Author for correspondence:
Jessica L. McKenzie
e-mail: mckenzie@zoology.ubc.ca

†Present address: Department of Biomolecular Chemistry, Epigenetics Theme, Wisconsin Institute for Discovery, University of Wisconsin, Madison, WI, USA.

THE ROYAL SOCIETY PUBLISHING

Evidence for a bimodal distribution of hybrid indices in a hybrid zone with high admixture

Jessica L. McKenzie[1,2], Rashpal S. Dhillon[1,†]
and Patricia M. Schulte[1]

[1]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada
[2]Centre for Aquaculture and Environmental Research, Fisheries and Oceans Canada, West Vancouver, British Columbia, Canada

The genetic structure of a hybrid zone can provide insights into the relative roles of the various factors that maintain the zone. Here, we use a multilocus approach to characterize a hybrid zone between two subspecies of killifish (*Fundulus heteroclitus*, Walbaum 1792) found along the Atlantic coast of North America. We first analysed clinal variation along the Atlantic coast using a single-nucleotide polymorphism in the mitochondrial DNA (mtDNA) displacement loop (D-loop) and a panel of nine nuclear microsatellite markers. A model constraining all clines to the same width and centre was not significantly different from a model in which the clines were allowed to vary independently. Locus-by-locus analysis indicated that the majority of nuclear clines shared the same centre as the mtDNA cline, and the widths of these clines were also narrower than that predicted by a neutral model, suggesting that selection is operating to maintain the hybrid zone. However, two of the nuclear clines had widths greater than the neutral prediction and had centres that were displaced relative to the mtDNA cline centre. We also found that a marsh located near the centre of the mtDNA cline demonstrated a bimodal distribution of nuclear hybrid index values, suggesting a deficit of first-generation hybrids and backcrossed genotypes. Thus, selection against hybrid genotypes may be playing a role in maintaining this hybrid zone and the associated steep nuclear and mtDNA clines.

1. Introduction

When two recently diverged taxa meet and reproductive isolation is not complete, hybridization can occur, which causes a transition between the morphological, physiological

and ultimately genetic traits of the two taxa as one crosses the hybrid zone [1,2]. Hybrid zones, and the resulting phenotypic and genetic clines that they create [1,3,4], provide natural laboratories for the study of speciation and natural selection [2] and can provide insight into the genes underlying morphological, physiological and reproductive differences among taxa (e.g. [5]).

Patterns of genetic variation within a hybrid zone can help illuminate the factors that maintain the zone [6–8]. For example, analysis of the shape of clines can be used to infer the strength of selection on various regions of the genome [9], because steep clines with cline widths narrower than predicted by a neutral model suggest that the differentiated alleles at these loci (or closely linked loci) are being maintained by selection [10,11]. Another approach to the analysis of hybrid zones involves explicit consideration of multilocus genetic variation through the calculation of the hybrid index, which is an estimate of the similarity of each hybrid individual to the parental types. By examining the frequency distribution of hybrid indices, it is possible to assess the extent of reproductive isolation operating within the zone [7]. If there is limited reproductive isolation between the hybridizing taxa, a unimodal pattern of hybrid indices will be observed at the centre of the zone, indicative of a hybrid swarm, whereas if there is substantial reproductive isolation between the taxa, a bimodal pattern of hybrid indices will be observed [7]. Observations of Hardy–Weinberg equilibrium (HWE) and linkage equilibrium at multiple loci are typical in unimodal zones, whereas the presence of heterozygote deficit and linkage disequilibrium (and cytonuclear disequilibrium) suggests that reproductive isolation and/or local adaptation of either parental type to divergent habitats are acting to reduce the success of hybrid offspring [7].

It is important to recognize that within a given hybrid zone different loci may show differing degrees of bimodality, depending on the strength of selection acting on a particular locus [7]. Similarly, the locations of the cline centres and steepness of clines may vary among loci, with some markers yielding steep, abrupt clines and others more gradual cline shapes, depending on the involvement or linkage of these markers to traits involved in reproductive isolation [12,13]. Thus, determining the pattern of genetic variation across and within a hybrid zone is an important step in identifying the factors that are responsible for the maintenance of the zone.

Here, we examine the patterns of genetic variation in a hybrid zone between two subspecies of Atlantic killifish, *Fundulus heteroclitus*. Morphological, physiological and genetic clines in *F. heteroclitus* have been extensively studied, and this species is cited as a 'textbook example' of intraspecific latitudinal genetic variation (e.g. [14,15]). Previous biochemical genetic studies in killifish have identified clines along the Atlantic coast at various enzyme-coding loci, with northern alleles being replaced by southern alleles along the New Jersey shoreline [16,17]. Subsequent work at the molecular level has detected clines in mitochondrial DNA (mtDNA) [18,19], putatively neutral nuclear microsatellite loci [20,21] and a variety of nuclear single-nucleotide polymorphisms (SNPs) [22]. Most of these genetic and character clines show breaks from one predominant type to the other between 40° N and 41° N, which has been used as evidence to suggest that *F. heteroclitus* consists of two subspecies: *Fundulus heteroclitus heteroclitus* Linnaeus 1766 (hereafter referred to as 'the southern subspecies') ranging from New Jersey south to northern Florida and *Fundulus heteroclitus macrolepidotus* Walbaum 1792 (hereafter referred to as 'the northern subspecies') ranging from Connecticut north to Newfoundland [23], with a hybrid zone located along the coast of New Jersey connecting the two subspecies.

For example, there has been thorough sampling of the area comprising the suspected transition zone and analysis using restriction fragment site polymorphisms (RFLPs) in the mitochondrial genome [19], and these authors concluded that the zone of intergradation occurred in a 60 km stretch between Barnegat Light, New Jersey (39.77° N, 74.12° W) and Great Egg, New Jersey (39.30° N, 74.63° W). Interestingly, in a more recent study employing nuclear markers [20], data from eight microsatellite loci were used to estimate q-bar (STRUCTURE) [24,25] leading to the conclusion that the centre of the hybrid zone was located approximately at Newark Bay, New Jersey (40.69° N, 74.11° W), roughly 130 km north of the most northern limit described for the mitochondrial inflection point. Despite the extensive work that has been conducted in this system, no previous study has simultaneously examined patterns of variation in both the nuclear and mitochondrial genomes in the putative zone of contact, nor have multilocus approaches been applied.

Here, we aimed to increase the sampling intensity in the putative hybrid zone, focusing on sampling from marshes located close to the mitochondrial cline centre, as the mtDNA cline is the steepest that has yet been detected in this species [22]. We used a combination of mtDNA and nine microsatellite markers to examine populations of killifish from marshes located within and immediately surrounding this region, conducting clinal analyses and assessing coincidence and concordance of these clines. We then used the microsatellite data to calculate multilocus hybrid indices for all individuals sampled to

determine the pattern of the genetic variation within the hybrid zone with respect to the unimodal–bimodal continuum, to provide additional insight into the nature of this hybrid zone.

2. Material and methods

2.1. Fish collection

Fundulus heteroclitus (520 individuals) were collected from various locations along the Atlantic coast of North America, with sample collection concentrated in and immediately adjacent to the presumed hybrid zone between the two subspecies in New Jersey (table 1 and figure 1). Samples from the contact zone were collected from May to June of 2008 from 11 locations along the coast (at the mouth of each marsh). Minnow traps (G-type) were used and trap set time ranged from 2 to 6 h. Fish length was recorded and a fin clip was taken from each individual and preserved in 95% ethanol. Additional samples representative of the pure populations of the two subspecies from previous studies [26] were collected near Brunswick, GA and Hampton, NH (table 1).

