

# Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*

DAVID D. DUVERNELL,\* JAMES B. LINDMEIER,\* KATHERINE E. FAUST† and ANDREW WHITEHEAD†

\*Department of Biological Sciences, Southern Illinois University Edwardsville, Edwardsville, IL 62026, USA,

†Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

## Abstract

A major goal of population genetics research is to identify the relative influences of historical and contemporary processes that serve to structure genetic variation. Most population genetic models assume that populations exist in a state of migration-drift equilibrium. However, in the past this assumption has rarely been verified, and is likely rarely achieved in natural populations. We assessed the equilibrium status at both local and regional scales of the Atlantic killifish, *Fundulus heteroclitus*. This species is a model organism for the study of adaptive clinal variation, but has also experienced a complicated history of range expansion and secondary contact following allopatric divergence, potentially obscuring the influence of contemporary evolutionary processes. Presumptively neutral genetic markers (microsatellites) demonstrated zones of secondary intergradation among coastal populations centred around northern New Jersey and the Chesapeake Bay region. Analysis of genetic variation indicated isolation by distance among some populations and provided supporting evidence that the Delaware Bay, but not the Chesapeake Bay, has acted as a barrier to dispersal among coastal populations. Bayesian estimates indicated large effective population sizes and low migration rates, and were in good agreement with empirically derived estimates of population and neighbourhood size from mark–recapture studies. These data indicate that populations are not in migration-drift equilibrium at a regional scale, and suggest that contributing factors include large population size combined with relatively low migration rates. These conditions should be considered when interpreting the evolutionary significance of the distribution of genetic variation among *F. heteroclitus* populations.

*Keywords:* effective population size, equilibrium, gene flow, introgression, isolation by distance, microsatellite, secondary contact

Received 10 September 2007; revision accepted 17 November 2007

## Introduction

A major ambition of population genetic studies is to understand how environments shape the evolution of populations. However, a basic understanding of how contemporary environments shape populations is often confounded by the legacy of population history. In general, genetic drift and local adaptation are both counteracted by the homogenizing effects of gene flow. If genetic drift and

gene flow are in a state of equilibrium, where genetic variance among populations stabilizes over time because the homogenizing effect of migration balances the diversifying effect of genetic drift, then interactions among populations may be reflected in the distributions of genetic variation. Deviations of genetic patterns from biogeographical expectations may be explained by the influence of contemporary environments (e.g. local adaptation). However, populations are rarely in a state of migration-drift equilibrium because of widespread demographic instability (Whitlock 1992). Historical events such as range expansion and secondary contact can complicate the interpretation of

Correspondence: David Duvernell, Fax: 618-650-3174; E-mail: dduvern@siue.edu

the effects of contemporary natural selection across heterogeneous environments and obscure current patterns of migration and genetic exchange. Recent range expansions may result in weak population structure that could be misinterpreted as evidence of extensive gene flow (e.g. Pogson *et al.* 2001). Conversely, secondary contact may yield populations that exhibit pronounced geographical structure (e.g. Turgeon & Bernatchez 2001; Rey & Turgeon in review) and clinal patterns associated with secondary intergradation (Endler 1977). For both scenarios, the time required to reach equilibrium is directly proportional to population size and inversely proportional to gene flow (Whitlock 1992). Consequently, basic assumptions of population genetic models, most notably that of migration-drift equilibrium, must be carefully evaluated at multiple spatial scales before current patterns of genetic structure may be interpreted with confidence (Hutchison & Templeton 1999; Pogson *et al.* 2001; Austin *et al.* 2004).

For species in which dispersal is constrained by distance, the stepping-stone model (Kimura & Weiss 1964) makes the prediction that, under genetic equilibrium conditions, populations will exhibit a pattern of isolation by distance (IBD). Therefore, a common approach for assessing migration-drift equilibrium is to measure correlations between genetic divergence and spatial distance among population samples (Hutchison & Templeton 1999) to test for IBD. However, the presence of an IBD pattern does not necessarily imply migration-drift equilibrium as nonequilibrium populations may also exhibit patterns of IBD (Slatkin 1993; Bossart & Prowell 1998). Furthermore, the stepping-stone model assumes that environments are uniform such that distance is the only factor that influences genetic exchange among populations. If environments are not uniform (e.g. identifiable barriers to dispersal exist), then the presence of a pattern of IBD could also imply that populations are not in migration-drift equilibrium. Therefore, one way of assessing whether the presence of an IBD pattern reflects ongoing migration-drift equilibrium conditions is to identify critical geographical features that may either limit or facilitate dispersal and test for the expected signal of those features within the distribution of genetic variation among populations (Poissant *et al.* 2005; Crispo *et al.* 2006; Rey & Turgeon in review). In the present study, we assess patterns of genetic structure and IBD in a broadly distributed estuarine killifish species and estimate population parameters of effective population size ( $N_e$ ) and migration rate ( $m$ ) to infer the relative influences of historical events vs. contemporary influences on the distribution of genetic variation.

The mummichog, *Fundulus heteroclitus*, has long been a model for the study of adaptive clinal variation, and provides an excellent system in which to examine the relative influences of historical and contemporary forces on structuring of genetic variation. *F. heteroclitus* is a small (< 10 cm) teleost killifish distributed from Newfoundland to Northern

Florida along the Atlantic seashore in coastal salt marsh habitats, and is also found in brackish, low-salinity habitats in suitable bay and river environments. The species is non-migratory, generally confined to salt marshes and tidal creek habitats, and individuals do not venture far beyond the shore line, preferring water less than a couple metres in depth (Bigelow & Schroeder 1953). Females produce dimersal eggs in the salt marsh habitats, and there is no planktonic larval stage (Kneib 1997). Mark-recapture studies have indicated that seasonal dispersal of individuals is less than 1 km (Lotrich 1975; Sweeney *et al.* 1998). This species is considered a model for studying adaptive clinal variation across large spatial scales (reviewed in Powers *et al.* 1986, 1991; Powers & Schulte 1998) as well as derived adaptive tolerance to pollution at local scales (e.g. Eisler 1986; Cohen 2002; Powell *et al.* 2000; Meyer & Di Giulio 2003; Hahn *et al.* 2004; McMillan *et al.* 2006).

*Fundulus heteroclitus* exhibits extensive latitudinal clinal variation in a number of physiological and biochemical traits, coupled with phylogeographical patterns at mitochondrial and nuclear DNA loci that indicate a complicated history of spatially variable selection and secondary intergradation. Concordance of clinal patterns in multiple morphological features (Able & Felley 1986), allelic isozymes (Ropson *et al.* 1990), mtDNA haplotypes (Gonzalez-Villasenor & Powers 1990; Smith *et al.* 1998), and presumptively neutral microsatellite loci (Adams *et al.* 2006) all indicate a major zone of secondary contact between 'northern' and 'southern' Pleistocene relict populations of the species in northern New Jersey, approximately corresponding to the region where the Hudson River enters the Atlantic Ocean. South of New Jersey, disjunct 'northern' populations also occur in the extreme upper-estuary reaches of the Chesapeake and Delaware Bays (Smith *et al.* 1998). Based on this evidence, it has been suggested that the distribution of northern and southern populations of the species have fluctuated during the past several hundred thousand years, and that their present distributions may reflect multiple population expansion and contraction events (Smith *et al.* 1998).

What remains unresolved is the nature and magnitude of the contemporary forces (e.g. drift, gene flow) that govern the distribution of genetic variation within and among populations of this species. Brown & Chapman (1991) performed a study of mtDNA genetic structure in a network of local populations over the geographical range of tens of kilometres of continuously distributed favourable *F. heteroclitus* habitat in the Chesapeake Bay region. Based on their estimates at this local scale, they concluded that the number of effective migrants (i.e. the product  $N_e m$ ) should be sufficient to homogenize distribution-wide gene frequency clines, and that selective forces are therefore most likely responsible for maintaining current distributions of these genetic variants (Brown & Chapman 1991). However, while  $N_e m$  may inform the extent to which populations will

diverge at migration–drift equilibrium, the product of these two key parameters is not sufficient to address how quickly populations may shift equilibrium in response to environmental change or disturbance. Furthermore, the conclusions of this study ignored the impact of environmental discontinuity on gene flow at broader geographical scales.

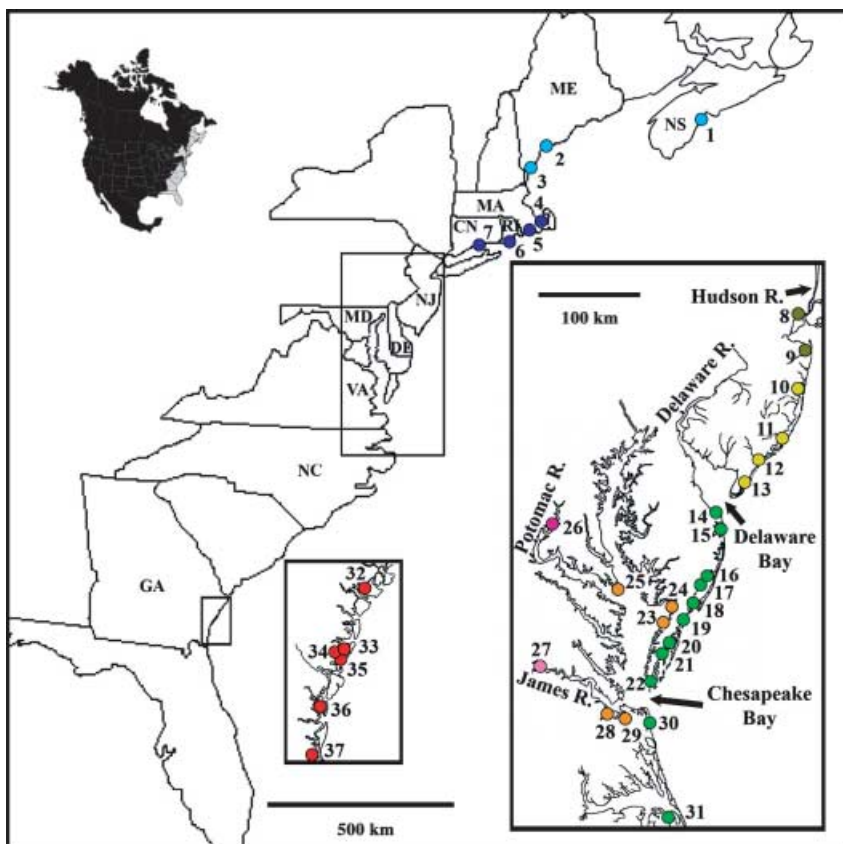
The goal of the present study was to examine the distribution of genetic variation among *F. heteroclitus* populations at multiple geographical scales and with respect to known barriers to contemporary gene flow and known zones of secondary contact. The objectives of this study were: (i) to assess presumptive zones of secondary contact to assess the status of populations in those regions with regard to genetic equilibrium; (ii) to identify the regions and spatial scales at which populations exhibit a pattern of IBD and to assess whether those patterns reflect a state of genetic equilibrium; (iii) to obtain genetic estimates of  $N_e$  and  $m$  and assess how these parameters have influenced current distributions of genetic variation among populations; and (iv) to compare genetic estimates of  $N_e$  and  $m$  with available direct estimates from mark–recapture studies. We find that the distribution of genetic variation among populations throughout the range of *F. heteroclitus* appears to be more greatly influenced by historical contingencies than by contemporary gene flow. Therefore, care must be taken to avoid discounting

population history when interpreting geographical patterns of variation in this species.

