Assessment of the Cape Cod Phylogeographic Break Using the Bamboo Worm *Clymenella torquata* Reveals the Role of Regional Water Masses in Dispersal

ROBERT M. JENNINGS, TIMOTHY M. SHANK, LAUREN S. MULLINEAUX, AND KENNETH M. HALANYCH

From the Biology Department, University of Massachusetts at Boston, 100 Morrissey Boulevard, Boston, MA 02125 (Jennings); the Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 (Shank and Mullineaux); and the Department of Biological Sciences, Auburn University, Auburn, AL 36849 (Halanych).

Address correspondence to Robert M. Jennings at the address above, or e-mail: rob.jennings@umb.edu.

Abstract

Previous genetic studies suggest Cape Cod, MA, as a phylogenetic break for benthic marine invertebrates; however, diffuse sampling in this area has hindered fine-scale determination of the break's location and underlying causes. Furthermore, some species exhibit breaks in different places, and others exhibit no breaks in this region. We analyze the phylogeographic patterns of 2 mitochondrial genes from 10 populations of the bamboo worm *Clymenella torquata* (Annelida: Maldanidae) focused around Cape Cod but extending from the Bay of Fundy, Canada, to New Jersey. A common invertebrate along the US coast, *C. torquata*, possesses life-history characteristics that should make it sensitive to factors such as dispersal barriers, bottlenecks, and founder events. As an inhabitant of soft sediments, *C. torquata* offers a unique contrast to existing research dominated by organisms dwelling on hard substrates. Our genetic data show a clear phylogenetic break and a cline of haplotype frequencies from north to south. Fine-scale sampling of populations on Cape Cod, combined with other sampled populations, confirm that this distinct break is not on the Cape Cod peninsula itself but to the south near a boundary of oceanic water masses. Low levels of gene flow occur in these populations, in an asymmetric manner congruent with coastal current patterns. No significant effect of Pleistocene glaciation was seen in the pattern of genetic diversity over the sampled range.

Key words: asymmetric migratin, atp6, maldanid, nad4

Phylogeographic studies of coastal invertebrate species in the northwestern Atlantic Ocean have revealed barriers to gene flow near Cape Cod, MA (reviewed in Wares 2002). Species with a broad range of life-history characteristics, from strongly dispersive (the polychaete *Marenzellaria viridis*, Bastrop et al. 1998) to weakly dispersive (the amphipod *Idotea baltica*, Wares and Cunningham 2001), show genetic discontinuities in this region. In general, the sharper genetic breaks tend to occur in species possessing weakly or non-dispersive larvae, whereas species with highly dispersive larvae show no breaks, as in the ocean quahog *Arctica islandica* (Dahlgren et al. 2000), the surf clam *Spisula solidissima solidissima* (Hare and Weinberg 2005), and the slipper snail *Crepidula fornicata* (Collin 2001). However, these patterns are by no means universal; even though the clam

Mercenaria mercenaria has planktonic larvae, Dillon and Manzi (1992) detected a transition in gene frequencies in this region, which indicated reduced gene flow. Similarly, a genetic break was observed in populations of the hermit crab *Pagurus longicarpus*, even though it produces larvae that are planktonic for 2–3 weeks (Young et al. 2002). Although the lack of a consistent pattern is hardly surprising, on the whole, the frequent observation of a break near Cape Cod suggests a pervasive environmental or geographic restriction to gene flow that only highly dispersive organisms can overcome.

Even the exact location of a genetic break in this region is subject to debate. The gene flow barrier may not be the Cape peninsula itself but may be displaced to the south (Wares 2002); however, this question is difficult to answer

from currently available data. Much of this work has focused on larger scale connectivity, and sampling around Cape Cod has typically been diffuse, hindering fine-scale determination of the location of breaks. Samples have often included only a single population representing Cape Cod, with the next sample to the south as far away as Chesapeake Bay or North Carolina. Because the phenomenon in question is a change (extreme or subtle) in genetic composition between one population and the next, the spacing of the neighboring populations limits precision in locating the break.

With this caveat in mind, several species do exhibit breaks between Cape Cod and the next site sampled to the south, rather than on the peninsula itself (e.g., Franz et al. 1981; Dillon and Manzi 1992; Vogler and DeSalle 1993; Bastrop et al. 1998; Lee 1999). Moreover, Kelly et al. (2006) observed significantly different genetic frequencies in the brooding amphipod *Gammarus tigrinus* among populations in 1) the Gulf of Maine (New Brunswick, Maine, and New Hampshire), 2) Rhode Island, and 3) the middle Atlantic (New York, Delaware, and Virginia), indicating that genetic structure may be more complex than implied by the larger scale phylogeographies. Similarly, Caudill and Bucklin (2004) found widely differing haplotype frequencies in populations of the copepod *Acartia tonsa* in New Hampshire; New Bedford, MA (south of Cape Cod); and Rhode Island.

