113	0.827454	0.718724	0.994794	0.123642
FB	0.222413	0.396079	0.233934	NA	0.433126	0.789238	-0.00755	0.003611	0.142573	0.80105	0.515373	0.38918	0.786363	0.047333
FN	0.029065	-0.02371	0.737446	0.412959	NA	0.244923	0.477066	0.501814	0.656962	0.253907	-0.003	-0.01493	0.237996	0.596015
H	0.368405	0.248352	0.998582	0.775284	0.226964	NA	0.800295	0.832487	0.998071	0.017548	0.169793	0.256773	0.059259	0.896509
HB	0.267419	0.442006	0.178736	-0.01205	0.456865	0.788274	NA	-0.00129	0.106121	0.809819	0.555788	0.435953	0.796877	0.023534
HK	0.286826	0.460436	0.179315	-0.00473	0.475368	0.816458	-0.00849	NA	0.105967	0.839848	0.581018	0.455345	0.829435	0.022059
IR	0.455745	0.636909	0.056957	0.127801	0.635878	0.995645	0.093164	0.088713	NA	0.991556	0.739157	0.611455	0.990005	0.047924
MI	0.382966	0.25781	0.992012	0.785493	0.233451	0.004777	0.797306	0.822843	0.986396	NA	0.175372	0.268313	0.010995	0.900259
ML	0.073396	-0.021	0.821442	0.496512	-0.01193	0.161049	0.536067	0.558821	0.729487	0.164225	NA	-0.00566	0.159231	0.675881
MS	0.003026	-0.03335	0.711293	0.371217	-0.01904	0.247116	0.414874	0.434801	0.599821	0.256199	-0.00952	NA	0.250551	0.553457
SH	0.360212	0.236256	0.990532	0.770967	0.216324	0.010641	0.783024	0.812026	0.982816	-0.00079	0.147043	0.235834	NA	0.893833
TB	0.390633	0.569961	0.101974	0.03165	0.575106	0.884611	0.013804	0.00883	0.034125	0.888002	0.657254	0.538696	0.88039	NA

References

Agapow, P.M. and A. Burt. 2001. Indices of Multilocus Linkage Disequilibrium. *Molecular Ecology*

Notes 1: 101-102.

Ahrens, C.W., P.D. Rymer, A. Stow, J. Bragg, S. Dillon, K.D. Umbers, and R.Y. Dudaniec. 2018. The

Search for Loci Under Selection: Trends, Biases, and Progress. *Molecular Ecology* 27: 1342

-1356.

- Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.
- Andrews, K.R., J.M. Good, M.R. Miller, G. Luikart and P.A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Review Genetics* 17: 81-92.
- Arnold, M.L. and N.H. Martin. 2010. Hybrid Fitness Across Time and Habitats. *Trends in Ecology and Evolution* 25: 530-536.
- Arnold, M.L. and S.A. Hodges. 1995. Are Natural Hybrids Fit or Unfit Relative to Their Parents? *Trends in Ecology and Evolution* 10: 67-71.
- Arnold, M.L. 1997. *Natural Hybridization and Evolution*. Oxford University Press, New York, NY.
- Arntzen, J.W., B. Wielstra, and G.P. Wallis. 2014. The modality of nine *Triturus* newt hybrid zones assessed with nuclear, mitochondrial, and morphological data. *Biological Journal of the Linnean Society* 113: 604-622.
- Arntzen, J.W. and G.P. Wallis. 1991. Restricted Gene Flow in a Moving Hybrid Zone of the Newts *Triturus Cristatus* and *T. Marmoratus* in Western France. *Society for the Study of Evolution* 45: 805-826.
- Bailey, R.I. 2022. ribailey/gg hybrid: gg hybrid R package for Bayesian hybrid index and genomic cline estimation (v2.0.0). Zenodo.
- Barbas, R.E. and M.R. Gilg. 2018. Quantification of reproductive isolating barriers between two naturally hybridizing killifish species. *Evolutionary Biology* 45: 425-436.
- Barton, N.H. and G.M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* 16: 113-148.
- Bell, R.C., and C.G. Irian. 2019. Phenotypic and genetic divergence in reed frogs across a mosaic hybrid zone on São Tomé Island. *Biological Journal of the Linnean Society* 128: 672-680.
- Bert, T.M. and W.S. Arnold. 1995. An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution* 49: 276-289.
- Bigelow, R. S. 1965. Hybrid zones and reproductive isolation. *Evolution*, 19: 449-458.