2.2. Mitochondrial D-loop PCR and SNP genotyping

DNA was extracted from all samples using Qiagen's DNeasy® Blood and Tissue Kit and all polymerase chain reactions were performed using an MJ Research PTC-200 Peltier Thermal Cycler. A restriction digestion assay was developed to allow the rapid diagnosis of mitochondrial-type based on a single fixed difference found during a preliminary screening of northern ($n = 20$) and southern ($n = 20$) mtDNA D-loop sequences. Preliminary screening was performed using primers K (forward) and G (reverse) [27] to amplify approximately 1100 base pairs (bp) of the mtDNA D-loop and reactions were prepared in 12.5 µl volumes containing 10× Taq Buffer with KCl (Fermentas), 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM each primer, 1 unit Taq DNA polymerase (Fermentas), the appropriate volume of ddH_2O and 1.0 µl genomic DNA of varying concentrations. The amplification profile consisted of an initial denaturing step of 2 min at 95°C followed by five cycles of denaturation at 94.0°C for 30 s, primer annealing at 60°C for 45 s, extension at 72°C for 1 min 30 s, and 34 cycles of denaturation at 94°C for 30 s, primer annealing at 65°C for 45 s, extension at 72°C for 1 min 30 s. The final extension was at 72°C for 5 min. The resulting PCR products were then electrophoresed on a 1% agarose gel pre-stained with ethidium bromide and visualized using a GeneGenius Bio Imaging System (Syngene). Successful amplifications were purified using Qiagen's QIAquick® PCR Purification Kit and direct sequenced using the BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Corporation, Foster City, CA, USA). This reaction contained 3.0 µl BigDye® dilution mix, 2.0 µl purified PCR product, 1.0 µM of K primer and ddH_2O for a final volume of 10 µl. The cycle sequencing profile consisted of an initial denaturing step of 96°C for 1 min followed by 26 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. The sequences were edited and aligned using the program MEGA 4.0.2 [28]. The online program NEBCUTTER v.2.0 [29] was then used to identify a restriction enzyme that would cut at one of the several diagnostic sites that exhibited fixed differences between northern and southern fish.

The restriction digestion assay was performed by PCR, amplifying the mtDNA D-loop as above. 8.0 µl reactions containing 5 µl of PCR product, 10× Buffer ScaI (Fermentas), 0.25 units of the restriction enzyme *Sca*I (Fermentas) and the appropriate volume of ddH_2O were incubated for 3 h at 37°C. The undigested 1100 bp fragment represented individuals with a northern mitochondrial haplotype while the digested 900 bp fragment was diagnostic of the southern mitotype. This PCR-RFLP assay was then applied to all 520 fish collected from 13 locations (table 1). The fragment polymorphisms generated by the restriction digest were easily discernible on a 1.5% agarose gel pre-stained with ethidium bromide. Ten northern, nine southern and 28 hybrid zone fish were then sequenced at random to confirm the results of the RFLP assay. The resulting sequences have been submitted to GenBank (Accession Numbers JQ518219–JQ518265).

2.3. Microsatellite genotyping

Individuals from locations 1 to 13 inclusive were also genotyped at nine microsatellite loci [30]: FhCA-1, FhATGB101, FhATGB128, FhATG-2, FhATG-4, FhATG-6, FhATG-17, FhATG-18 and FhATG-20. PCR conditions for each of these microsatellite markers required modification from those described by Adams *et al.* [30] for successful amplification (electronic supplementary material, table S1). Reactions involving a single round of denaturation, annealing and extension were performed for 35 cycles. For any reaction

Table 1. Sampling location names, coordinates and sample sizes (*n*) along the Atlantic coast of North America.

| location name	distance from Georgia (km)	latitude (°N)	longitude (°W)	*n*
1. Brunswick, GA	0	31.116130	−81.462250	19
2. Rutgers University Marine Field Station (RUMFS), NJ	1145.77	39.512905	−74.320450	50
3. Laurel Harbour, NJ	1186.03	39.750349	−74.192734	50
4. Metedeconk Creek, NJ	1208.94	40.065616	−74.065320	49
5. Navesink River, NJ	1228.41	40.376297	−73.993993	13
6. Sandy Hook Bay, NJ	1231.88	40.412904	−73.969487	50
7. Belford Creek, NJ	1242.06	40.440968	−74.103963	49
8. Cheesequake Creek, NJ	1257.54	40.463420	−74.258888	50
9. Wantagh, NY	1285.31	40.638689	−73.563080	39
10. Housatonic River, CT	1357.23	41.173856	−73.103081	22
11. Roger's Creek, CT	1363.41	41.213379	−73.057564	12
12. Hammonassett Creek, CT	1393.09	41.251971	−72.537634	48
13. Hampton, NH	1631.00	42.937055	−70.839844	20

involving two rounds of denaturation, annealing and extension, the first round proceeded for five cycles, followed by a second round of 30 cycles. All final extensions were performed at 72° for 5 min. FhATG-17, FhATG-18 and FhATG-20 were combined in a single multiplex reaction, as were FhATG-2 and FhATG-4. Although FhATGB101 and FhATGB-128 used identical reaction conditions as FhATG-2 and FhATG-4, they were combined in a separate multiplex reaction to avoid complications as a result of size-fragment overlap. Final reaction volumes were 10 μl with reagent concentrations as follows: 10× Taq Buffer with KCl (Fermentas), 2.0 mM MgCl$_2$, 0.2 mM dNTPs, 0.4 μM LDH-F1, 0.4 μM LDH-R2, 0.5 units Taq DNA polymerase (Fermentas), 1.0 μl genomic DNA and ddH$_2$O for a final reaction volume of 10 μl. All reactions were performed with these concentrations except for the multiplex reaction involving FhATG-2 and FhATG-4 which required 1 unit of Taq polymerase. PCR products were then dried down in 96-well plates and sent to the Boston Children's Hospital Molecular Genetics Core Facility for genotyping on an Applied Biosystems 3730 DNA Analyzer. Alleles were scored using PEAK SCANNER™ software v. 1.0 (Applied Biosystems). Multilocus microsatellite genotypes for all individuals used in this study can be found in the electronic supplementary material, table S2.

2.4. Analyses

MICROCHECKER v. 2.2.3 [31] was used to screen the microsatellite data for the presence of null alleles, large allele dropout and scoring errors. The program GENEPOP v. 4.0.10 [32] was used to conduct the exact test for Hardy–Weinberg equilibrium. Tests for heterozygote excess and heterozygote deficiency were performed for each locus in each population. Markov chain parameters were set to the defaults (dememorization number = 1000; 100 batches; 1000 iterations per batch). GENETIX v. 4.05.2 [33] was used to test for linkage disequilibrium among the different microsatellite loci. Each locus pair in each population was tested for evidence of linkage disequilibrium using the Black and Krafsur procedure [34]. The program CNDM (available at http://statgen.ncsu.edu/cnd/CNDd.php [35,36]) was used to test for cytonuclear disequilibrium between mitochondrial-type and each microsatellite locus within marshes 2–8 (table 1).