Wares (2002) reviewed 3 possible explanations for the presence of genetic breaks in the vicinity of Cape Cod: 1) differences in water mass characteristics north and south of the break prevent haplotypes from mixing, 2) hydrographic currents along the coast of North America control the direction and extent of gene flow, and 3) the genetic differentiation observed today results from historical dynamics during and after the last glacial maximum (LGM), some 20 000 years ago (Pielou 1991).

Temperature and salinity variation between water masses are known to effect organismal distributions (Hutchins 1947; Valentine 1966) and can effect genetic differentiation as well (Koehn et al. 1976; Gardner and Palmer 1998). In the Northwest Atlantic, the correlation of water temperature and salinity differences with boundaries in species ranges led to the definition (Hutchins 1947; Hayden and Dolan 1976; Engle and Summers 1999) of Cape Cod's southern coast as the boundary between the Acadian coastal biogeographic province to the north (extending to the northeast corner of Maine) and the Virginian province to the south (extending to the southern edge of Virginia). Waters in the Acadian province are uniformly colder than those of the more variable Virginian, and differences in average salinity are also known (Hayden and Dolan 1976; Engle and Summers 1999). Some species' ranges do cross the Acadian/Virginian boundary, but this line seems to mark the northern limit of species' ranges more frequently than it marks species' southern limits (Engle and Summers 1999), suggesting that many species from the Virginian are unable to persist in the colder, saltier Acadian waters. Greater seasonal variation in the Virginian may also prohibit some species adapted to the less variable Acadian from extending southward.

Coastal currents in the northwestern Atlantic are hypothesized to cause genetic structure by controlling the degree and direction of larval movements. The Gulf of Maine circulation brings cold water south along the New England coast, into Cape Cod Bay, and south along the Atlantic coast of Cape Cod (Lynch et al. 1996, 1997; Lermusiaux 2001). Although coastal currents near Cape Cod are tidally driven and strongly affected by storm events, the mean flow is southwestward along the south coast of Cape Cod and westward along both coasts of Long Island (Figure 1, Spaulding and Gordon 1982; Churchill 1985; Vieira 2000).

Finally, Pleistocene glaciation has been shown to affect some North Atlantic intertidal invertebrates by reducing genetic diversity in regions where it destroyed the organisms' habitat (usually northern sites). Species that could not survive glaciation and associated climatic changes either disappeared from glaciated sites or were forced into refugia outside of the ice sheets (Pielou 1991; Holder et al. 1999). After reemergence of glaciated habitat, the founder effect resulting from recolonization resulted in lower genetic diversity in modern populations in the north (e.g., Bernatchez and Wilson 1998; Cunningham and Collins 1998; Hare and Weinberg 2005). In other species, lower genetic diversity has been detected in southern sites, indicating recolonization from northern refugia (e.g., Dahlgren et al. 2000; Govindarajan et al. 2005). Organisms that live in soft substrates and have limited dispersal abilities should have been especially susceptible to the LGM because soft sediment habitat is limited in the Northwest Atlantic and the genetic effects of founder events persist longer without strong gene flow.