- Bossu, C.M. and T.J. Near. 2013. Characterization of a contemporaneous hybrid zone between two darter species (*Etheostoma bison* and *E. caeruleum*) in the Buffalo River System. *Genetica* 141: 75-88.
- Bourret, V., M. Dionne, and L. Bernatchez. 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic Salmon: polygenic multilocus analysis surpass genome scan. *Molecular Ecology* 23: 4444-4457.
- Brown, A. H. D., M.W. Feldman, and E. Nevo. 1980. Multilocus structure of natural populations of *Hordeum Spontaneum**. *Genetics* 96: 523-536.
- Brown, C.A., C.T. Gothreaux, C.C. Green. 2011. Effects of temperature and salinity during incubation on hatching and yolk utilization of Gulf Killifish *Fundulus grandis* embryos. *Aquaculture* 315: 335-339.
- Buggs, R.J.A. 2007. Empirical study of hybrid zone movement. *Heredity* 99: 301-312.
- Campbell, D.R. and N.M. Waser. 2007. Evolutionary dynamics of an *Ipomopsis* hybrid zone: confronting models with lifetime fitness data. *The American Naturalist* 169: 298-310.
- Carl, G., and I. Kuhn. 2007. Analyzing spatial autocorrelation in species distributions using Gaussian and Logit models. *Ecological Modeling* 207: 159-170.
- Cassens, I, R. Tiedmann, F. Suchentrunk, and GB Hartl. 2000. Mitochondrial DNA variation in the European Otter (*Lutra lutra*) and the use of spatial autocorrelation analysis in conservation. *Journal of Heredity* 91: 31-35.
- Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124-3140.
- Davey, John W. and M.L. Blaxter. 2010. RADSeq: next-generation population genetics. *Briefings in Functional Genomics* 9: 416-423.
- Dobzhansky, T.H. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21: 113-135.
- Dobzhansky, T.H. 1940. Speciation as a stage in evolutionary divergence. *American Naturalist* 74(753): 404-420.

- Dormann, C.F., J.Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. Garcia Marquez, B. Gruber, B. Lafourcade, P. J. Leitao, T. Munkemuller, C. McClean, P.E. Osborne, B. Reineking, B. Shroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27-46.
- Duvernell, D., J.F. Schaefer, D.C. Hancks, and J.A. Fonoti. 2007. Hybridization and reproductive isolation among syntopic populations of the topminnows *Fundulus notatus* and *F. olivaceus*. *Journal of Evolutionary Biology* 20: 152-164.
- Ellis, W.L., and S.S. Bell. 2004. Conditional use of mangrove habitats by fishes: depth as a cue to avoid predators. *Estuaries* 27: 966-976.
- Gallego-Tevar, B., B.J. Grewell, C.J. Futrell, R.E. Drenovsky, J.M. Castillo. 2020. Interactive effects of salinity and inundation on native *Spartina foliosa*, invasive *S. densiflora* and their hybrid from San Francisco Estuary, California. *Annals of Botany* 125: 377-389.
- Galleher, S. N., I. Gonzalez, M. R. Gilg, and K. J. Smith. 2009. Abundance and distribution of larval and juvenile *Fundulus heteroclitus* in northeast Florida marshes. *Southeastern Naturalist* 8: 495-502.
- Galleher, S.N., M.R. Gilg, and K.J. Smith. 2010. Comparison of larval thermal maxima between *Fundulus heteroclitus* and *F. grandis*. *Fish Physiology and Biochemistry* 36: 731-740.
- Gilg, M.R., E.V. Kerns, N.E. Gutierrez-Bayona, C. Kooyomjian, and N.A. Hinojosa. 2022. Dynamic cohort analysis reveals fluctuating patterns of selection within a hybrid zone between the killifish *Fundulus heteroclitus* and *F. grandis*. *Evolutionary Biology* 49: 1-14.
- Gonzalez, I., M. Levin, S. Jermanus, B. Watson, M.R. Gilg. 2009. Genetic assessment of species ranges in *Fundulus heteroclitus* and *F. grandis* in northeastern Florida salt marshes. *Southeastern Naturalist*: 8: 227-243.
- Gonzalez, R.J., C.H. Mason, and W.A. Dunson. 1989. Anomalous tolerance to low pH in the estuarine killifish *Fundulus heteroclitus*. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology* 94: 169-172.