2.5. Clinal analyses

The program HZAR (hybrid zone analysis for R) [37], implemented in the R programming environment [38], was used to fit all clines and perform all clinal analyses. First, the best of three models (as determined through lowest AIC value) was fit to the allele frequency data and used to predict cline parameters (centre, width, p_{min}, p_{max}). In addition to the null model, the models used for the cline fitting of the molecular markers were: model I: fixed p_{min}/p_{max} values and tails not fitted, model II:
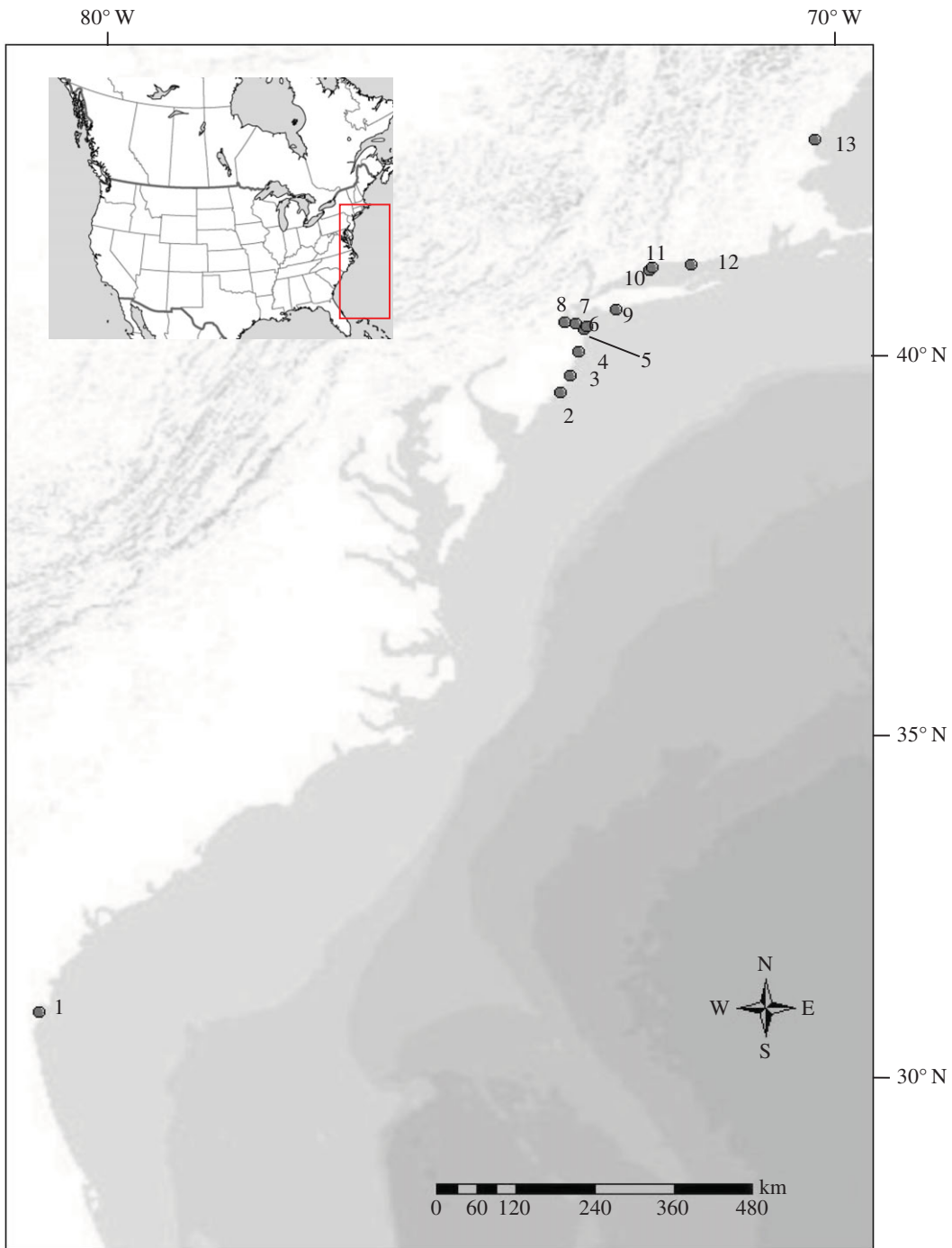
Figure 1. Map of collection locations. Numbers correspond to those listed in table 1. Inset of North America indicates location of sampling area.

p_{min}/p_{max} estimated and tails not fitted, and model III: p_{min}/p_{max} estimated and tails fitted. More detailed descriptions of these models can be found in Brumfield *et al.* [39]. Once the best model had been selected, it was used to plot the best fit cline for the observed allele frequencies as well as the 2-unit support envelope around the cline. In order to compare the estimate for cline widths to that which would be expected under a neutral model of selection (w), the following equation [40], was solved for w:

$$T = \left(\frac{w}{(\sqrt{2\pi})\sigma} \right)^2,$$

(2.1)

where T represents the number of generations since contact under the assumption of a neutral cline, and σ represents dispersal per generation.

Clines were tested for coincidence and concordance using the program CFIT7 [41]. Briefly, this program uses the scaled logit shape to fit the clines [42]. Two models were compared: one in which mitochondrial and microsatellite clines were constrained to a common centre and width and one in which the cline shapes were allowed to vary independently. The likelihoods of these two models were then compared using the likelihood ratio test [43].

2.6. Hybrid index

The program STRUCTURE 2.3.4 [24,25] was used to first identify the number of admixed populations among the 13 locations sampled, and second to approximate which of the sampled locations best represented the edges of the hybrid zone. The program was run using the default parameters, beginning with a 50 000 iteration burn-in followed by 150 000 Markov chain Monte Carlo iterations. The number of genetic clusters (k) tested ranged from one to 10, with each analysis being replicated 10 times. The number of clusters that best described the genetic admixture in our samples was determined using the method of Evanno *et al.* [44] as implemented in the online program HARVESTER [45]. Values of q were then examined to identify the populations bounding the hybrid zone.

Populations were deemed to represent the edges of the hybrid zone based on a shift in q-values from those indicative of either parental type predominating ($0.10 > q > 0.90$) to an increase in values representing admixed individuals ($0.10 < q < 0.90$). Individuals from these border marshes were then used as parental types to train the program INTROGRESS as implemented in the R programming environment [38,46,47]. This program uses allele frequency differentials (δ) to calculate a hybrid index for each individual based on their multilocus microsatellite genotypes. Hybrid index values for each individual in each marsh were then binned in 0.10 increments and compiled in a histogram representing the microsatellite hybrid indices present in each marsh. The shape of the frequency distribution (either unimodal or non-unimodal) was then determined by Hartigan's diptest [48,49] using the package DIPTEST as implemented in the R programming environment [38].

3. Results

3.1. Hardy–Weinberg Equilibrium and linkage disequilibrium

Among the microsatellites, there was no evidence of null alleles or large allele dropout, but there was some evidence of significant departures from HWE (table 2). Significant heterozygote deficit was observed for at least one locus in each of marshes 2, 3, 6, 11–13 (table 2). Significant heterozygote excess was observed at three loci (FhATG18, FhATG17 and FhCA-1) in marsh 12 and one locus (FhCA-1) among individuals collected from marsh 4. However, only one of these values (heterozygote deficit at locus FhATGB101 in marsh 6) remained significant following an adjustment for false discovery rate (FDR = 0.05) using the classical one-stage method [50].

Among the 13 locations sampled, a total of 61 pairs of microsatellite loci were found to be in significant linkage disequilibrium prior to correction for multiple testing, with 18 pairs of linkage disequilibria remaining significant after FDR-adjustment for multiple comparisons (table 2). No microsatellite loci were found to be in significant disequilibrium with the mitochondrial genome at any location.

4. Characterization of the hybrid zone

4.1. Clinal analyses

Five hundred and twenty individuals from 13 locations were genotyped at one mtDNA SNP and nine microsatellite markers to confirm the previously observed locations of the cline centres for these nuclear and mitochondrial markers [18,20,21,51].

All microsatellite loci and the mtDNA SNP marker exhibited clinal variation along the coast (figure 2; electronic supplementary material, table S3 and figure S1), as has been previously observed [18,20,21,51]. Clines were also fitted independently for each sex and did not differ. The mtDNA cline centre was extremely close to our sampled location no. 4 (Metedeconk Creek). The clines generated by the microsatellite loci FhATGB101 and FhATG18 had their centres within the width of the mitochondrial cline (table 3), while FhATG20, FhATG4, FhATG2 and FhATGB128 had similar cline centres that were located slightly to the north of this range (table 3). Further to the north were the cline centres for the
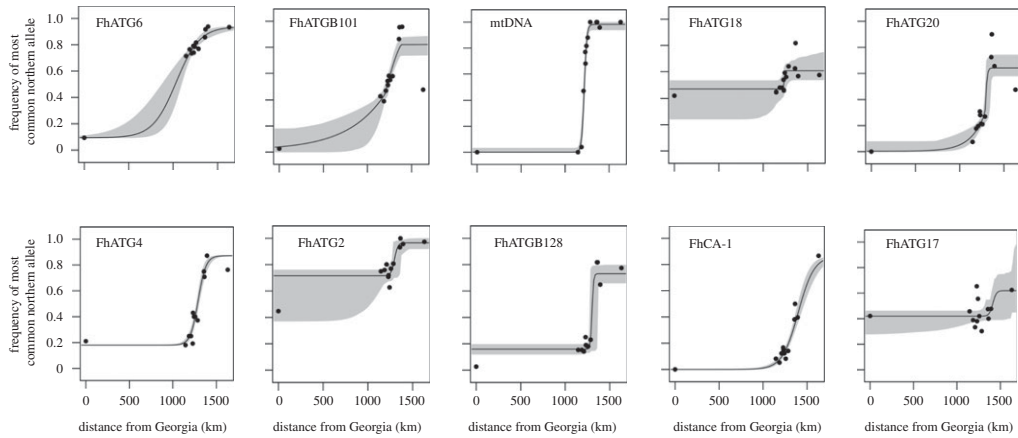
Figure 2. Shape of the cline of nine microsatellite loci and an mtDNA SNP of *F. heteroclitus* populations distributed along the coast of North America as determined using HZAR. Grey shading indicates the 2-unit support envelope around the cline. Distance refers to distance from Georgia as in table 1.