## Materials and methods

Overall, population genetic structure was characterized among 37 spatially distributed populations collected from across the native range of *Fundulus heteroclitus* using microsatellites. Our choice of analytical methods was centred around the objective of characterizing population structure at both local and regional scales and characterizing the evolutionary mechanisms likely responsible for local and regional structure. Bayesian methods were used to characterize large-scale geographical structure, and admixture models were used to assess zones of secondary contact. IBD was measured at local and regional scales using Mantel tests, and the presence of contemporary barriers to dispersal was evaluated using multiple Mantel tests, analysis of molecular variance (AMOVA), and computational geometry approaches. Migration rates and effective population sizes were estimated using Bayesian models.

Population samples were collected in the summers of 2002, 2004, 2005 and 2006 using Gee-type minnow traps, or seines (Table 1, Fig. 1). Populations sampled north of the Hudson River (1–7) are referred as northern samples, those



**Fig. 1** Locations for 37 population samples of *Fundulus heteroclitus*. Samples numbered according to Table 1. Sample colours correspond to 9 genetic clusters identified by BAPS. Abbreviations: NS, Nova Scotia; ME, Maine; MA, Massachusetts; RI, Rhode Island; CN, Connecticut; NJ, New Jersey; DE, Delaware; MD, Maryland; VA, Virginia; NC, North Carolina; GA, Georgia.

**Table 1** Sample abbreviations, geographical locations and sample sizes for *Fundulus heteroclitus* population samples

Sample	Abbreviation	General collection site	Latitude (N), longitude (W)	Year	<i>n</i>
1	HAVNS*	Bridgewater, NS	44°22.0', 64°31.0'	2002	17
2	CME*	Chewonki, ME	43°57.3', 69°43.2'	2002	50
3	WME*	Wells, ME	43°19.2', 70°34.2'	2002	32
4	BSMA*	Barnstable, MA	41°44.0', 70°23.0'	2002	50
5	NBMA*	New Bedford, MA	41°34.0', 70°54.9'	2002	20
6	JRI	Jerusalem, RI	41°23.1', 71°31.5'	2005	33
7	CCN*	Clinton, CN	41°15.3', 72°32.8'	2002	50
8	NBNJ*	Newark Bay, NJ	40°41.2', 74°06.7'	2002	51
9	REDNJ	Red Bank, NJ	40°20.9', 74°05.0'	2004	60
10	TOMNJ	Holiday City, NJ	39°59.7', 74°09.0'	2004	66
11	TNJ*	Tuckerton, NJ	39°32.2', 74°19.4'	2002	50
12	EGGNJ	Somers Point, NJ	39°18.8', 74°33.8'	2004	66
13	SHNJ*	Stone Harbor, NJ	39°03.6', 74°46.4'	2002	99
14	LEWDE	Lewes, DE	38°46.5', 75°08.1'	2005	80
15	IRBDE	Indian River Bay, DE	38°38.2', 75°04.1'	2005	73
16	SPEMD	Speace, MD	38°09.1', 75°17.2'	2005	60
17	WMD	Stockton, MD	38°04.6', 75°21.9'	2005	57
18	CHIVA	Chincoteague, VA	37°56.1', 75°25.1'	2005	57
19	GARVA	Gargathy, VA	37°46.5', 75°33.7'	2005	53
20	QUIVA	Upshur, VA	37°32.9', 75°43.9'	2005	58
21	HOGVA	Hog Bay, VA	37°26.7', 75°50.4'	2005	59
22	MVA*	Magotha, VA	37°10.6', 75°56.5'	2002	22
23	CHEVA	Chesconessex Creek, VA	37°44.9', 75°46.4'	2005	25
24	HAMVA	Hammock Landing, VA	37°54.2', 75°41.0'	2005	30
25	FhMC4	Point Lookout, MD	38°3.5', 76°19.5'	2006	26
26	FhMC3	Piscataway Park, MD	38°41.7', 77°3.2'	2006	42
27	FhVC2	Shirley Plantation, VA	37°19.7', 77°15.7'	2006	62
28	CHVA	Suffolk, VA	36°51.8', 76°28.7'	2006	20
29	NVA*	Norfolk, VA	36°48.5', 76°17.7'	2002	60
30	RUDVA	Rudee Heights, VA	36°49.4', 75°58.9'	2006	60
31	RINC*	Roanoke Island, NC	35°53.8', 75°36.9'	2002	52
32	SIGA	Skidaway Island, GA	31°56.8', 81°04.2'	2004	60
33	SICGA*	Sapelo Island, GA	31°29.3', 81°15.9'	2002	77
34	SIFGA*	Sapelo Island, GA	31°27.2', 81°21.8'	2002	50
35	SISGA	Sapelo Island, GA	31°24.8', 81°17.8'	2002	71
36	JIGA	Jekyll Island, GA	31°04.0', 81°27.5'	2004	70
37	SMGA	St Marys, GA	30°45.2', 81°35.0'	2004	62

\*Reported in Adams *et al.* (2006).

sampled from New Jersey to North Carolina (9–31, Fig. 1 large inset) are referred as mid-Atlantic samples, those sampled along the coast between Delaware Bay and Chesapeake Bay (14–22) are coastal Delmarva (Delaware, Maryland, Virginia) samples, and those from Georgia (32–37, Fig. 1, small inset) are southern samples. Sampling density was focused primarily in the mid-Atlantic region in order to explore the effects of population history in the portion of the species distribution where clinal patterns have previously been described using other genetic markers (e.g. mtDNA, Smith *et al.* 1998). The southern end of the species distribution was also sampled at a fine geographical scale to provide comparative data on local genetic structure among populations

presumed to be least impacted by climatic fluctuations in recent population history.

Samples from 2002 were reported in Adams *et al.* (2006). For samples collected in 2004–2006, fin tissues were removed at the collection site from the pectoral or caudal fin, and either preserved in 95% ethanol or dried in manila envelopes. Genomic DNA was extracted with the aid of AquaPure Genomic DNA Tissue Kits (Bio-Rad) or with DNeasy tissue kits (QIAGEN). Individuals were genotyped using eight trinucleotide (ATG) microsatellite loci described by Adams *et al.* (2006). These loci were selected from a larger panel of di- and trinucleotide loci (Adams *et al.* 2005) specifically for their relatively low allelic diversity (8–19 alleles per locus;

Adams *et al.* 2006). Within collection years, sample orders were randomized with respect to location of origin for data collection purposes, and markers were amplified in separate reactions using the polymerase chain reaction (PCR) protocols described in Adams *et al.* (2005). One primer was fluorescently labelled and PCR products were electrophoresed on either an ABI 310 (SIUE) or an ABI 3130XL (LSU) (Applied Biosystems). DNA fragments were analysed, and genotypic data were generated using GENEMAPPER version 3.0 (Applied Biosystems). A reference set of individuals was employed at both SIUE and LSU to establish allele size identity across all samples. Tests for Hardy–Weinberg equilibrium within samples were conducted using the global test implemented in FSTAT version 2.9.3 (Goudet 1995) with 10 000 permutations. Tests for linkage disequilibrium were conducted over all samples using the log-likelihood  $G$ -statistic implemented in FSTAT with the nominal level of statistical significance set at 0.05. Population pairwise multilocus estimates of  $F_{ST}$  (Weir and Cockerham's  $\theta$ ; Weir & Cockerham 1984) were also calculated using FSTAT. Throughout this report, Weir and Cockerham's  $\theta$  is referred as  $F_{ST}$  to avoid confusion with the population parameter  $\theta = 4N_e\mu$ , which is also extensively used.

Given the wide geographical distribution of *F. heteroclitus*, and the potential for long coalescent times among some populations, the assumption that stepwise mutation (the stepwise mutation model, SMM) has not significantly affected population divergence was evaluated. We tested the influence of stepwise mutation relative to drift on population divergence both globally and for population pairs at all geographical scales using the randomization test of Hardy *et al.* (2003). This test compares observed estimates of  $R_{ST}$ , which are based on allele size differences, to simulated values of  $\rho R_{ST}$  in which allele size classes have been randomized, resulting in measures analogous to  $F_{ST}$ . If stepwise mutations have contributed to population divergence, then  $R_{ST}$  will be greater than  $\rho R_{ST}$ , and  $R_{ST}$  is a more appropriate measure of population divergence than  $F_{ST}$ . If stepwise mutation has not contributed to population divergence, then studies have shown that  $F_{ST}$  is preferable to  $R_{ST}$  (Gaggiotti *et al.* 1999). The tests were performed based on 1000 permutations using SPAGEDI version 1.1 (Hardy & Vekemans 2002).

#### *Geographical patterns of genetic structure*

The Bayesian inference method performed by the program BAPS version 4.4 (Corander *et al.* 2003) was employed to assess genetic structure across the entire geographical range sampled and to assign population samples to population clusters. The data were analysed with individuals assigned to population samples and the number of predefined groups set to a range of  $k = 2$ –12. The best-supported cluster assignment was identified by assessment of the posterior probability of the log (ml) values. For comparative purposes,

the program STRUCTURE version 2.1 (Pritchard *et al.* 2000) was employed to calculate the proportion of membership of each population sample into a predefined set of groups that also ranged from  $k = 2$ –12. The admixture model was used with run parameters of 1 million Markov chain Monte Carlo (MCMC) replicates and a burn-in period of 50 000.

#### *Detecting contemporary barriers to dispersal*

Regions were assessed for conformance to migration-drift equilibrium according to a stepping-stone model of gene flow. According to the IBD model, if populations are in migration-drift equilibrium, a positive monotonic relationship is expected between population divergence and geographical distance (assuming mutation is negligible relative to migration) (Hutchison & Templeton 1999). A second prediction of the model, when populations are distributed in two dimensions, is that the variance around estimates of  $F_{ST}$  should increase with geographical distance (Hutchison & Templeton 1999). However, when populations are distributed in one-dimensional environments (e.g. coastal habitats), this latter prediction does not hold, and in fact, variance in  $F_{ST}$  estimates may be reduced at longer geographical distances as the genetic distances among populations at the ends of a linear array of samples are all correlated. To conservatively minimize the effect of these longer geographical distances on correlation tests, we performed a log transformation of geographical distances.