To pinpoint the location of genetic barriers near Cape Cod and to investigate their underlying causes, we examined genetic patterns in the intertidal bamboo worm *Clymenella torquata* (Annelida: Maldanidae). Several aspects of the life history of *C. torquata* make it well suited to this task: First, it is a common coastal invertebrate from New Brunswick, Canada, to Florida (Mangum 1962). Often forming dense aggregations, adults of the species live head down in tubes constructed of sand grains and use their posterior end to push surface sediments and associated interstitial organisms down the interior of the tube for ingestion at the anterior end. Second, the species' short larval life span (only a few days, Newell 1951) likely allows only limited dispersal, which increases the chances of observing genetic discontinuities around any impediments to dispersal. Third, *C. torquata* reproduces synchronously once per year when waters warm in the spring and probably lives for only a few years (Mangum 1964), conditions which approximate conventional population genetics models. Lastly, its habitation in soft sediments offers a unique comparison with existing hard-substrate studies. Although the populations analyzed herein do not represent the full geographic range of *C. torquata*, our goal is not to describe the complete phylogeography of *C. torquata* but to use its desirable life-history characteristics as a test case to investigate a specific question focused on a limited geographic region. For a species as weakly dispersive as *C. torquata*, the phylogeography examined here is likely affected minimally by
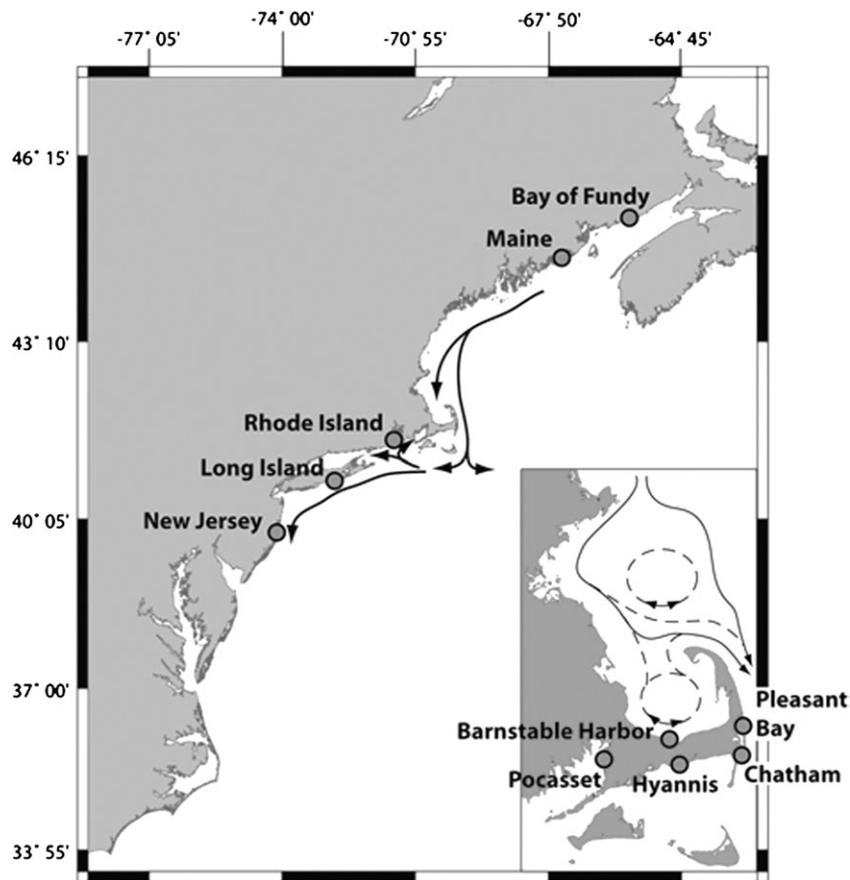
Figure 1. Populations sampled along the Atlantic coast of North America in this study, and coastal currents in the Northwest Atlantic. Gulf of Maine circulation adapted from Lynch et al. (1997); Massachusetts Bay and Cape Cod Bay circulation redrawn from Lermusiaux (2001); Middle Atlantic Bight circulation based on Spaulding and Gordon (1982) and Churchill (1985). Solid lines indicate temporally stable currents, whereas dashed lines indicate more variable currents. Gray circles show locations of populations of *Clymenella torquata* sampled in this study.

haplotypes present farther north or south in its range. We have combined samples from several closely spaced populations along the coasts of Cape Cod with multiple samples to the north and south in order to determine the location of phylogeographic breaks near Cape Cod at a finer scale. Analysis of these samples also offers the opportunity to explore more fully the possibility that environmental or historical forces can affect phylogeography in the northwestern Atlantic.

Materials and Methods

Sample Collection and Sequencing

The majority of *C. torquata* populations were sampled from May to November 2002 (Table 1 and Figure 1); these include 5 Cape Cod populations and a population from New Jersey. A population from Maine was sampled in January 2003, one from the Bay of Fundy in July 2003, and one population each from Rhode Island and Long Island in September 2003.

We obtained worm tubes by shovel, separated them from sediment, and kept them in seawater chilled on ice until sorting in the laboratory. We then removed worms from their tubes and transferred them to finger bowls containing isotonic magnesium chloride ($MgCl_2$) in seawater for species diagnosis. *Clymenella torquata* was almost always found in monospecific stands; however, we confirmed every individual used in this study to be *C. torquata* by the presence of a collarette on setiger 4 and the absence of red bands in the midregion (Mangum 1962). Incomplete anterior ends were only used if insufficient numbers of complete worms were collected and only if they were intact to setiger 4. We fixed 2 complete worms from each location in formalin overnight and stored them in 95% ethanol as vouchers. For DNA extraction, we removed tissue samples (~25 mg) from the midsection of individual worms using sterile techniques. We extracted genomic DNA with the DNeasy Extraction Kit (Promega, Madison, WI) following the manufacturer's protocol.