Table 2. Deviations from HWE (F_{IS}) for microsatellite loci among coastal populations of *F. heteroclitus*. Location numbers correspond to those given in table 1. Positive F_{IS} indicates heterozygote deficit while negative F_{IS} indicates heterozygote excess. F_{IS} values significant at $\alpha = 0.05$ in bold. The final column indicates the number of pairs of loci showing significant linkage disequilibrium. The number of loci remaining significant after FDR-adjustment is shown in parentheses.

location	locus name									no. pairs of loci in linkage disequilibrium
	FhATG18	FhATG20	FhATG17	FhATGB101	FhATGB128	FhATG2	FhATG4	FhATG6	FhATG-CA1	
1.	0.081	0.029	0.161	−0.057	−0.186	−0.049	−0.067	−0.041	−0.000	6 (0)
2.	**0.154**	0.030	0.196	0.018	0.000	−0.015	−0.055	−0.018	0.009	7 (2)
3.	0.091	−0.068	0.123	−0.025	−0.086	−0.055	**0.061**	−0.025	−**0.083**	7 (2)
4.	0.041	−0.057	0.123	0.068	0.134	0.038	−0.099	0.044	−**0.112**	5 (2)
5.	−0.206	−0.039	0.189	0.028	0.052	−0.375	0.219	−0.146	−0.143	2 (0)
6.	−0.055	0.019	0.103	**0.127**[a]	0.045	−0.005	0.050	−0.069	−0.050	11 (8)
7.	0.018	0.042	0.201	0.044	−0.019	0.217	−0.047	−0.053	0.007	2 (0)
8.	−**0.168**	0.064	−0.061	−0.055	0.066	0.100	−0.006	−0.079	−0.067	4 (2)
9.	−0.110	0.083	0.070	**0.254**	−0.0041	−0.226	−0.021	0.002	−0.055	7 (2)
10.	−0.176	−0.097	−0.301	0.091	−0.147	−0.050	0.216	−0.086	−0.314	3 (0)
11.	−0.127	−0.029	0.018	n.a.	−0.177	n.a.	0.160	**0.500**	−0.065	2 (0)
12.	−**0.211**	−**0.097**	−**0.235**	−0.025	0.029	**0.490**	−0.090	−0.029	−**0.147**	2 (0)
13.	0.101	−0.162	−0.106	−0.062	**0.360**	n.a.	0.060	−0.018	−0.087	3 (0)

[a]F_{IS} values significant after FDR-adjustment.

loci FhCA-1 and FhATG17, while the cline centre for the marker FhATG6 was substantially south of the mtDNA cline centre (table 3). The likelihood ratio test showed no significant difference between a model where all clines were constrained to the same centre and width (log-likelihood = −4869.10; parameters = 6) and a model where these parameters were allowed to vary independently among loci (log-likelihood = −4645.82; parameters = 5; likelihood ratio test: $\chi^2 = 0.094$, d.f. = 6, $p > 0.05$; electronic supplementary material, table S4 and figure S4).

Assuming that the secondary contact of the northern and southern subspecies of *F. heteroclitus* occurred sometime after the beginning of the most recent glacial retreat, approximately 15 000 generations have passed since the secondary contact of these killifish forms. Using a conservative dispersal estimate of 2 km, as derived from mark–recapture data [52], and solving equation (2.1) for w, a similarly conservative estimate for the width of the neutral cline is 614 km. Of the estimates for

Table 3. Parameters for mitochondrial and microsatellite allele frequency clines (with 2-unit support limits) as estimated with HZAR. Model selection based on lowest AIC value, as indicated in bold for each locus.

| locus name	centre (km)	width (km)	p_{min}	p_{max}	model AIC			
					null	I	II	III
FhATG6	1035.57 (872.60, 1102.30)	525.77 (347.17, 950.62)	0.09 (fixed)	0.94 (fixed)	104.387	**11.734**	15.031	22.600
FhATGB101	1216.19 (1207.30, 1250.95)	377.07 (257.38, 457.69)	0.01 (0.01, 0.10)	0.96 (0.89, 1.0)	193.265	99.29	75.779	**67.566**
mtDNA	1217.47 (1212.52, 1222.23)	57.89 (46.13, 72.11)	0.00 (0.00, 0.02)	0.98 (0.95, 1.00)	366.149	38.951	29.046	29.046
FhATG18	1238.46 (1177.50, 1268.31)	7.23 (1.47, 297.40)	0.47 (0.34, 0.51)	0.61 (0.56, 0.67)	26.395	21.223	**21.727**	24.875
FhATG20	1288.67 (1281.02, 1327.94)	43.59 (8.91, 115.26)	0.003 (0, 0.05)	0.64 (0.59, 0.72)	195.100	68.82	**15.922**	**37.43**
FhATG4	1291.37 (1277.59, 1306.67)	179.23 (139.44, 229.79)	0.18 (fixed)	0.87 (fixed)	197.64	**24.630**	46.03	34.347
FhATG2	1297.11 (1199.43, 1356.68)	73.24 (0.038, 467.43)	0.72 (0.47, 0.75)	0.96 (0.93, 1.0)	92.462	32.921	**32.855**	32.919
FhATGB128	1301.02 (1285.85, 1356.91)	33.98 (0.004, 99.28)	0.16 (0.13, 0.19)	0.73 (0.67, 0.79)	242.700	39.626	**24.446**	28.374
FhCA-1	1398.94 (1371.88, 1434.90)	363.30 (296.29, 459.45)	0 (fixed)	1.0 (fixed)	164.162	**18.286**	21.263	26.942
FhATG17	1413.30 (1295.85, 1699.83)	80.05 (0.33, 1522.26)	0.42 (0.30, 0.45)	0.62 (0.47, 1.00)	30.973	30.254	**29.783**	38.029
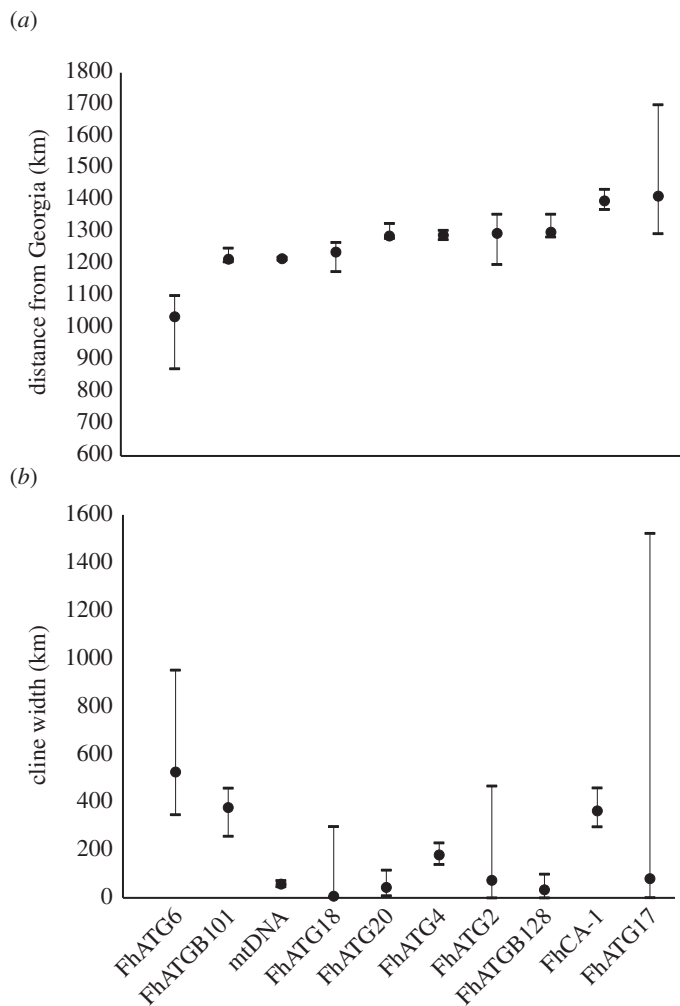
Figure 3. Centre as distance in kilometres from Georgia (*a*) and width (*b*) estimates with 2-unit support limits as estimated using HZAR.
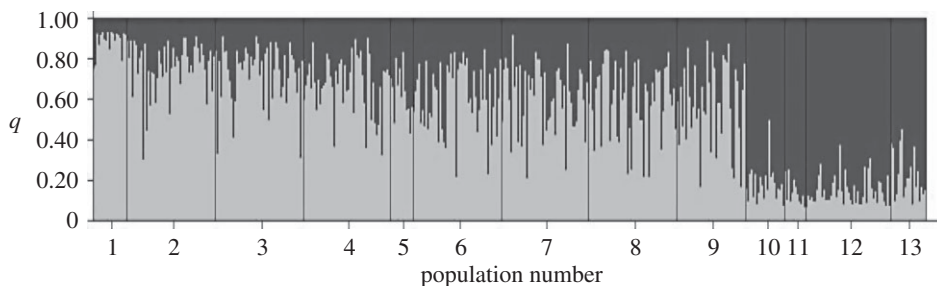
Figure 4. Barplot of admixture values (*q*) from STRUCTURE, *k* = 2. Light grey bars indicate proportion of genome inherited from southern-type parents and black bars indicate proportion of genome inherited from northern-type parents. Population numbering is as in table 1.