IBD patterns were assessed by calculating the significance of correlations between linearized pairwise  $F_{ST}$  ( $F_{ST}/1 - F_{ST}$ ) and log-geographical distances using Mantel's test (Mantel 1967) as implemented in IBD version 1.52 (Bohonak 2002) with 10 000 permutations. For simplicity, and repeatability of the analyses, geographical distances were determined as straight-line distances among collection sites. Given the linearity of the coastal habitats that were assessed, these distances did not differ substantially from distances that were calculated by following the shoreline distance between pairs of populations and did not affect the outcome of the analyses (shoreline distance analyses not shown). Distances were estimated using MAPSOURCE version 3.02 (Garmin). A multiple factor Mantel permutation test design was employed to separate the effects of historical divergence resulting from putative dispersal barriers (e.g. large bays) from ongoing IBD (Smouse *et al.* 1986; Sacks *et al.* 2004; Telles & Diniz-Filho 2005). The matrices analysed included linearized genetic distances, log-geographical distances, and binary matrices expressing regional assignment where 0 indicated two sample locations in the same region and 1 indicated two sampling locations in different regions. In order to test robustness of these associations, an AMOVA (using ARLEQUIN 3.11; Excoffier *et al.* 1992) was also performed.

Computational geometry routines as implemented in the program BARRIER version 2.2 (Manni *et al.* 2004) were

employed to perform an additional test for the presence of dispersal barriers among coastal populations. In contrast to the multiple Mantel and AMOVA tests, this analysis made no a priori predictions as to where dispersal barriers would occur among populations. BARRIER uses a multiple matrix approach to test the shape of the genetic ( $F_{ST}$ ) relative to geographical landscapes to identify regions of reduced gene flow.

#### Estimation of effective population size and migration rates

Traditional estimates of population structure and gene flow were obtained by applying Wright's (1969) approximation to pairwise estimates of  $F_{ST}$  (Weir and Cockerham's  $\theta$ ; Weir & Cockerham 1984). According to Wright's island model, if appropriate assumptions are met (including no mutation and migration drift-equilibrium among others; reviewed by Whitlock & McCauley 1999), the variance in allele frequencies among populations is related to the number of migrants entering a population each generation according to the relationship  $N_e m = 0.25(1/F_{ST} - 1)$ .

Bayesian-scaled estimates of long-term effective population size ( $N_e$ ) and migration rates ( $m$ ), were obtained using the program MIGRATE version 2.1 (Beerli & Felsenstein 2001; Beerli 2006). MIGRATE simultaneously estimates the parameters  $\theta_i$  ( $\theta = 4N_e\mu$ ) and  $M_{j \rightarrow i}$  ( $M = m/\mu$ ) by exploring all possible genealogies for a set of population samples assuming that all populations are at migration-drift equilibrium and have maintained constant size over the coalescent period (Beerli & Felsenstein 1999). The program was implemented with the Bayesian approach using the CBSU web computing interface at Cornell University (<http://cbsuapps.tc.cornell.edu/migrate.aspx>).

Estimates of  $\theta$  ( $= 4N_e\mu$ ) were obtained for all population samples by performing a series of separate analyses for each population sample. For each sample, 30 subsampled individuals (when sample sizes were  $> 30$ ) were analysed and the results of two or more independent repetitions of the analysis were averaged. Sub-samples were employed because larger numbers of individuals would necessitate prohibitively longer runs of the MCMC method to fully explore tree parameter space and would not appreciably increase precision (Jue 2006). Consistency of the obtained results was verified by independently subsampling some population samples multiple times and comparing the 95% confidence intervals of the estimated parameters. The effect of initial parameter values was explored by repeating a subset of analyses with starting values of  $\theta$  equal to 1, 10 and 50. Reported results were obtained setting the starting value of  $\theta$  to 10 for all replicate analyses of all samples with one chain of 20 million visited and 1 million recorded genealogies (burn-in = 50 000). A static heating scheme was employed with temperatures set to 1.0, 1.5, 3.0 and 6.0. Mutation rates were scaled by number of alleles per locus across all samples (following Jue 2006).

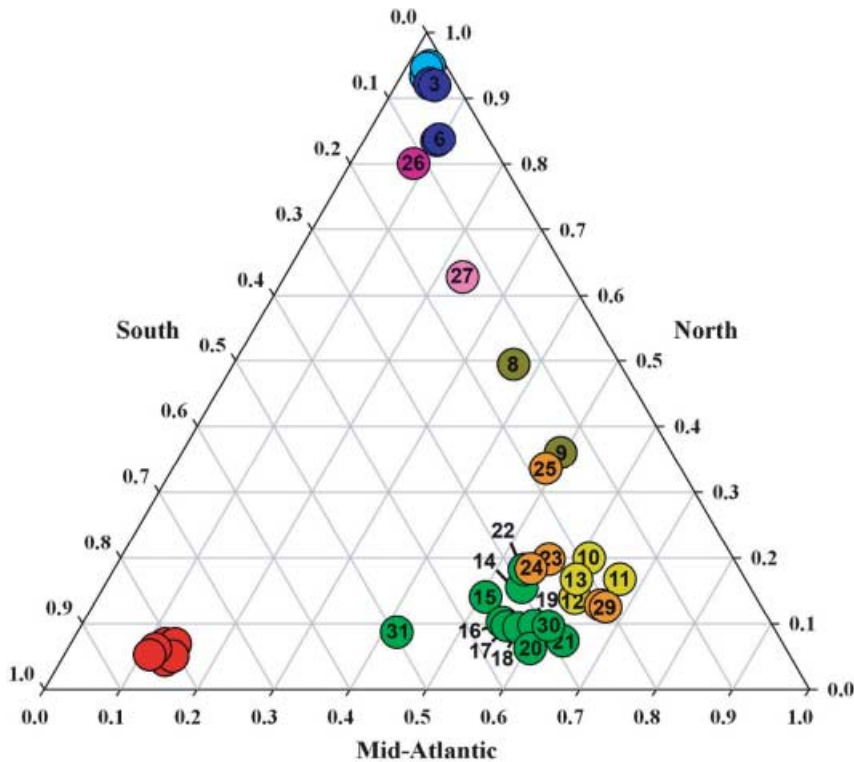
Estimates of  $M$  were obtained for selected pairs of population samples by performing a minimum of two replicate analyses of two independent subsamples of 20 individuals from each population sample (four analyses for each pair of populations). Results of analyses were accepted if the 95% confidence intervals overlapped. Starting values of  $\theta$  were derived from the single-population analyses, and starting values of  $M$  were set to 10.0 for reported results (values of  $M$  equal to 1 and 50 were also explored with similar results). The Bayesian analysis included one chain of 35 million visited and 700 000 recorded genealogies (burn-in = 50 000) with a static heating scheme and scaled mutation rates as described above. Estimates of  $N_e m$  were obtained as  $N_e m_i = 0.25 \times \theta_i \times M_{j \rightarrow i}$ .

Mutation rates were estimated following the protocol of Turner *et al.* (2002) and Jue (2006). This method assumes that mutation rates vary from  $1 \times 10^{-3}$  to  $1 \times 10^{-5}$  based on empirically derived estimates in a variety of species (Jarne & Lagoda 1996; Vigouroux *et al.* 2002). Assuming that  $N_e$  is the same across all loci, differences in  $\theta$  among loci should be directly related to mutation rate ( $\mu$ ). To infer the range of mutation rates, the locus with the largest value of  $\theta$  was assumed to have a mutation rate  $\mu_{\max} = 10^{-3}$ . Mutation rates for other loci were then scaled by the ratio  $\theta_j/\theta_{\max}$ . Similarly, the locus with the smallest value of  $\theta$  was assumed to have a mutation rate  $\mu_{\min} = 10^{-5}$  and mutation rates for other loci were then scaled by the ratio  $\theta_j/\theta_{\min}$ . The midpoint of the interval of harmonic mean estimates of  $\theta_{\min}$  and  $\theta_{\max}$  scaled mutation rates was then applied to MIGRATE-generated values of  $\theta$  and associated 95% confidence intervals.

One caveat to this approach of estimating mutation rates was that the microsatellite loci employed in this study had previously been selected specifically to minimize polymorphism (Adams *et al.* 2006). Therefore, to appropriately capture the full range of mutation rates, we used MIGRATE to estimate  $\theta$  for 14 loci that encompassed the full range of variability among microsatellites in *F. heteroclitus* (data derived from population 33, reported in Adams *et al.* 2005). After scaling mutation rates for all 14 loci, the midpoint of the harmonic mean of mutation rates was calculated using only those loci employed in this study.

## Results

Tests of Hardy-Weinberg equilibrium and linkage disequilibrium were conducted on those samples not previously reported in Adams *et al.* (2006). Following Bonferroni correction, only one of the 184 single-locus tests of Hardy-Weinberg equilibrium exhibited a significant deficit of heterozygotes. This deficit was observed at locus FhATG4 in population FhVC2 ( $F_{IS} = 0.64$ ). None of the remaining populations exhibited a statistically significant elevated  $F_{IS}$  value at this locus, and overall, heterozygote deficiency was not a pervasive problem in any population samples or



**Fig. 2** Ternary plot of population admixture proportions from STRUCTURE analysis with  $k = 3$ . Population numbers correspond to Fig. 1.

at any loci. This is consistent with results for the previously reported samples (Adams *et al.* 2006). No evidence of linkage disequilibrium was observed among loci, consistent with previous reports (Adams *et al.* 2005; 2006). The complete data set, including samples reported in Adams *et al.* (2006), is available as Table S1, Supplementary material.