- Goudet, Jérôme. 2005. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5: 184-186.
- Griffith, R.W. 1974. Environment and salinity tolerance in the genus *Fundulus*. *American Society of Ichthyologists and Herpetologists (ASIH)* 1974: 319-331.
- Hardy, J. D. Jr. 1978. Development of fishes of the mid-atlantic bight, Vol. II. *Anguillidae* through *Syngnathidae*. US Fish and Wildlife Service Publication FWS/OBS-78/12.
- Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. *Heredity* 56: 337-349.
- Harrison, R.G. 1993. Hybrid zones and the evolutionary process. *Oxford University Press* 1993.
- Jensen, O.P., C.W. Martin, K.L. Oken, F.J. Fodrie, P.C. Lopez-Duarte, K.W. Able, and B.J. Roberts. 2019. Simultaneous estimation of dispersal and survival of the Gulf Killifish *Fundulus grandis* from a batch-tagging experiment. *Marine Ecology Progress Series* 624: 183-194.
- Joost, S., A. Bonin, M.W. Bruford, L. Despres, C. Conord, G. Erhardt, and P. Taberlet. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* 18: 3955-3969.
- Kangas, P.C. and A.E. Lugo. 1990. The distribution of mangroves and saltmarsh in Florida. *Tropical Ecology* 31: 32-39.
- Kamvar, Z.N., J.F. Tabima, and N.J. Grünwald. 2014. The Irish Potato Famine pathogen *phytophthora infestans* originated in Central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences* 111: 8791 – 8796.
- Kneib, R. T. 1986. The role of *Fundulus heteroclitus* in salt marsh trophic dynamics. *American Zoology* 26: 259-269.
- Lotrich, V.A. 1975. Summer home range and movements of *Fundulus heteroclitus* (Pisces: Cyprinodontidae) in a tidal creek. *Ecology* 56: 191-198.
- MacCallum, C.J., B. Nurnberger, N.H. Barton, and J.M. Szymura. 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52: 227-239.
- Manel, Stephanie, M.K. Schwartz, G. Luikart, and Pierre Taberlet. 2003. Landscape genetics: combining

- landscape ecology and population genetics. *Trends in Ecology and Evolution*. 18: 189-197.
- McKenzie, J.L., R.S. Dhillon, and P.M. Schulte. 2016. Steep, coincident, and concordant clines in mitochondrial and nuclear-encoded genes in a hybrid zone between subspecies of Atlantic killifish, *Fundulus heteroclitus*. *Ecology and Evolution* 6: 5771-5787.
- Montgomery, D.C., and E.A. Peck. 1992. Introduction to Linear Regression Analysis, 2nd edition. Wiley Interscience. ISBN: 9780471533870.
- Moore, W.S. 1977. An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* 52: 263-277.
- Muller, H.J. 1942. Isolating mechanisms, evolution, and temperature. *Biology Symposium* 6: 71-125.
- Mulvey, M., M.C. Newman, W.K. Vogelbein, M.A. Unger, and D.R. Ownby. 2003. Genetic structure and mtDNA diversity of *Fundulus heteroclitus* populations from polycyclic aromatic hydrocarbon-contaminated sites. *Environmental Toxicology and Chemistry* 22: 671-677.
- Nelson, T.R., D. Sutton, and D.R. DeVries. 2014. Summer movements of the Gulf Killifish (*Fundulus grandis*) in a northern Gulf of Mexico salt marsh. *Estuaries and Coasts* 37: 1295-1300.
- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlenn, P. R. Minchin, R. B. O'Hara, G.L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. Wagner. 2020. *vegan: Community Ecology R Package v 2.5-7*.
- Orr, H.A. and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55: 1085-1094.
- Paris, J.R., J.R. Stevens, and J.M. Catchen. 2017. Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution* 8: 1360-1373.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7: e37135.
- Piertney, SB, AD MacColl, PJ Bacon, and JF Dallas. 1998. Local genetic structure in Red Grouse