cline width in table 3, only two loci (FhATG6 and FhATG17) have cline widths that are not significantly different from the neutral cline estimate, having upper support limits that greatly exceed the neutral estimate of 614 km. It is notable that the remaining clines all have similarly significantly steep clines and quite similar centres (figure 3).

Analysis with STRUCTURE identified the presence of two admixed gene pools as indicated by the elevated Δ*k* calculated using Evanno *et al.*'s method [44] (electronic supplementary material, figure S2). The barplot of *q*-values for *k* = 2 (figure 4) clearly shows a transition from pure southern-type individuals dominating the southern-most marsh to an increased proportion of admixed individuals in marsh 2 (RUMFS). Similarly, northern-type individuals predominate in marshes 10–13, with admixed genotypes becoming more common in marshes 9 (Wantagh) and 8 (Cheesequake).
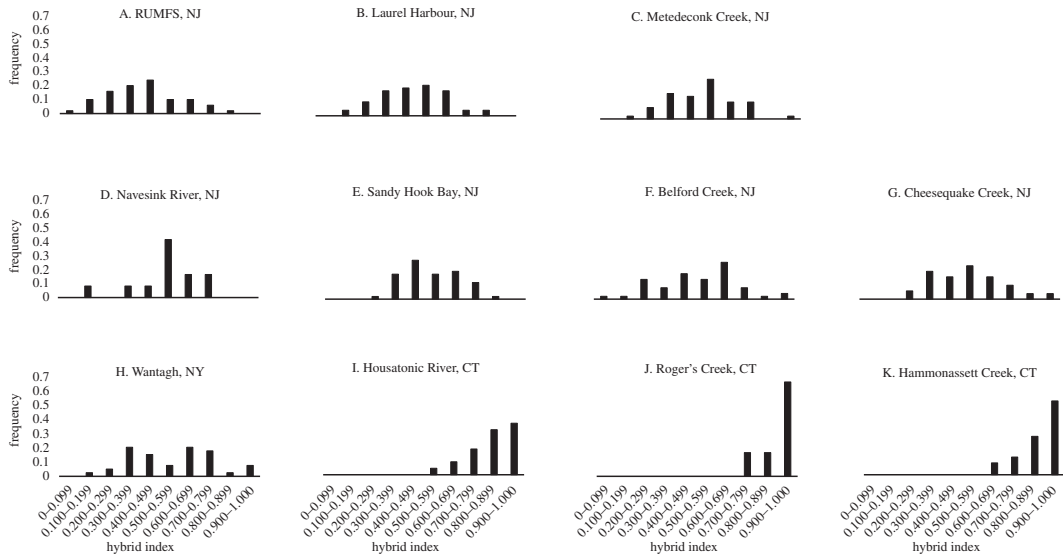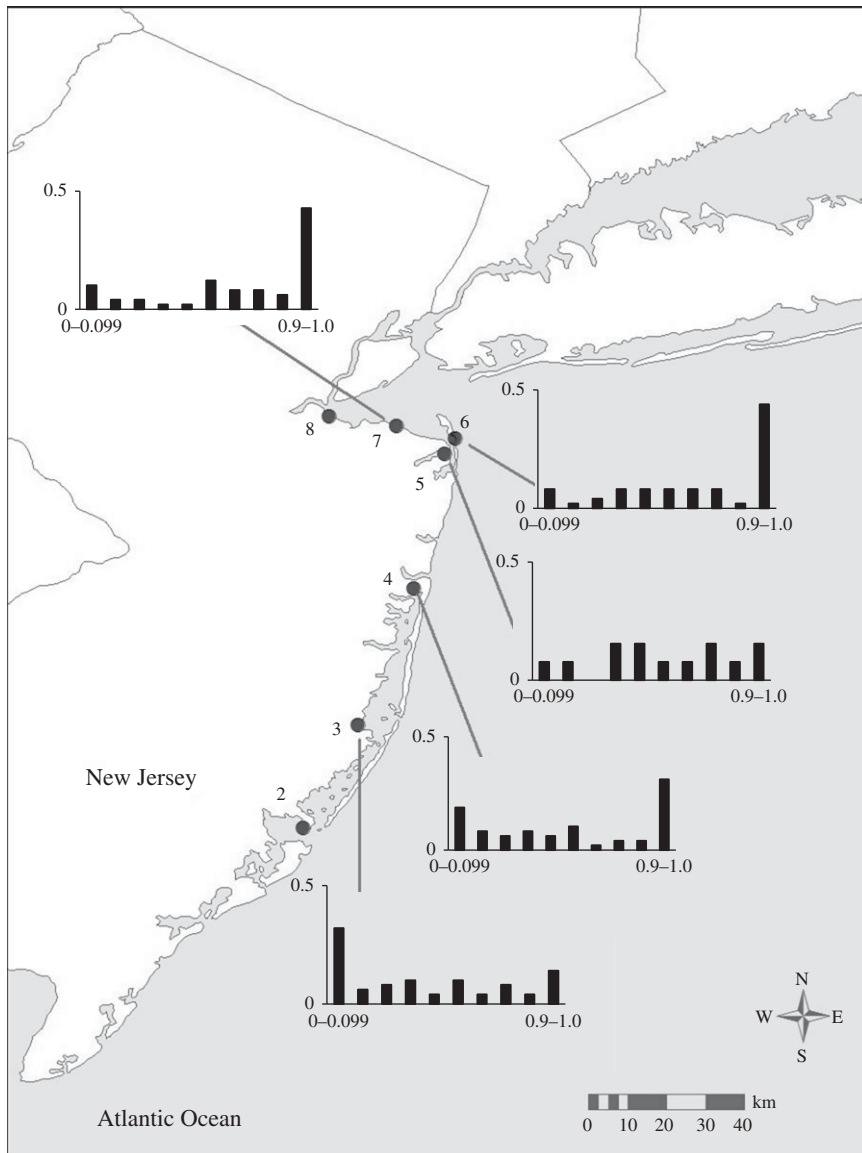
Figure 5. Frequency distributions of hybrid indices as calculated with the program INTROGRESS using individuals from Hampton, NH as northern parental types and those from Brunswick, GA as southern parental types. Individuals with a hybrid index of 1.0 are representative of the northern subspecies while those with a hybrid index value of 0.0 identify with the southern subspecies.
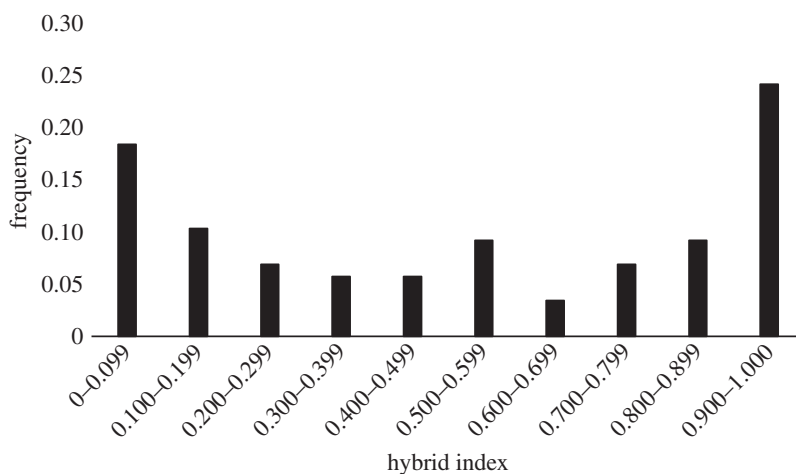
4.2. Hybrid index