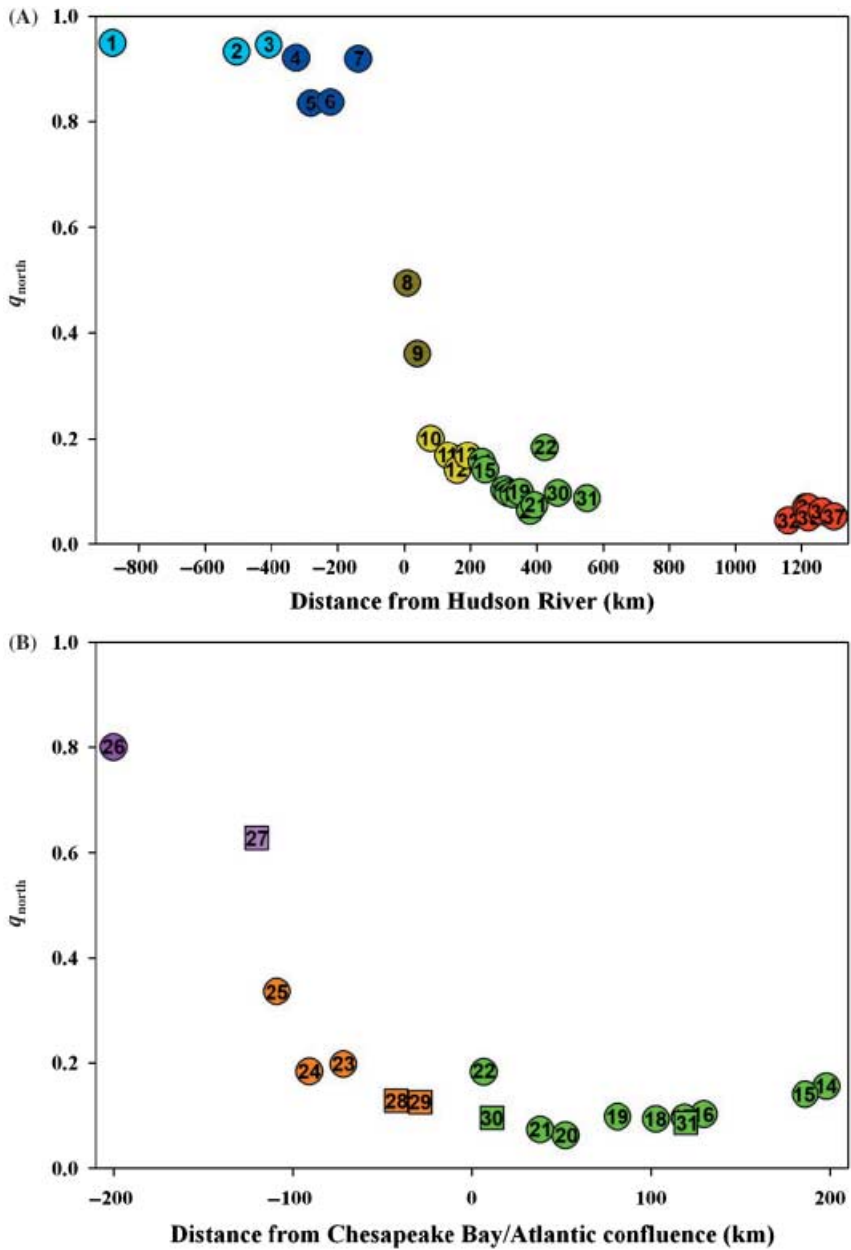
There was no evidence that stepwise mutation contributed to population divergence at any geographical scale, and the SMM estimator of  $R_{ST}$  did not perform better than  $F_{ST}$  in estimating population differentiation. In the global tests,  $R_{ST}$  was not significantly greater than  $\rho R_{ST}$  for any individual locus or over all loci. Accordingly,  $F_{ST}$  was used for all subsequent analyses.

#### *Geographical patterns of genetic variation*

The BAPS test assigned all samples to nine clusters ( $-\ln k(9) = 42\,951$ , with a 0.946 probability compared with all other tested values of  $k$  ranging from 2 to 12). These clusters are colour-coded in Figs 1 and 2. The maximum-likelihood value using STRUCTURE was obtained with  $k = 5$  (range of values tested:  $k = 2$ –12). Overall, clusters obtained using the BAPS and STRUCTURE algorithms were consistent. When the STRUCTURE analysis was performed with  $k = 3$  (Fig. 2), most samples exhibited a strong affinity for either a southern, mid-Atlantic, or northern cluster. Within the mid-Atlantic cluster, there was a distinct separation between New Jersey vs. Delmarva samples. The two northernmost New Jersey

samples (8 and 9) exhibited admixture values intermediate between the northern and mid-Atlantic clusters. Chesapeake Bay samples tended to cluster more closely with New Jersey samples than the geographically more proximate Delmarva coastal samples and the Potomac and James River samples (26 and 27) exhibited a clear association with the northern cluster.

A steep cline in the northern admixture proportion ( $q_{north}$ ) (based on the STRUCTURE analysis with  $k = 3$ ) vs. geographical distance from the Hudson River supported a zone of intergradation in northern New Jersey as previously reported (Adams *et al.* 2006), and placed the two northernmost New Jersey samples (8 and 9) within that intergradation zone (Fig. 3a). Other studies have reported the presence of both northern allozyme alleles (Ropson *et al.* 1990) and mtDNA haplotypes (Smith *et al.* 1998) within the upper reaches of the Chesapeake and Delaware Bays. A second cline in northern admixture proportion, with similar geographical range and slope, was apparent in the Chesapeake Bay region when  $q_{north}$  was plotted against geographical distance from the confluence between the Chesapeake Bay and the Atlantic Coast (Fig. 3b). Among the Delmarva samples (14–22), the northern admixture proportion observed for sample 22 appeared anomalously high (Fig. 3a, b). This result, coupled with the proximity to the mouth of the Chesapeake Bay, suggested that this sample has been influenced by introgression with Chesapeake Bay populations. Therefore, sample 22 was excluded from subsequent analyses that were restricted to coastal Delmarva samples.



**Fig. 3** Clinal relationships of northern admixture proportions from STRUCTURE analysis ( $k = 3$ ) for (A), Atlantic coastal samples measured from the confluence of the Hudson River with the Atlantic Ocean, and B, Chesapeake Bay and Delmarva samples measured from the confluence of the Chesapeake Bay with the Atlantic Ocean. In plot (B), distances measured between the sample site and the north shore of the Chesapeake Bay/Atlantic Ocean confluence shown as circles; distances from south shore shown as squares. Population numbers correspond to Fig. 1.

*Contemporary vs. historical influences on distributions of genetic variation*

Populations of *Fundulus heteroclitus* generally inhabit shallow brackish water coves and tidal creeks where they have ready access to salt marshes (Bigelow & Schroeder 1953; Lotrich 1975). Within these environments, deep (i.e. > 2 m in depth) and wide waters likely provide effective barriers to contemporary migration. In order to partition components of genetic divergence according to contemporary isolation-by-distance vs. long-term restricted movement effects imposed by the presence of dispersal barriers, a series of AMOVA tests and multiple Mantel tests were performed.

Contemporary coastal populations flanking the Delaware Bay and the Chesapeake Bay were posited to be isolated by these respective marine features and were analysed in separate tests.

Both the multiple Mantel test and the AMOVA indicated that genetic variation is partitioned to either side of the Delaware Bay (samples 10–13 vs. 14–21, Table 2). The multiple Mantel test demonstrated that, although much of the genetic variation was accounted for by geographical distance (73% of variation, Fig. 4a), the proportion of variation explained was significantly increased (to 80%) by adding the grouping effect. The AMOVA test supported subdivision across Delaware Bay with a statistically

**Table 2** Distribution of genetic variation among groups of populations separated by bays

	(11–13) (14–21)	(14–21) (30–31)
AMOVA:		
$F_{CT}$	0.014***	-0.001
$F_{SC}$	0.007*	0.009***
$F_{ST}$	0.021***	0.008***
Variance (%):		
Among groups	1.39	-0.11
Among populations	0.73	0.92
Within groups		
Within populations	97.88	99.2
Multiple Mantel tests:		
$r_{geodis}$	0.854***	0.619***
$r_{grouping}$	0.707*	0.057
$r_{geodis\_grouping}$	0.748***	0.762***
$r_{grouping\_geodis}$	0.431*	-0.568**
Variance (%):		
Geodis	72.9	38.3
Group	49.9	0.3
Total	79.9	58.2

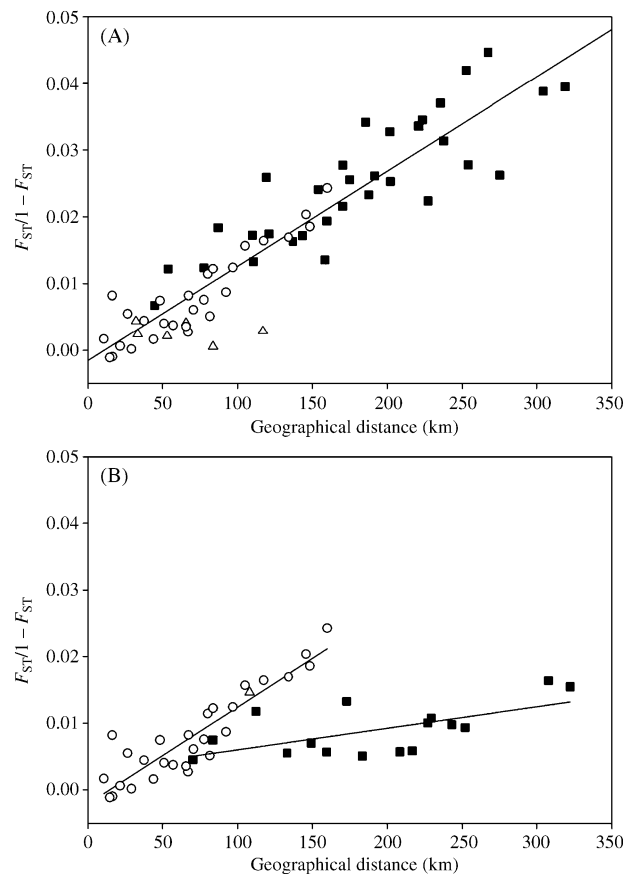
\* $P < 0.01$ ; \*\* $P < 0.001$ ; \*\*\* $P < 0.0001$ .

significant 1.4% of the variation distributed among groups (Table 2).

A multiple Mantel test of samples flanking the Chesapeake Bay (samples 14–21 vs. 30–31, Table 2) also revealed a significant correlation between genetic and geographical distance ( $r_{geodis} = 0.62$ ). The correlation between genetic distance and group was not significant ( $r_{group} = 0.057$ ), but the partial correlation between genetic distance and group controlling for geographical distance was significant and negative ( $r_{group\_geodis} = -0.568$ ), indicating that populations found on opposite sides of the Chesapeake Bay were actually more genetically similar than would have been predicted based on geographical distance alone. This can be seen in the two distinct trend lines for intragroup and intergroup comparisons (Fig. 4b). Similarly, an AMOVA test revealed that no variance was attributable to differences among groups north and south of the Chesapeake Bay (Table 2). In corroboration, the BARRIERS analysis also indicated that Delaware Bay presented a significant barrier to gene flow, but not Chesapeake Bay (data not shown).

#### Effective population size and contemporary gene flow estimates

Per-locus estimates of  $\theta (= 4N_e\mu)$  among 14 loci in population sample 33 ranged more than an order of magnitude from 4.6 to 67.2 between the least and most variable loci, respectively. Among the eight loci employed in this study, estimates of  $\theta$  ranged from 4.6 to 24.7 with an inferred



**Fig. 4** Relationship between genetic distance and geographical distance. (A) Samples 10 through 21. Pairwise comparisons among samples 10–12 shown as triangles, Pairwise comparisons among samples 13–21 shown as circles. Pairwise comparisons between these two groups shown as squares. (B) Samples 13 through 21, 30, 31. Pairwise comparisons among samples 13–21 shown as circles. Comparison between 30 and 31 shown as a triangle. Pairwise comparisons between these two groups shown as squares. Separate trend lines shown for intragroup and intergroup comparisons.

mutation rate of  $8.2 \times 10^{-5}$ . This mutation rate was conservatively assumed to be  $10^{-4}$  in all subsequent analyses to limit underestimation bias.