To determine which markers were best suited for phylogeographic analyses, we screened a number of genes

Table 1. Sampling and gene amplification information for sites in this study

Location	Collection date	Latitude Longitude	Abbreviation	N atp6	N nad4	N Both
Bay of Fundy	July 2003	45°06′00.0″N 66°24′00.0″W	BF	15	15	15
Maine	January 2003	44°57′13.0″N 67°09′45.0″W	ME	28	28	28
Barnstable Harbor	May 2002	41°42′39.6″N 70°19′29.4″W	BH	28	28	27
Pleasant Bay	August 2002	41°42′24.7″N 69°58′24.1″W	PB	24	23	22
Chatham	August 2002	41°40′00.0″N 69°58′34.0″W	CH	9	9	9
Hyannis	August 2002	41°37′57.9″N 70°19′18.3″W	HY	14	13	11
Pocasset	September 2002	41°40′27.9″N 70°38′27.7″W	P	6	4	4
Rhode Island	September 2003	41°26′57.1″N 71°27′04.7″W	R	24	24	22
Long Island	September 2003	40°47′06.8″N 72°47′26.0″W	L	17	18	15
New Jersey	September 2002	40°11′11.6″N 74°01′50.8″W	NJ	24	14	13
Total				189	176	166

N, number of individuals sequenced for *atp6*, *nad4*, or both genes.

using primers designed from *C. torquata*'s complete mitochondrial genome (Jennings and Halanych 2005). We collected DNA sequence data from 10 individuals each from Maine, Barnstable Harbor (or Hyannis), and New Jersey and compared genetic diversity among these markers (data not shown). We selected 2 mitochondrial genes, the ATPase F0 subunit 6 (*atp6*) and the nicotinamide adenine dinucleotide dehydrogenase subunit 4 (*nad4*) because they consistently amplified well and possessed the greatest sequence divergence among these populations. Nuclear internal transcribed spacer1 and internal transcribed spacer2 were also screened but exhibited no spatially informative variation in the range of *C. torquata* sampled. Primers were designed from mitochondrial genome data to amplify 650 bp of the *atp6* gene (Ctatp6f, 5′-GACCCTGCTACTAACTCTTTT-3′ and CtArgR, 5′-TTGCCACCTTTTAATGAATGA-3′) and 630 bp of the *nad4* gene (Ctnad4PGf, 5′-TATTTCTTATTCTAGGGY-GAGGTT-3′ and Ctnad4PGr, 5′-TCTTCGTGATTGG-GYGGTTTC-3′). We amplified each fragment in separate 50-μl polymerase chain reaction (PCR) reactions containing 1× PCR buffer (Promega), 2.5 mM $MgCl_2$, 2 pmol deoxynucleoside triphosphates, 1.2 pmol of each primer, 2 μl extracted DNA template, and 1 U of *Taq* polymerase (Promega). PCR conditions for both genes consisted of the following: initial denaturation, 94 °C, 1 min; 35 cycles of 94 °C, 45 s; 49 °C, 45 s; 72 °C, 1 min; final extension, 72 °C, 5 min; final hold, 4 °C. Successfully amplified DNA was purified directly from the reactions using the SV Gel and PCR Cleanup Kit (Promega) and eluted in 30 μl sterile water. We employed standard one-eighth format sequencing reactions using BigDye Terminators (version 3, Applied Biosystems, Foster City, CA) in 96-well plates and purified

sequencing reactions by isopropanol precipitation before sequencing on an ABI 3730 Capillary Sequencer. We sequenced PCR products bidirectionally and proofread the sequence reads in Auto Assembler (Applied Biosystems, Foster City, CA). Sequences from all individuals were aligned in MACCLADE 4.06 (Maddison DR and Maddison WP 2000), with the aid of inferred amino acid translations using the *Drosophila* mitochondrial code. Sequences were trimmed to minimize gaps at the beginning and end of the alignments, and all differences among individuals were verified in the electropherograms.