Hybrid indices for the multilocus microsatellite data were first calculated for individuals ($n = 432$) from 11 populations using individuals from the northernmost (Hampton, NH) and southern-most (Brunswick, GA) sampling locations as parental populations to train the program INTROGRESS (figure 5). Consistent with the results of the STRUCTURE analysis, all populations within the putative hybrid zone were genotypically intermediate between the two extreme populations. Within the putative hybrid zone, the frequency distributions in hybrid index exhibited unimodal or flat patterns (figure 5*a*–*h*) and were progressively more similar to the extreme northern population at sites north of Long Island (figure 5*i*–*k*).

The individuals collected from the extremes of the species' distribution represent the pure forms of the northern and southern subspecies of *F. heteroclitus*, but it is unlikely that fish from these locations represent the true parental genotypes of individuals within the hybrid zone, due to their great distance from the zone and the clinal patterns of allele frequencies at the microsatellite loci. Thus, based on the results from the STRUCTURE analysis (above), individuals from marshes 2 (RUMFS) and 8 (Cheesequake) were chosen to represent the most likely 'southern' and 'northern' parents, respectively, with which to train INTROGRESS for the estimation of hybrid indices for populations of killifish located within the hybrid zone (locations 3–7; $n = 311$). The sample from Wantagh, NY (Marsh 9) yielded individual admixture proportions, *q*, and mean-*q* values (figure 4; electronic supplementary material, figure S3) similar to those observed among the samples from Cheesequake and fish from these two populations are not genetically distinct ($F_{ST} = 0.00297; p = 0.17117$). However, we elected to use the individuals from Cheesequake as representative of the northern parental types in the hybrid index analysis due to the relative proximity of these individuals to the central populations of interest in this analysis as it is unlikely that these killifish disperse long distances along the coast. There was a gradual transition from a more southern hybrid index in the south of the hybrid zone to a more northern hybrid index towards the northern edge of the zone (figure 6). However, in the central marsh (4, Metedeconk), the microsatellite data show a lower frequency of intermediate genotypes and a higher frequency of pure northern and southern types upon visual inspection. Formal analysis using the R package DIPTEST led to the rejection of the null hypothesis of unimodality ($D = 0.0927, p = 0.0015$) [49]. The frequency distribution of hybrid indices in marsh 4 (Metedeconk) is suggestive of a bimodal pattern. To strengthen this conclusion, we analysed a combined dataset containing an additional 40 individuals sampled at the same location the following year (November 2009). This sample was genetically indistinguishable ($F_{ST} = 0.00616, p > 0.05$) from the original samples collected in June 2008. The resulting distribution of hybrid indices generated using this larger sample size ($n = 89$) yielded an even more pronounced bimodal pattern ($D = 0.0805, p < 0.0001$; figure 7). Thus, the pattern of hybrid indices observed in Metedeconk cannot be explained by

Figure 6. Frequency distributions of hybrid indices calculated using the most likely 'southern' and 'northern' parents. A hybrid index of 0 represents a pure RUMFS (location 2) genotype while a value of 1 indicates a pure Cheesequake (location 8) genotype.

Figure 7. Hybrid index as calculated using INTROGRESS for individuals collected in Metedeconk across two years. A hybrid index of 0 represents a pure RUMFS (location 2) genotype while a value of 1 indicates a pure Cheesequake (location 8) genotype.

a unimodal distribution, and is best explained by a bimodal distribution. By contrast, a nearby marsh (5, Navesink) had an essentially even distribution of all hybrid indices. However, there was a relatively small sample size from this marsh ($n = 13$), so it is difficult to make firm conclusions about the shape of the distribution of hybrid indices based on this location. Owing to its low sample size, Navesink was excluded from subsequent analyses.

To assess whether the bimodal distribution of hybrid indices among individuals collected from the centrally located Metedeconk marsh was an artefact of the choice of parental populations for training INTROGRESS, we also calculated hybrid index for Metedeconk using all possible 'northern' and 'southern' parental combinations from within the hybrid zone (electronic supplementary material, figure S5). All comparisons yielded varying degrees of a bimodal pattern in frequency distribution of hybrid index at this marsh.

5. Discussion

In this study, we used a combination of mitochondrial DNA and microsatellite markers and a multilocus analytical approach to characterize genetic variation in killifish, *F. heteroclitus*. We thus extended the previous analyses of Adams *et al.* [20] and Duvernell *et al.* [21], who examined variation at these microsatellites but did not calculate a hybrid index, and those of Gonzáles-Villasenor & Powers [18] and Smith *et al.* [19], who examined mtDNA variation but did not examine variation in the nuclear genome. We confirmed that the location of the centre of the mtDNA cline is in north central New Jersey in agreement with previous studies [18,19]. Analysis of mtDNA and microsatellite cline shapes with CFIT7 suggests that a model constraining all clines to the same centre and width has equal likelihood to a model allowing these parameters to vary among loci. This is in direct contrast to previous analyses of these markers [20], which suggested that nuclear and mtDNA cline centres were substantially discordant. Locus-by-locus examination of our cline centre estimates (from HZAR) suggests that the majority of the microsatellite clines are concordant with the mtDNA cline, based on examination of 2-unit support limits. In addition, the loci that have their centres located closest to the mtDNA cline centre all have widths that are less than that predicted under a neutral model, indicating that these clines may be maintained by the same strength and type of selection. By contrast, the two clines with widths that are not significantly different from a neutral cline width (FhATG6 and FhCA-1) have centres (and associated 2-unit support limits) located outside the borders of the hybrid zone, distant from the mitochondrial cline centre. Thus, amalgamating these markers together into a single nuclear cline may result in a false estimate of the nuclear cline centre. This observation suggests that there is additional complexity to the genetic patterns in this species that has not been fully considered in previous analyses.

In addition, we found evidence of a bimodal distribution of nuclear hybrid indices at a marsh at the mtDNA cline centre. This bimodal pattern was strengthened when additional genetically indistinguishable samples from a subsequent year were added to the analysis and persisted when varying combinations of populations from within and immediately adjacent to the hybrid zone were used as reference parental populations. Together, these data suggest the possibility of reproductive isolating barriers acting against first generation and backcrossed types in populations close to the mtDNA cline centre to maintain this genetic structure. However, this pattern was not evident when populations from the extremes of the species distribution were used as parental populations, indicating that there is substantial gene flow across this hybrid zone at a larger scale.