Among coastal samples, there was a steep several-fold transition in estimates of  $\theta$  (Fig. 5) that closely paralleled the cline in northern admixture proportions (Fig. 3a). Bayesian estimates of median  $\theta$  ranged from 1.4 in the northernmost Nova Scotia sample (sample 1) to 11.9 in sample 35 from Sapello Island, Georgia, representing approximately an 8.5-fold difference in long-term historical effective population size across the range of the species. Additionally, estimates of  $\theta$  were substantially lower in the Chesapeake Bay region relative to adjacent Delmarva coastal samples. These single-population estimates of  $\theta$  were convergent across replicate runs of the analyses,

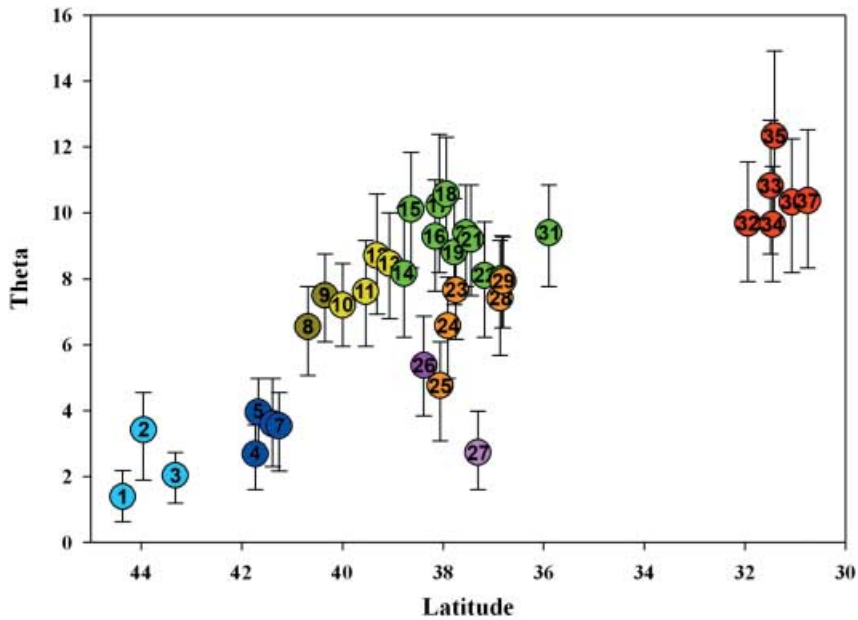


Fig. 5 Bayesian estimates of  $\theta$  and 95% confidence intervals based on single-population analyses with MIGRATE. Numbers correspond to Fig. 1.

including when independent subsamples of each location were assessed, and when starting values of  $\theta$  were varied from 1 to 50.

In order to assess the importance of contemporary gene flow on the distribution of genetic variation among populations, we analysed a subset of populations within the species distribution that may be at or near migration drift equilibrium. We examined two regions that did not appear to exhibit genetic structure based on STRUCTURE and BAPS tests. These were Delmarva samples 13 through 21 and Georgia samples 32 through 37. The Delmarva samples exhibited a highly significant correlation between genetic and geographical distances ( $r = 0.79$ ,  $P < 0.0001$ ), and the Georgia samples also exhibited a statistically significant correlation ( $r = 0.64$ ,  $P = 0.01$ ).

Pairwise population estimates of the effective number of migrants per generation ( $N_e m$ ) inferred from  $F_{ST}$  were very high in both the Delmarva and Georgia regions. The minimum observed  $N_e m$  value was 7.1 among all Delmarva samples and 37.9 among Georgia samples (Table 3). Bayesian estimates of  $N_e m$  derived using MIGRATE ranged from 4.4 to 5.8 among select Delmarva populations and from 5.3 to 7.2 among Georgia populations (Table 4), and were generally an order of magnitude or more lower than traditional estimates based on  $F_{ST}$  for the same population pairs (see Table 3). Single-population estimates of  $\theta$  were in agreement with paired-population estimates, and convergence of estimates of both  $\theta$  and  $M$  were very good across replicate paired-population analyses of the same as well as independent subsamples of each population pair (Table 4). Bayesian estimates of  $M (= m/\mu)$  indicated migration rates exceeding mutation rates by a factor ranging from 1.4 to 3.2 with a mean of 2.4.

## Discussion

If genetic drift and gene flow are in a state of equilibrium, then the influence of contemporary environments on populations (for example, through natural selection leading to local adaptation) may be distinguished from neutral phylogeographical expectations. Although migration-drift equilibrium is often assumed in population genetic studies, few studies verify that populations exist in this state, and few populations are likely to exist in this state because of widespread demographic instability (Whitlock 1992). *Fundulus heteroclitus* is considered a model organism for the study of adaptive clinal variation. However, the evolutionary forces responsible for the distribution of genetic variation among populations, and in particular the relative influences of gene flow and natural selection at varying geographical scales, remains an important and largely unresolved issue. Steep clinal patterns in allozyme allelic and mtDNA haplotype frequencies, both along the Atlantic coast and within the Chesapeake and Delaware Bay regions, have been attributed to spatially variable selection pressures against a background of gene flow in these respective regions at broad geographical scales (Brown & Chapman 1991; Smith *et al.* 1998). However, the relatively recent (post-Pleistocene) historical influences of population expansions and secondary contacts as well as the influences of variable population sizes and the presence of contemporary barriers to migration and gene flow have remained largely discounted in these studies.

In this report, we confront the prevailing assumptions that *F. heteroclitus* populations are in migration-drift equilibrium, and that contemporary processes primarily govern present distributions of genetic variation. Data

**Table 3** Genetic distance between pairs of populations.  $F_{ST}$  is provided below the diagonal and the corresponding estimate of  $N_e m$  is shown above the diagonal.  $F_{ST}$  estimates in bold were significantly greater than zero ( $P < 0.05$ ) based on permutation tests

	11	12	13	14	15	16	17	18	19	20	21	30	31	32	33	34	35	36	37
11 TNJ	—	56.6	60.7	14.5	9.7	9.8	7.3	7.6	7.2	6.0	5.6	7.0	4.8	3.0	2.9	3.3	2.9	2.5	2.4
12 EGGNJ	0.004	—	103.9	20.2	13.6	14.5	10.4	11.6	9.6	7.4	6.8	9.8	6.7	3.8	3.9	4.5	3.7	3.2	3.0
13 SHNJ	<b>0.004</b>	0.002	—	37.1	20.6	18.8	14.4	15.3	18.4	10.8	9.9	11.9	8.9	4.3	4.0	4.5	4.0	3.4	3.2
14 LEWDE	<b>0.017</b>	<b>0.012</b>	0.007	—	NA	40.7	21.9	20.1	15.2	12.3	10.3	23.3	16.2	6.2	6.1	6.5	5.8	4.6	4.5
15 IRBDE	<b>0.025</b>	<b>0.018</b>	<b>0.012</b>	−0.001	—	67.3	30.6	20.4	16.0	14.8	13.5	42.9	15.3	6.8	6.5	6.9	6.8	5.2	4.8
16 SPEMD	<b>0.025</b>	<b>0.017</b>	<b>0.013</b>	<b>0.006</b>	0.004	—	146.8	45.2	33.5	33.1	28.8	44.4	26.9	7.6	7.2	8.1	7.6	6.0	5.3
17 WMD	<b>0.033</b>	<b>0.024</b>	<b>0.017</b>	<b>0.011</b>	<b>0.008</b>	0.002	—	30.6	56.6	92.3	48.8	35.5	25.5	6.8	5.7	6.3	6.4	5.1	4.3
18 CHIVA	<b>0.032</b>	<b>0.021</b>	<b>0.016</b>	<b>0.012</b>	<b>0.012</b>	<b>0.006</b>	<b>0.008</b>	—	416.4	62.3	71.2	45.2	25.0	8.4	7.6	7.4	7.8	5.6	5.4
19 GARVA	<b>0.033</b>	<b>0.026</b>	<b>0.013</b>	<b>0.016</b>	<b>0.015</b>	<b>0.007</b>	0.004	0.001	—	1249.8	156.0	21.3	43.6	7.2	6.1	6.7	7.2	5.4	4.7
20 QUIVA	<b>0.040</b>	<b>0.033</b>	<b>0.023</b>	<b>0.020</b>	<b>0.017</b>	<b>0.008</b>	0.003	0.004	0.000	—	NA	33.5	49.8	8.4	6.3	6.5	7.2	5.0	4.4
21 HOGVA	<b>0.043</b>	<b>0.036</b>	<b>0.025</b>	<b>0.024</b>	<b>0.018</b>	<b>0.009</b>	<b>0.005</b>	0.004	0.002	−0.001	—	55.3	18.8	6.5	4.8	4.8	5.7	4.2	3.8
30 RUDVA	<b>0.035</b>	<b>0.025</b>	<b>0.021</b>	<b>0.011</b>	<b>0.006</b>	<b>0.006</b>	<b>0.007</b>	<b>0.006</b>	<b>0.012</b>	<b>0.007</b>	<b>0.005</b>	—	17.1	6.3	5.4	5.6	6.0	4.1	4.2
31 RINC	<b>0.050</b>	<b>0.036</b>	<b>0.027</b>	<b>0.015</b>	<b>0.016</b>	<b>0.009</b>	<b>0.010</b>	<b>0.010</b>	0.006	0.005	<b>0.013</b>	<b>0.014</b>	—	18.1	11.7	12.6	12.8	8.5	8.1
32 SIGA	<b>0.077</b>	<b>0.062</b>	<b>0.055</b>	<b>0.039</b>	<b>0.035</b>	<b>0.032</b>	<b>0.035</b>	<b>0.029</b>	<b>0.034</b>	<b>0.029</b>	<b>0.037</b>	<b>0.038</b>	<b>0.014</b>	—	108.4	99.8	NA	38.2	46.0
33 SICGA	<b>0.078</b>	<b>0.061</b>	<b>0.059</b>	<b>0.039</b>	<b>0.037</b>	<b>0.034</b>	<b>0.042</b>	<b>0.032</b>	<b>0.039</b>	<b>0.038</b>	<b>0.049</b>	<b>0.044</b>	<b>0.021</b>	0.002	—	NA	624.8	192.1	113.4
34 SIFGA	<b>0.070</b>	<b>0.052</b>	<b>0.052</b>	<b>0.037</b>	<b>0.035</b>	<b>0.030</b>	<b>0.038</b>	<b>0.033</b>	<b>0.036</b>	<b>0.037</b>	<b>0.049</b>	<b>0.043</b>	<b>0.019</b>	0.003	−0.002	—	NA	NA	NA
35 SISGA	<b>0.080</b>	<b>0.063</b>	<b>0.058</b>	<b>0.041</b>	<b>0.036</b>	<b>0.032</b>	<b>0.038</b>	<b>0.031</b>	<b>0.034</b>	<b>0.033</b>	<b>0.042</b>	<b>0.040</b>	<b>0.019</b>	−0.001	0.000	−0.001	—	227.0	77.9
36 JIGA	<b>0.091</b>	<b>0.073</b>	<b>0.069</b>	<b>0.052</b>	<b>0.046</b>	<b>0.040</b>	<b>0.047</b>	<b>0.043</b>	<b>0.044</b>	<b>0.047</b>	<b>0.057</b>	<b>0.057</b>	<b>0.029</b>	<b>0.007</b>	0.001	−0.001	0.001	—	277.5
37 SMGA	<b>0.094</b>	<b>0.076</b>	<b>0.074</b>	<b>0.053</b>	<b>0.050</b>	<b>0.045</b>	<b>0.055</b>	<b>0.044</b>	<b>0.051</b>	<b>0.053</b>	<b>0.061</b>	<b>0.056</b>	<b>0.030</b>	<b>0.005</b>	0.002	−0.001	0.003	0.001	—