- (*Lagopus lagopus scoticus*): evidence from microsattelite DNA markers. *Molecular Ecology* 7: 1645-1654.
- Powell, D.L. 2019. The causes and consequences of hybridization in two northern swordtail fishes. Office of Graduate and Professional Studies of Texas A&M University Dissertation.
- Rand, D.M. and R.G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43: 432-449.
- Rellstab, C., F. Gugerli, A.J. Eckert, A.M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* 24: 4348-4370.
- Revelle, W. 2021. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.9.
- Riginos, C. and C.W. Cunningham. 2005. Local adaptation and species segregation in two mussel (*Mytilus edulis* x *Mytilus trossulus*) hybrid zones. *Molecular Ecology* 14: 381-400.
- Rochette, N.C., A.G. Rivera-Colon, and J. M. Catchen. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* 28: 4737-4754.
- Rozas, L.P., and D.J. Reed. 1993. Nekton use of marsh-surface habitats in Louisiana (USA) deltaic salt marshes undergoing submergence. *Marine Ecology Progress Series* 96: 147-157.
- Rozas, L.P. and M.W. LaSalle. 1990. A comparison of the diets of Gulf Killifish, *Fundulus grandis* Baird and Girard, entering and leaving a Mississippi brackish marsh. *Estuaries* 13: 332-336.
- Sambrook, J. and D.W. Russell. 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Scott, G.R., J.T. Rogers, J.G. Richards, C.M. Wood, P.M. Schulte. 2004. Intraspecific divergence of ionoregulatory physiology in the euryhaline teleost *Fundulus heteroclitus*: possible mechanisms of freshwater adaptation. *Journal of Experimental Biology* 207: 3399-3410.
- Schaefer, J., D. Duvernell, and D.C. Campbell. 2016. Hybridization and introgression in two ecologically dissimilar *Fundulus* hybrid zones. *Evolution* 70: 959-1163.
- Segelbacher G., J. Tomiuk, and S. Manel. 2008. Temporal and spatial analyses disclose consequences

- of habitat fragmentation on the genetic diversity in *Capercaillie (Tetraourogallus)*. *Molecular Ecology* 17: 2356–2367.
- Sinama, M., A. Gilles, C. Costedoat, E. Corse, J.M. Oliver, R. Chappaz, and N. Pech. 2013. Non homogenous combination of two porous genomes induces complex body shape trajectories in cyprinid hybrids. *Frontiers in Zoology* 10: 1-16.
- Sweeney, J. L. Deegan, and R. Garritt. 1998. Population size and site fidelity of *Fundulus heteroclitus* in a macrotidal saltmarsh creek. *The Biological Bulletin* 195: 238 – 239.
- Swenson, N.G. 2008. The past and future influence of Geographic Information Systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology* 21: 421-434.
- Taylor, S.A. and E.L. Larson. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution* 3: 170-177.
- Taylor, S.A., T.A. White, W.M. Hochachka, V. Ferretti, R.L. Curry, and I. Lovette. 2014. Climate mediated movement of an avian hybrid zone. *Current Biology* 24: 671-676.
- Whitehead, A. 2010. The evolutionary radiation of diverse osmotolerant physiologies in Killifish (*Fundulus SP.*). *Society for the Study of Evolution* 64: 2070-2085.