Bimodal hybrid zones have been described across a wide range of taxa including fish (e.g. [53,54]), molluscs (e.g. [55,56]), insects (e.g. [57,58]), birds (e.g. [41,59]), plants (e.g. [60,61]), reptiles (e.g. [62,63]) and mammals (e.g. [64]). The presence of a bimodal hybrid zone is an important genetic signature of reproductive isolation, in which parental types dominate a particular location, and a deficit of hybrid types is observed. The results of our hybrid index analyses coupled with the observation of elevated occurrences of linkage disequilibrium in these marshes near the centre of the mtDNA and several microsatellite clines reinforce the conclusion that there is some form of reproductive isolation between the northern and southern types of *F. heteroclitus* influencing population structure in this region. However, our results also indicate that this reproductive isolation is not complete, and that there is still substantial gene flow between the subspecies. This conclusion is consistent with the broad clines observed at some loci in *F. heteroclitus* [17,22,65].

The main factors that could promote the formation of the patterns we observed are incomplete premating and/or pre-zygotic isolation, selection against backcross types, and the inability of recombination to breakdown co-adapted gene complexes (as evidenced by the presence of linkage

disequilibrium among genetic markers within the hybrid zone [7]). This maintenance of non-random associations of particular genotypic combinations in an area where two subspecies are known to be sympatric can be the result of several phenomena. For example, selection against hybrid offspring produced by matings between 'more pure' northern-type and southern-type parents that are constantly migrating in from neighbouring populations can create a balance between dispersal and selection [9]. Alternatively, phenomena such as positive assortative mating can also produce these genetic associations and distributions of hybrid indices in the hybrid zone [7]. Indeed, these phenomena are not mutually exclusive, and both may be operating. For example, strong selection against hybrids harbouring maladapted combinations of genotypes as the result of matings between genetically distinct parental types is likely to result in reinforcement through the strengthening of pre-zygotic barriers (i.e. assortative mating, ecological character displacement) [66]. Such pre-zygotic barriers tend to be favoured in response to existing post-zygotic barriers (i.e. hybrid sterility, hybrid inviability) because they operate early in an organism's life cycle and as such will reduce an individual's incurred cost of wasting gametes in production of less fit hybrid offspring [67]. At this point, it is reasonable to conclude that killifish residing in the hybrid zone are experiencing some combination of pre-zygotic and/or post-zygotic isolation that results in reduced numbers of F_1 and backcross genotypes in the population. It is important to note that at this time we are unable to distinguish whether a pre-zygotic or post-zygotic barrier is the main contributor to the reproductive isolation we are seeing among these subspecies, because the minimum size of fish retained by the minnow trap was 32 mm. Thus, we cannot differentiate among a scenario in which the hybrids are not being produced at all (pre-zygotic isolation) or one in which the hybrids are being produced but are not surviving beyond this minimum size (post-zygotic isolation).

6. Conclusion

In summary, we have found evidence of a bimodal hybrid zone connecting the northern and southern subspecies of the killifish, *F. heteroclitus*, for multilocus genotypes assessed with highly variable microsatellite markers. This bimodal pattern in the centrally located marshes near the centres of the steep mtDNA D-loop and microsatellite clines suggests some level of pre- and/or post-zygotic reproductive isolation is operating to prevent the complete merging of these two subspecies [7]. Together, these results emphasize the importance of collecting data on both the nuclear and cytoplasmic genomes and considering multilocus genotypes when trying to uncover the underlying genetic structure of closely related hybridizing taxa.

References

1. Dobzhansky T. 1940 Speciation as a stage in evolutionary divergence. *Am. Nat.* **74**, 312–321. (doi:10.1086/280899)

2. Hewitt GM. 1988 Hybrid zones—natural laboratories for evolutionary studies. *Trends Ecol. Evol.* **3**, 158–167. (doi:10.1016/0169-5347(88)90033-X)

3. Barton NH, Hewitt GM. 1989 Adaptation, speciation, and hybrid zones. *Nature* **341**, 497–503. (doi:10.1038/341497a0)

4. Sotka EE, Palumbi SR. 2006 The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology* **87**, 1094–1103. (doi:10.1890/0012-9658(2006)87[1094:TUOGCT]2.0.CO;2)

5. Teeter KC *et al*. 2008 Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* **18**, 67–76. (doi:10.1101/gr.6757907)

6. Szymura JM, Barton NH. 1986 Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* **40**, 1141–1159. (doi:10.2307/2408943)

7. Jiggins CD, Mallet J. 2000 Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* **15**, 250–255. (doi:10.1016/S0169-5347(00)01873-5)

8. Cheviron ZA, Brumfield RT. 2009 Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution* **63**, 1593–1605. (doi:10.1111/j.1558-5646.2009.00644.x)

9. Barton NH, Hewitt GM. 1985 Analysis of hybrid zones. *Annu. Rev. Ecol. Evol. Syst.* **16**, 113–148. (doi:10.1146/annurev.es.16.110185.000553)

10. Haldane JBS. 1948 The theory of a cline. *J. Genet.* **7**, 277–284. (doi:10.1007/BF02986626)

11. Slatkin M. 1973 Gene flow and selection in a cline. *Genetics* **75**, 733–756.

12. Larson EL, White TA, Ross CL, Harrison RG. 2014 Gene flow and the maintenance of species boundaries. *Mol. Ecol.* **23**, 1668–1678. (doi:10.1111/mec.12601)

13. Vines TH. 2014 Stuck in the middle with you: close concordance between geographical clines in a cricket hybrid zone. *Mol. Ecol.* **23**, 1647–1649. (doi:10.1111/mec.12692)

14. Avise JC. 2004 *Molecular markers, natural history, and evolution*, 2nd edn. Sunderland, MA: Sinauer Associates.

15. Freeman S, Herron JC. 2004 *Evolutionary analysis*, 3rd edn. Upper Saddle River, NJ: Pearson Education, Inc.

16. Powers DA, Ropson I, Brown DC, Van Beneden R, Cashon R, Gonzáles-Villasenor LI, DiMichele JA. 1986 Genetic variation in *Fundulus heteroclitus*: geographic distribution. *Am. Zool.* **26**, 131–144. (doi:10.1093/icb/26.1.131)

17. Powers DA, Place AR. 1978 Biochemical genetics of *Fundulus heteroclitus* (L.). I. Temporal and spatial variation in gene frequencies of Ldh-B, Mdh-A, Gpi-B, and Pgm-A. *Biochem. Genet.* **16**, 593–607. (doi:10.1007/BF00484222)

18. Gonzáles-Villasenor LI, Powers DA. 1990 Mitochondrial-DNA restriction-site polymorphisms in the teleost *Fundulus heteroclitus* support secondary intergradation. *Evolution* **44**, 27–37. (doi:10.2307/2409522)

19. Smith MW, Chapman RW, Powers DA. 1998 Mitochondrial DNA analysis of Atlantic Coast, Chesapeake Bay, and Delaware Bay populations of the teleost *Fundulus heteroclitus* indicates temporally unstable distributions over geologic time. *Mol. Mar. Biol. Biotechnol.* **7**, 79–87.

20. Adams SM, Lindmeier JB, Duvernell DD. 2006 Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Mol. Ecol.* **15**, 1109–1123. (doi:10.1111/j.1365-294X.2006.02859.x)

21. Duvernell DD, Lindmeier JB, Faust KE, Whitehead A. 2008 Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. *Mol. Ecol.* **17**, 1344–1360. (doi:10.1111/j.1365-294X.2007.03648.x)

22. Strand AE, Williams LM, Oleksiak MF, Sotka EE. 2012 Can diversifying selection be distinguished from history in geographic clines? A population genomic study of killifish (*Fundulus heteroclitus*). *PLoS ONE* **7**, e45138 (doi:10.1371/journal.pone.0045138)

23. Able KW, Felley JD. 1986 Geographical variation in *Fundulus heteroclitus*: tests for concordance between egg and adult morphologies. *Am Zool.* **26**, 145–157. (doi:10.1093/icb/26.1.145)

24. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

25. Falush D, Stephens M, Pritchard JK. 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574–578. (doi:10.1111/j.1471-8286.2007.01758.x)

26. Fangue NA, Hofmeister M, Schulte PM. 2006 Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Funudulus heteroclitus*. *J. Exp. Biol.* **209**, 2859–2872. (doi:10.1242/jeb.02260)