**Table 4** Estimates of  $\theta$ ,  $M$ , and  $N_e m$  from Bayesian analyses of genotype genealogies using MIGRATE

Populations* $i, j$	Replicate†	$\theta_i$ (95% CI)	$M_{j \rightarrow i}$ (95% CI)	$N_e m_{j \rightarrow i}$ (95% CI)	$\theta_j$ (95% CI)	$M_{i \rightarrow j}$ (95% CI)	$N_e m_{i \rightarrow j}$ (95% CI)
11, 12	1a	5.4 (4.0, 6.7)	1.3 (0.1, 2.1)	1.8 (0.1, 3.5)	5.9 (4.0, 7.9)	3.7 (2.1, 5.3)	5.5 (2.1, 10.5)
	1b	6.8 (4.6, 8.8)	1.5 (0.5, 2.3)	2.6 (0.6, 5.0)	5.4 (3.3, 7.6)	3.9 (2.7, 5.5)	5.3 (2.2, 10.5)
	2a	8.4 (4.6, 12.5)	3.4 (1.9, 4.7)	7.0 (2.2, 14.7)	9.5 (7.5, 11.7)	0.6 (0.1, 1.3)	1.5 (0.2, 3.8)
	2b	6.2 (4.8, 7.4)	1.6 (0.7, 2.5)	2.5 (0.8, 4.6)	9.0 (6.4, 11.7)	3.9 (1.5, 5.3)	8.8 (2.4, 15.5)
	Mean	6.7	2.0	3.3	7.5	3.0	5.7
Single population‡		7.6 (6.0, 9.2)			8.7 (6.9, 10.6)		Mean $N_e m = 4.5$
16, 17	1a	8.1 (6.5, 12.1)	5.5 (2.7, 9.7)	11.1 (4.4, 29.4)	9.1 (7.5, 10.6)	0.98 (0.1, 1.7)	2.2 (0.2, 4.5)
	1b	7.2 (4.1, 9.6)	1.4 (0.5, 2.1)	2.5 (0.5, 5.0)	7.1 (4.7, 9.6)	3.4 (2.1, 4.5)	6.0 (2.5, 10.8)
	2a	9.1 (6.4, 12.7)	4.3 (3.1, 5.3)	9.9 (4.9, 16.8)	9.7 (7.4, 12.3)	1.5 (0.7, 2.3)	3.7 (1.3, 7.0)
	2b	10.1 (6.8, 13.5)	4.1 (2.5, 5.3)	10.4 (4.2, 17.9)	6.4 (4.6, 8.6)	0.6 (0.1, 1.3)	1.0 (0.1, 2.8)
	Mean	8.6	3.8	8.3	8.1	1.6	3.3
Single population:		9.3 (7.6, 11.0)			10.2 (8.2, 12.4)		Mean $N_e m = 5.8$
17, 18	1a	10.1 (7.4, 13.4)	2.3 (1.4, 4.3)	5.7 (2.2, 14.4)	12.0 (6.8, 17.3)	2.2 (1.3, 3.1)	6.7 (2.2, 13.4)
	1b	8.0 (2.2, 5.8)	3.8 (2.7, 4.7)	7.5 (1.5, 6.8)	7.1 (5.3, 8.9)	1.4 (0.5, 2.1)	2.4 (0.7, 4.7)
	2a	8.2 (6.0, 10.3)	2.8 (1.7, 3.5)	5.7 (2.5, 9.0)	8.4 (5.8, 11.0)	2.0 (1.1, 2.7)	4.1 (1.6, 7.4)
	2b	10.2 (6.4, 13.8)	3.2 (0.7, 4.9)	8.1 (1.1, 16.9)	7.8 (4.6, 10.9)	1.5 (0.5, 2.3)	2.8 (0.6, 6.2)
	Mean	9.1	3.0	6.9	8.8	1.8	3.9
Single population:		10.2 (8.2, 12.4)			10.6 (8.9, 12.3)		Mean $N_e m = 5.4$
18, 19	1a	8.8 (6.2, 11.6)	1.4 (0.3, 2.1)	3.0 (0.5, 6.1)	9.5 (5.8, 12.1)	2.5 (1.1, 4.1)	5.9 (1.6, 12.4)
	1b	7.8 (5.7, 9.9)	1.2 (0.1, 1.9)	2.4 (0.1, 4.7)	9.7 (6.5, 12.3)	2.9 (0.9, 4.9)	7.0 (1.5, 15.0)
	2a	9.5 (7.5, 11.4)	1.1 (0.1, 1.7)	2.6 (0.2, 4.8)	7.6 (5.7, 9.6)	6.3 (5.1, 7.5)	12.0 (7.2, 18.0)
	2b	7.4 (5.4, 10.0)	4.6 (3.3, 5.9)	8.6 (4.4, 14.8)	7.9 (5.7, 9.9)	1.0 (0.1, 1.7)	1.9 (0.1, 4.2)
	Mean	8.4	2.1	4.3	8.7	3.2	6.9
Single population:		10.6 (8.9, 12.3)			8.8 (7.2, 10.4)		Mean $N_e m = 5.6$
19, 20	1a	6.7 (5.0, 8.6)	3.9 (3.7, 5.7)	6.5 (4.6, 12.3)	5.6 (2.7, 8.2)	1.5 (0.5, 2.5)	2.1 (0.3, 5.1)
	1b	8.9 (6.2, 11.4)	0.9 (0.1, 1.7)	2.1 (0.2, 4.8)	9.3 (6.1, 12.1)	2.3 (1.1, 3.1)	5.3 (1.7, 9.4)
	2a	7.5 (5.8, 9.0)	1.6 (0.3, 2.7)	3.0 (0.4, 6.1)	7.8 (5.0, 10.4)	2.2 (0.7, 3.5)	4.2 (0.9, 9.1)
	2b	6.5 (5.0, 7.8)	1.6 (0.5, 2.3)	2.6 (0.6, 4.5)	11.0 (6.9, 14.8)	3.8 (2.7, 4.7)	10.4 (4.7, 17.4)
	Mean	7.4	2.0	3.7	8.4	2.4	5.1
Single population:		8.8 (7.2, 10.4)			9.4 (7.8, 10.9)		Mean $N_e m = 4.4$
33, 32	1a	9.5 (7.4, 11.8)	1.1 (0.1, 1.9)	2.6 (0.2, 5.6)	10.0 (7.1, 12.8)	5.5 (4.1, 6.7)	13.6 (7.2, 21.5)
	1b	13.5 (9.3, 17.6)	4.1 (2.5, 5.3)	13.8 (5.8, 23.3)	7.5 (5.7, 9.5)	2.5 (1.3, 3.5)	4.7 (1.8, 8.3)
	2a	6.2 (3.3, 9.6)	1.4 (0.1, 3.1)	2.2 (0.1, 7.4)	13.7 (10.9, 16.6)	0.9 (0.1, 1.5)	3.1 (0.3, 6.2)
	2b	12.6 (7.8, 15.8)	1.1 (0.1, 1.7)	3.4 (0.2, 6.7)	6.9 (4.8, 8.9)	2.5 (1.3, 3.3)	4.2 (1.6, 7.3)
	Mean	10.5	1.9	5.0	9.5	2.8	6.8
Single population:		10.8 (8.8, 12.8)			9.7 (7.9, 11.6)		Mean $N_e m = 5.9$
33, 34	1a	10.8 (6.7, 15.2)	0.8 (0.1, 1.5)	2.0 (0.2, 5.7)	12.4 (8.2, 16.2)	3.1 (1.5, 4.7)	9.6 (3.1, 19.0)
	1b	10.5 (7.2, 14.4)	4.9 (3.3, 6.5)	12.9 (5.9, 23.3)	8.3 (6.5, 10.0)	0.6 (0.1, 1.3)	1.2 (0.2, 3.3)
	2a	9.2 (7.2, 11.1)	1.2 (0.3, 1.9)	2.7 (0.5, 5.3)	3.7 (2.3, 5.0)	3.6 (2.5, 4.7)	3.3 (1.4, 5.8)
	2b	5.9 (3.7, 7.9)	4.3 (2.7, 5.3)	6.2 (2.5, 10.5)	9.5 (7.5, 11.3)	0.8 (0.1, 1.5)	2.0 (0.2, 4.2)
	Mean	9.2	2.8	6.3	8.5	2.0	4.3
Single population:		10.8 (8.8, 12.8)			9.6 (7.9, 11.4)		Mean $N_e m = 5.3$
33, 35	1a	8.6 (6.5, 12.1)	1.2 (0.3, 1.9)	2.6 (0.5, 5.8)	13.7 (8.2, 19.0)	3.9 (2.7, 4.9)	13.3 (5.5, 23.2)
	1b	7.5 (5.7, 9.2)	1.7 (0.7, 2.3)	3.2 (1.0, 5.3)	15.2 (11.8, 18.8)	2.9 (2.1, 5.5)	11.1 (6.2, 25.9)
	2a	12.8 (9.2, 17.0)	4.6 (3.5, 5.5)	14.7 (8.0, 23.4)	11.2 (8.6, 13.9)	1.0 (0.1, 1.7)	2.9 (0.2, 5.9)
	2b	20.0 (16.5, 23.6)	2.1 (0.5, 3.5)	10.3 (2.1, 20.6)	9.6 (7.8, 11.3)	1.1 (0.1, 1.7)	2.5 (0.2, 4.8)
	Mean	12.2	2.4	7.3	12.5	2.2	6.9
Single population:		10.8 (8.8, 12.8)			12.3 (9.7, 14.9)		Mean $N_e m = 7.2$
34, 35	1a	7.8 (6.0, 9.7)	1.8 (0.7, 2.9)	3.6 (1.0, 7.1)	10.6 (8.2, 13.1)	1.4 (0.3, 2.3)	3.6 (0.6, 7.5)
	1b	8.3 (6.5, 10.0)	0.6 (0.1, 1.3)	1.2 (0.2, 3.3)	10.5 (7.2, 14.3)	4.9 (3.3, 6.5)	12.9 (5.9, 23.3)
	2a	12.4 (10.1, 14.6)	0.6 (0.1, 1.1)	1.8 (0.3, 4.0)	15.7 (11.7, 19.8)	4.3 (1.9, 6.9)	17.1 (5.6, 34.2)
	2b	4.5 (2.9, 6.4)	2.5 (0.7, 4.1)	2.9 (0.5, 6.5)	9.1 (5.0, 12.5)	2.2 (1.1, 3.3)	2.9 (0.5, 6.5)
	Mean	8.3	1.4	2.9	11.5	3.2	9.2
Single population:		9.6 (7.9, 11.4)			12.3 (9.7, 14.9)		Mean $N_e m = 6.1$