27. Lee W, Conroy J, Howell WH, Kocher TD. 1995 Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol.* **41**, 54–66.

28. Tamura K, Dudley J, Nei M, Kumar S. 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599. (doi:10.1093/molbev/msm092)

29. Vincze T, Posfai J, Roberts RJ. 2003 NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* **31**, 3688–3691. (doi:10.1093/nar/gkg526)

30. Adams SM, Oleksiak MF, Duvernell DD. 2005 Microsatellite primers for the Atlantic coastal killifish, *Fundulus heteroclitus*, with applicability to related *Fundulus* species. *Mol. Ecol. Notes* **5**, 275–277. (doi:10.1111/j.1471-8286.2005.00898.x)

31. Van Oosterhout C, Hutchinson WF, Willis DPM, Shipley P. 2004 MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538. (doi:10.1111/j.1471-8286.2004.00684.x)

32. Raymond M, Rousset F. 1995 Genepop (v. 1.2), population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249.

33. Belkhir K, Borsa P, Goudet J, Bonhomme F. 1999 *GENETIX 4.04*. Montpellier, France: Université de Montpellier II.

34. Cockerham CC, Weir BS. 1977 Quadratic analyses of reciprocal crosses. *Biometrics* **33**, 187–203. (doi:10.2307/2529312)

35. Asmussen MA, Basten CJ. 1996 Constraints and normalized measures for cytonuclear disequilibria. *Heredity* **76**, 207–214. (doi:10.1038/hdy.1996.33)

36. Basten CJ, Asmussen MA. 1997 The exact test for cytonuclear disequilibria. *Genetics* **146**, 1165–1171.

37. Derryberry EP, Deryberry GE, Maley JM, Brumfield RT. 2014 HZAR: hybrid zone analysis using an R software package. *Mol. Ecol. Res.* **14**, 652–663. (doi:10.1111/1755-0998.12209)

38. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

39. Brumfield RT, Jernigan RW, McDonald DB, Braun MJ. 2001 Evolutionary implications of divergent clines in an avian (*Mancus*: Aves) hybrid zone. *Evolution* **55**, 2070–2087. (doi:10.1111/j.0014-3820.2001.tb01322.x)

40. Barton NH, Gale KS. 1993 Genetic analysis of hybrid zones. In *Hybrid zones and the evolutionary process* (ed. RG Harrison), pp. 13–45. New York, NY: Oxford University Press.

41. Gay L, Crochet P, Bell D, Lenormand T. 2008 Comparing genetic and phenotypic clines in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789–2806. (doi:10.1111/j.1558-5646.2008.00491.x)

42. Rieux A, Lenormand T, Carlier J, de Lapeyre de Bellaire L, Ravigné V. 2013 Using neutral cline decay to estimate contemporary dispersal: a generic tool and its implication to a major crop pathogen. *Ecol. Lett.* **16**, 721–730. (doi:10.1111/ele.12090)

43. Whitlock M, Schluter D. 2009 *The analysis of biological data*. Greenwood Village: Roberts and Co. Publishers.

44. Evanno G, Regnaut S, Goudet J. 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620. (doi:10.1111/j.1365-294X.2005.02553.x)

45. Earl DA, vonHoldt BM. 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. (doi:10.1007/s12686-011-9548-7)

46. Gompert Z, Buerkle CA. 2009 A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.* **18**, 1207–1224. (doi:10.1111/j.1365-294X.2009.04098.x)

47. Gompert Z, Buerkle CA. 2010 INTROGRESS: a software package for mapping components of isolation in hybrids. *Mol. Ecol. Res.* **10**, 378–384. (doi:10.1111/j.1755-0998.2009.02733.x)

48. Hartigan PM. 1985 Computation of the dip statistic to test for unimodality. *Appl. Stat. J. R. Stat. Soc. C* **34**, 320–325. (doi:10.2307/2347485)

49. Hartigan JA, Hartigan PM. 1985 The dip test of unimodality. *Ann. Stat.* **1**, 70–84. (doi:10.1214/aos/1176346577)

50. Pike N. 2010 Using false discovery rate for multiple comparisons in ecology and evolution. *Methods Ecol. Evol.* **2**, 278–282. (doi:10.1111/j.2041-210X.2010.00061.x)

51. Bernardi G, Sordino P, Powers DA. 1993 Concordant mitochondrial and nuclear-DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proc. Natl Acad. Sci. USA* **90**, 9271–9274. (doi:10.1073/pnas.90.20.9271)

52. Fritz ES, Meredith WH, Lotrich VA. 1975 Fall and winter movements and activity level of the mummichog, *Fundulus heteroclitus*, in a tidal creek. *Chesapeake Sci.* **16**, 211–215. (doi:10.2307/1350898)

53. Redenbach Z, Taylor EB. 2003 Evidence for bimodal hybrid zones between two species of char (Pisces: *Salvelinus*) in northwestern North America. *J. Evol. Biol.* **16**, 1135–1148. (doi:10.1046/j.1420-9101.2003.00619.x)

54. Aboim MA, Mavárez J, Bernatchez L, Coelho MM. 2010 Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. *J. Evol. Biol.* **23**, 817–828. (doi:10.1111/j.1420-9101.2010.01953.x)

55. Johannesson K, Rolán-Alvarez E, Ekendahl A. 1995 Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**, 1180–1190. (doi:10.2307/2410443)

56. Riginos C, Cunningham CW. 2005 Invited review: local adaptation and species segregation in two mussel (*Mytilus edulis × Mytilus trossulus*) hybrid zones. *Mol. Ecol.* **14**, 381–400. (doi:10.1111/j.1365-294X.2004.02379.x)

57. Bridle JR, Butlin RK. 2002 Mating signal variation and bimodality in a mosaic hybrid zone between *Chorthippus* grasshopper species. *Evolution* **56**, 1184–1198. (doi:10.1111/j.0014-3820.2002.tb01431.x)

58. Harrison RG, Bogdanowicz SM. 1997 Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution* **51**, 493–505. (doi:10.2307/2411122)

59. Bensch S, Helbig AJ, Salomon M, Seibold I. 2002 Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Mol. Ecol.* **11**, 473–481. (doi:10.1046/j.0962-1083.2001.01455.x)

60. Cruzan MB, Arnold ML. 1993 Ecological and genetic associations in an *Iris* hybrid zone. *Evolution* **47**, 1432–1445. (doi:10.2307/2410158)

61. Minder AM, Rothenbuehler C, Widmer A. 2007 Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (Caryophyllaceae): evidence for introgressive hybridization. *Mol. Ecol.* **16**, 2504–2516. (doi:10.1111/j.1365-294X.2007.03292.x)

62. Fitzpatrick BM, Placyk Jr JS, Niemiller ML, Casper GS, Burghardt GM. 2008 Distinctiveness in the face of gene flow: hybridization between specialist and generalist gartersnakes. *Mol. Ecol.* **17**, 4107–4117. (doi:10.1111/j.1365-294X.2008.03885.x)

63. Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N. 2009 Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: evidence for limited natural hybridization and a bimodal hybrid zone. *J. Zool. Syst. Evol. Res.* **47**, 368–377. (doi:10.1111/j.1439-0469.2009.00532.x)

64. Bulatova NS, Shchipanov NA, Searle JB. 2007 The Seliger–Moscow hybrid zone between chromosome races of common shrews–an initial description. *Russ. J. Theriol.* **6**, 111–116.

65. Cashon RE, Van Beneden RJ, Powers DA. 1981 Biochemical genetics of *Fundulus heteroclitus* (L.). IV. Spatial variation in gene frequencies of Idh-A, Idh-B, 6-Pgdh-A, and Est-S. *Biochem. Genet.* **19**, 715–728. (doi:10.1007/BF00484004)

66. Servedio MR, Noor MAF. 2003 The role of reinforcement in speciation: theory and data. *Annu. Rev. Ecol. Evol. Syst.* **34**, 339–364. (doi:10.1146/annurev.ecolsys.34.011802.132412)

67. Coyne JA, Orr HA. 2004 *Speciation*. Sunderland, MA: Sinauer Associates.