\*Population samples numbered according to Table 1; †two independent replicate analyses (a,b) of each of two independent subsamples (1,2) of each pair of populations; ‡estimates of  $\theta$  (95% CI) when population samples were analysed separately.

presented here indicate that secondary contact is responsible for nonequilibrium conditions across large regional scales, and that apparent isolation by distance across certain formidable dispersal barriers may not be consistent with contemporary migration-drift equilibrium. Historical contingencies rather than contemporary processes appear to primarily account for partitioning of genetic variation within *F. heteroclitus*.

#### *Disequilibrium conditions associated with secondary contact events*

The Pleistocene distribution of populations of *F. heteroclitus* remains largely unsettled. However, the presence of steep clines in microsatellite allele frequencies at eight presumptively neutral loci along the Atlantic coast (data from this study combined with data previously reported by Adams *et al.* (2006) and within the Chesapeake Bay system (data from this study) provide strong evidence against widespread migration-drift equilibrium on a large geographical scale, and argue in favour of secondary contact between Pleistocene refugial northern and southern populations of the species. The secondary contact zone along the Atlantic coast in northern New Jersey argues in favour of northern and southern Pleistocene refugia during the most recent glacial maximum (Adams *et al.* 2006) while the presence of populations with northern genetic affinity in the Chesapeake Bay system (and likely the Delaware Bay as well; see Smith *et al.* 1998), supports the argument that northern populations were more southerly distributed some time during the Pleistocene period (Ropson *et al.* 1990). Reconciliation of these two contact zones likely requires multiple expansions and contractions of these populations during this period.

Using the diffusion approximation equation for cline width (Barton & Gale 1993), Adams *et al.* (2006) estimated the time for the initiation of secondary contact between northern and mid-Atlantic populations at approximately 15 000 years ago (assuming one generation per year). This value is in good agreement with the last glacial retreat. In this study, the width and shape of the Chesapeake Bay cline (Fig. 3b) was comparable to the more northern coastal contact zone (Fig. 3a), suggesting a similar time frame. However, a similar mathematical assessment of the Chesapeake Bay cline was not attempted because the size and shape of the Chesapeake Bay has changed dramatically over the same time period as the post-Pleistocene sea level has increased, inundating the tributaries of this bay region, making an assessment of the historically relevant geographical distances among population samples difficult to infer. Indeed, the Delmarva peninsula has been shifting since the Early Pleistocene and throughout the Quaternary period, during which time the Chesapeake Bay was also periodically open to the North Carolina sounds, and the

major channels excavated by the Potomac and Susquehanna Rivers have shifted north and south during sea-level regressions (Hobbs 2004).

#### *Atlantic coastal populations and patterns of isolation-by-distance*

Although populations near New Jersey and Chesapeake Bay are not in migration-drift equilibrium because of secondary contact, the patterns of IBD apparent among coastal mid-Atlantic as well as southern populations of *F. heteroclitus* were consistent with a condition of migration-drift equilibrium throughout these regions. Furthermore, low  $F_{ST}$  values were consistent with high estimates of  $N_e m$  and corresponding gene flow rates among populations. However, the presence of a pattern of IBD is not definitive evidence of migration-drift equilibrium. For example, if these populations were in migration-drift equilibrium, then one would predict that a pattern of IBD would be disrupted by any barriers to dispersal. While there was evidence that the Delaware Bay represented a subtle barrier to gene flow, there was no evidence that the presumably formidable mouth of the Chesapeake Bay presented a dispersal barrier between north and south. In order to examine why this may be the case, a more nuanced exploration of the process of population convergence to an equilibrium state is merited.

The time required for populations to reach migration-drift equilibrium is a function of both migration rate ( $m$ ) and population size ( $N$ ). While high values of  $N_e m$  will prevent populations in migration-drift equilibrium from diverging, they do not ensure rapid convergence upon equilibrium conditions. The time required for  $F_{ST}$  to reach halfway from an old to a new equilibrium value is  $\ln(1/2)/\ln[(1-m)^2(1-1/2N)]$ , and is substantially longer with lower  $m$  and higher  $N$  (Whitlock 1992; Whitlock & McCauley 1999).

The effective population size of *F. heteroclitus* varied substantially across its range. Distribution-wide estimates of  $\theta$  (1.4–11.9) correspond to historical effective population size estimates, ranging from approximately 3500 in Nova Scotia to 30 000 in Georgia (assuming  $\mu = 10^{-4}$ ). In the mid-Atlantic region, estimates of  $\theta$  were on the order of 8–10.5 (Fig. 5, Table 4) corresponding to effective population sizes of 20 000–25 000 individuals. Bayesian estimates of  $M (= m/\mu)$  for population samples ranging from 10 to 30 km apart in this region averaged about 2.5, corresponding to  $m$  on the order of 0.00025. Under such conditions of large population size and low migration rate, the half-life of an equilibrium shift in  $F_{ST}$  would be approximately 1300 generations. Across the Chesapeake Bay, where  $m$  is assumed to be 0, the half-life of an equilibrium shift would increase substantially to about 28 000 generations (assuming  $N_e = 20 000$ ). Based on estimates of  $\theta$  in populations flanking this presumed barrier (21 and 30), the time required for these two populations to

diverge in the absence of migration from a current value of  $F_{ST} = 0.0045$  (Table 3) to the level of divergence observed between populations flanking Delaware Bay ( $F_{ST} = 0.007$ , Table 3) would be approximately 18 000 generations, or the length of time since the last glacial maximum (assuming one generation per year).

Although high effective population size and low migration rates may explain why formidable dispersal barriers are not reflected in the distribution of genetic variation among populations flanking the Chesapeake Bay, this does not explain why populations south of the bay were actually more similar to populations north of the bay than would have been predicted by geographical distance alone. A possible explanation for this pattern lies in the fact that coastal Delmarva populations are situated between two independent contact zones corresponding to each of the two bays. Although samples from the upper reach of Delaware Bay were not included in this study, the presence of northern mtDNA variants in this bay region is known from other studies (Smith *et al.* 1998). The slightly elevated northern admixture proportions ( $q_{north}$ , Fig. 3a, b) in populations 14 and 15 (northernmost Delmarva populations) as well as 22 (southernmost Delmarva population) suggest the influence of introgression from populations in the bay environments. Consequently, the overall genetic similarity of populations flanking both sides of the Chesapeake Bay (i.e. 20, 21, 30, 31), contrasted by their equivalent level of divergence from populations adjacent to the south shore of Delaware Bay (i.e. 14, 15), implies that genetic relationships among these coastal populations reflects introgression from bay populations. Although the effects appear subtle, these results in combination call into question the notion that populations in the mid-Atlantic region (populations 8–31) are in migration-drift equilibrium beyond the local population level ( $\sim < 30$  km).

An important outcome of the Bayesian analysis with MIGRATE was that estimates of  $\theta$ , and therefore  $N_e$ , were exceptionally large and estimates of  $M$ , and therefore  $m$ , were exceptionally small. Estimates of the effective number of migrants,  $N_e m$ , based on the variance in allele frequencies (minimum within-region value = 10.3, Table 3) as well as Bayesian estimates from genealogies (range of estimated values = 4.4–7.2, Table 4) did not differ substantially between mid-Atlantic and Georgia populations (where migration-drift equilibrium was inferred to be reasonable) and were sufficient to counteract drift and prevent population divergence. However, estimates derived by each of the respective methods mostly differed by an order of magnitude or more. The likely explanation for this discrepancy is the recognized fact that estimates of  $N_e m$  based on  $F_{ST}$  (allele frequencies) are problematic. First, the relationship between  $F_{ST}$  and  $N_e m$  is not linear and small values of  $F_{ST}$  yield estimates of  $N_e m$  with high variance (Whitlock & McCauley 1999), thereby making estimates of  $N_e m$  greater than  $\sim 5$  difficult

to interpret. Additionally, the assumption that mutation rates are negligible (i.e.  $\mu \ll m$ ) may be violated in this system, as Bayesian estimates of  $M$  suggested that migration is only two or three times that of mutation. Homoplasy resulting from high mutation rates would result in a systematic underestimate of  $F_{ST}$  and overestimate of  $N_e m$  (Jarne & Lagoda 1996).

#### Comparisons to empirically derived field estimates

What do genetic model-based estimates of population size and migration rate tell us about the forces that affect contemporary populations? Population genetic model-based estimates of migration rate and population size are often at odds with estimates derived empirically in the field. Genetic model-based estimates of gene flow may differ substantially from empirically derived estimates because of differences in reproductive success between migrant and resident individuals, and because of the difficulty in empirically measuring rare but potentially important long-distance dispersal events (Koenig *et al.* 1996). Similarly, empirical estimates of population size are usually much larger than genetic model-based estimates because they do not account for the effects of factors such as historical population size fluctuation, sex-bias, and variance in reproductive success (Frankham 1995). On the other hand, estimates of effective population size and migration rates derived from genetic models of the coalescent process provide only long-term evolutionary averages of these population parameter values, and therefore may not address short-term ecological events that impact natural populations.

Empirically derived estimates of neighbourhood size and population size have been determined through mark-recapture studies of *F. heteroclitus* populations. How do these estimates compare to genetic model-based estimates in this study? Effective population size estimates exhibited unusually close agreement with mark-recapture estimates of census population size. Sweeney *et al.* (1998) reported average neighbourhood size estimates of about 26 000 individuals in creeks on Plum Island in Massachusetts. In this study, the estimate of  $\theta$  from Barnstable, Massachusetts (sample 4,  $\sim 20$  km from Plum Island) was 2.7, corresponding to a long-term effective population size of 6800 individuals. Therefore, for this location the ratio  $N_e/N$  was about 0.26, at the upper end of the distribution of such values among animals where the mean ratio is close to 0.11 (Frankham 1995).

Estimates of  $N_e m$  appeared also to be in concordance with neighbourhood sizes predicted based on empirical mark-recapture estimates in the same study by Sweeney *et al.* (1998). According to the stepping-stone model, the relationship between neighbourhood size ( $N_b$ ) and  $N_e m$  is  $N_b = 4\pi D\sigma^2 = 2\pi N_e m$  where  $D = N/\epsilon^2$  and  $\epsilon$  is the spacing between demes (Slatkin & Barton 1989). Solving for  $N_e m$  for a geographical distance of 30 km (30 000 m), assuming a

value of  $N$  of 6800 and a mean home range ( $\sigma$ ) of 650 m (Sweeney *et al.* 1998), yields an estimate of  $N_m = 6.3$ . The ninety-five percent confidence limits among Bayesian estimates of  $N_m$  in this study ranged from 1 to 12 (Table 4). Therefore, remarkable agreement was obtained for estimates of genetic exchange based on indirect genetic estimates in this study and empirically derived field-based estimates (Sweeney *et al.* 1998).

## Conclusions

A common assumption of population genetic studies is that populations are in migration-drift equilibrium, a condition that is likely rarely met in natural populations because of widespread environmental and demographic instability (Whitlock 1992). The assumption of migration-drift equilibrium is rarely verified in population genetic studies. This assumption is important for population genetic models that infer isolation by distance or that may implicate the influence of natural selection, and is therefore of particular relevance to the estuarine teleost fish *Fundulus heteroclitus*, which is a model species for the study of adaptive clinal variation.

Environmental clines have long provided important contexts for studying adaptive variation in nature (Endler 1986), especially for *F. heteroclitus*, although defining the genetic underpinnings of such variation is often elusive. The presence of a cline at any particular locus may be insufficient to invoke natural selection because of nonequilibrium conditions and processes such as secondary contact and introgression that imprint clinal variation at neutral loci as well. That is, population history can confound interpretation of the adaptive significance of clinal variation. Buried within all of this neutral clinal variation in species such as *F. heteroclitus* also lies adaptive variation (for review see Burnett *et al.* in press). Consideration of the phylogeographical history of a species can often lend greater scope for inference when seeking to distinguish the genetic targets of natural selection (for examples, see Pierce & Crawford 1997; and Whitehead & Crawford 2006).

## Acknowledgements

We would like to thank D. Crawford, M. Oleksiak, R. Kneib, T. Repicky, B. Rees and D. Hanks for assistance in field collection and sending DNA samples. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation. Funding for this project was provided by National Science Foundation grant OCE-0221879 to D.D.D., and National Science Foundation grant CBET-0652006 to A.W.

## References

Able KW, Felley JD (1986) Geographical variation in *Fundulus heteroclitus*: tests for concordance between egg and adult morphologies. *American Zoologist*, **26**, 145–157.

- Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **15**, 1109–1123.
- Adams SM, Oleksiak MF, Duvernell DD (2005) Microsatellite primers for the Atlantic coastal killifish, *Fundulus heteroclitus*, with applicability to related *Fundulus* species. *Molecular Ecology Notes*, **5**, 275–277.
- Austin JD, Loughheed SC, Boag PT (2004) Controlling for the effects of history and nonequilibrium conditions in gene flow estimates in northern bullfrog (*Rana catesbeiana*) populations. *Genetics*, **168**, 1491–1506.
- Barton N, Gale K (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison R), pp. 13–45. Oxford University Press, Oxford, UK.
- Berli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.
- Berli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.
- Berli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in  $n$  subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.
- Bigelow HB, Schroeder WC (1953) Fishes of the Gulf of Maine. Fishery Bulletin 74. *Fishery Bulletin of the Fish and Wildlife Service*, Vol. 53. United States Government Printing Office, Washington.
- Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.
- Bossart JL, Prowell DP (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution*, **13**, 202–206.
- Brown BL, Chapman RW (1991) Gene flow and mitochondrial DNA variation in the killifish, *Fundulus heteroclitus*. *Evolution*, **45**, 1147–1161.
- Burnett KG, Bain LJ, Baldwin WS *et al.* (2007) *Fundulus* as the premier teleost model in environmental biology: opportunities for new insights using genomics. *Comparative Biochemistry and Physiology. Part D: Genomics and Proteomics*, **2**, 257–286.
- Cohen S (2002) Strong positive selection and habitat-specific amino acid substitution patterns in *Mhc* from an estuarine fish under intense pollution stress. *Molecular Biology and Evolution*, **19**, 1870–1880.
- Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.
- Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, **15**, 49–62.
- Eisler R (1986) Use of *Fundulus heteroclitus*. Pollution studies. *American Zoologist*, **26**, 283–288.
- Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.
- Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Frankham R (1995) Effective population size / adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.

- Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513–1520.
- Gonzalez-Villasenor LI, Powers DA (1990) Mitochondrial DNA restriction site polymorphisms in the teleost *Fundulus heteroclitus* support secondary intergradation. *Evolution*, **44**, 27–37.
- Goudet J (1995) FSTAT version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Hahn ME, Karchner SI, Franks DG, Merson RR (2004) Aryl hydrocarbon receptor polymorphisms and dioxin resistance in Atlantic killifish (*Fundulus heteroclitus*). *Pharmacogenetics*, **14**, 131–143.
- Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.
- Hardy OJ, Vekemans X (2002) SPAGED1: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.
- Hobbs CH (2004) Geological history of Chesapeake Bay, USA. *Quaternary Science Reviews*, **23**, 641–661.
- Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.
- Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, **11**, 424–429.
- Jue NK (2006) Exploring the structure of genetic variation and the influences of demography on effective population size in the gag grouper *Mycteroperca microlepi* (Goode & Bean). *Journal of Fish Biology*, **69**, 217–224.
- Kimura M, Weiss GH (1964) Stepping-stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.
- Kneib RT (1997) The role of tidal marshes in the ecology of estuarine nekton. *Oceanography and Marine Biology: An Annual Review*, **35**, 163–220.
- Koenig WD, VanVuren D, Hooge PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology and Evolution*, **11**, 514–517.
- Lotrich VA (1975) Summer home range and movements of *Fundulus heteroclitus* (Pisces–Cyprinodontidae) in a tidal creek. *Ecology*, **56**, 191–198.
- Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.
- Mantel N (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- McMillan AM, Bagley MJ, Jackson SA, Nacci DE (2006) Genetic diversity and structure of an estuarine fish (*Fundulus heteroclitus*) indigenous to sites associated with a highly contaminated urban harbor. *Ecotoxicology*, **15**, 539–548.
- Meyer JN, Di Giulio RT (2003) Heritable adaptation and fitness costs in killifish (*Fundulus heteroclitus*) inhabiting a polluted estuary. *Ecological Applications*, **13**, 490–503.
- Pierce VA, Crawford DL (1997) Phylogenetic analysis of glycolytic enzyme expression. *Science*, **276**, 256–259.
- Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution*, **55**, 131–146.
- Poissant J, Knight TW, Ferguson MM (2005) Nonequilibrium conditions following landscape rearrangement: the relative contribution of past and current hydrological landscapes on the genetic structure of a stream-dwelling fish. *Molecular Ecology*, **14**, 1321–1331.
- Powell WH, Bright R, Bello SM, Hahn ME (2000) Developmental and tissue-specific expression of AHR1, AHR2, and ARNT2 in dioxin-sensitive and -resistant populations of the marine fish *Fundulus heteroclitus*. *Toxicological Sciences*, **57**, 229–239.
- Powers DA, Lauerman T, Crawford D, Dimichele L (1991) Genetic mechanisms for adapting to a changing environment. *Annual Review of Genetics*, **25**, 629–659.
- Powers DA, Ropson I, Brown DC *et al.* (1986) Genetic variation in *Fundulus heteroclitus*: geographic distribution. *American Zoologist*, **26**, 131–144.
- Powers DA, Schulte PM (1998) Evolutionary adaptations of gene structure and expression in natural populations in relation to a changing environment: a multidisciplinary approach to address the million-year saga of a small fish. *Journal of Experimental Zoology*, **282**, 71–94.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rey O, Turgeon J (2007) Influence of contemporary and historical factors on the genetic structure of *Fundulus diaphanus* (Cyprinodontidae) a long the St Lawrence River (Quebec, Canada). *Canadian Journal of Zoology*, **85**, 891–901.
- Ropson IJ, Brown DC, Powers DA (1990) Biochemical genetics of *Fundulus heteroclitus* (L.) 6. Geographical variation in the gene frequencies of 15 loci. *Evolution*, **44**, 16–26.
- Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**, 1349–1368.
- 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. *Molecular Marine Biology and Biotechnology*, **7**, 79–87.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.
- Sweeney J, Deegan L, Garritt R (1998) Population size and site fidelity of *Fundulus heteroclitus* in a macrotidal saltmarsh creek. *Biology Bulletin*, **195**, 238–239.
- Telles MP, Diniz-Filho JA (2005) Multiple Mantel tests and isolation-by-distance, taking into account long-term historical divergence. *Genetics and Molecular Research*, **4**, 742–748.
- Turgeon J, Bernatchez L (2001) Clinal variation at microsatellite loci reveals historical secondary intergradation between glacial races of *Coregonus artedii* (Teleostei: Coregoninae). *Evolution*, **55**, 2274–2286.
- Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics*, **162**, 1329–1339.

- Vigouroux Y, Jaqueth JS, Matsuoka Y *et al.* (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, **19**, 1251–1260.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences USA*, **103**, 5425–5430.
- Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, **46**, 608–615.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration:  $F_{ST}$  not equal  $1/(4Nm + 1)$ . *Heredity*, **82**, 117–125.
- Wright S (1969) *Evolution and Genetics of Populations, Vol. 2. The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

---

David Duvernell is an Associate Professor in the Biological Sciences at Southern Illinois University, Edwardsville (SIUE) with an interest in fish ecological and evolutionary genetics and genome evolution. Andrew Whitehead is an Assistant Professor in the Department of Biological Sciences at Louisiana State University (LSU), with interests in ecological and evolutionary functional genomics, ecophysiology, and genome evolution. James Lindmeier is a master's student at SIUE who participated in this study during the course of completing his thesis. Katherine Faust is a graduating senior at LSU and is currently applying to graduate schools to pursue a PhD in Evolutionary Biology.

---

## Supplementary material

The following supplementary material is available for this article:

**Table S1** Genotypic data at eight tri-nucleotide microsatellite repeat loci for 37 population samples of *Fundulus heteroclitus*

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03648.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.