Population Structure

To investigate the relationships among haplotypes, we constructed a parsimony network using the programs PAUP* (version 4.0b10, Swofford 2002) and CombineTree following the method of Cassens et al. (2005). First, we concatenated sequences from the 2 loci and removed all ambiguous characters from further analysis. We then employed PAUP* to construct maximum parsimony trees using a heuristic search algorithm with tree bisection reconnection branch swapping and 10 replicates with random sequence addition. Gaps were treated as missing data. We then used CombineTree to condense the population of equally parsimonious trees and obtain the best network estimate. CombineTree first merges all trees into a single consensus network (potentially with multiple paths between haplotypes) and then uses an algorithm to determine which paths can be pruned, resulting in the best network estimate. To determine the most likely ancestral haplotypes, we computed outgroup weights by hand using

the algorithm of Castelloe and Templeton (1994). The network was then transformed into a geographic map of haplotype distributions.

To evaluate the geographic isolation of populations and regional groups, we conducted an analysis of molecular variance (AMOVA) on combined loci using ARLEQUIN (Schneider et al. 2000). The AMOVA data consisted of pairwise genetic distances using 3 hierarchical levels: within each sampling location, among locations but within regional groups, and among regional groups. We defined the regional groups based on shortest overwater distances as follows. The 5 Cape Cod sites were clustered (i.e., closer to each other than to their nearest non-Cape neighbors) and were placed into a group with Rhode Island because the latter is closer to the nearest Cape Cod site (Pocasset) than it is to the nearest non-Cape site (Long Island). Placing Maine and the Bay of Fundy into a second group and Long Island and New Jersey into a third resulted in 3 groups, for which all intragroup distances were smaller than all intergroup distances. We used overwater routes through the Cape Cod Canal in calculating distances if they were shorter than routes around the peninsula. This operational definition is based on geographic proximity in a straightforward manner and is independent of any information on genetics or water circulation. All AMOVA analyses were bootstrapped 10 000 times to assess significance. Using the program ARLE-QUIN, we computed traditional F_{ST} values, estimates of the number of migrants per generation (Nm), and significance values for all population pairs using significance correction by the sequential Bonferroni procedure of Rice (1989). Finally, we analyzed the relationship between genetic and geographic distances by performing a Mantel test (Mantel 1967) on matrices of the estimated number of migrants (Nm) and shortest overwater distance between sampled populations, with significance tested using 10 000 random permutations. This test was conducted in ARLEQUIN as formulated by Smouse et al. (1986).

We analyzed the importance of currents as dispersal vectors using the program MIGRATE version 2.1.5 (Beerli 2002) to compare models employing asymmetric versus symmetric migration rates. MIGRATE uses a Markov Chain Monte Carlo search strategy to obtain maximum likelihood estimates (MLEs) of theta ($\theta = N\mu$, where N is the effective population size and μ the mutation rate) and the number of migrants per generation per population (Nm, where m is the migration rate). MIGRATE also provides flexibility in the specification of a migration model, so that various scenarios can be proposed based on biological or physical constraints and tested against each other. We obtained MLEs for 4 migration models and used hierarchical likelihood ratio tests (hLRTs) to choose the model best supported by the data. For the first 2 models, migration was allowed between all pairs of populations. Model 1 allowed migration to be asymmetric (i.e., $M_{12} \neq M_{21}$, where M_{12} is the migration rate from population 1 to 2), whereas Model 2 constrained migration to be symmetric ($M_{12} = M_{21}$). For the other 2 models, populations were treated as a linear series from the Bay of Fundy to New Jersey following the general

downcoast direction of coastal currents into Cape Cod Bay, around Cape Cod, and further south. Migration was allowed only between a population and its 2 nearest neighbors. Model 3 allowed asymmetric rates, whereas Model 4 constrained migration to be symmetric. Nesting among the models led to a hierarchical chain of LRTs (Figure 2) starting with Model 4 (most constrained), proceeding to Model 3, and then to Model 2 or 1 (least constrained). Using MIGRATE, we optimized all models on combined *atp6* and *nad4* sequences with initial parameters (Ti/Tv of 15:1 and gamma distribution shape parameter $\alpha = 0.2701$) from empirical estimates in PAUP* 4.0b (Swofford 2002). Migration rates and their likelihoods were estimated using the "quick and dirty" option per the MIGRATE user's manual (and P.Beerli, personal communication). We ran 15 short chains with an increment of 50 and a sample of 2000 trees, followed by 3 long chains with an increment of 500 and a sample of 20 000 trees after a burn-in of 1000 trees. Several options were employed to ensure convergence and an efficient search of the parameter space (MIGRATE manual and Beerli, personal communication). The adaptive heating option was employed on short chains with initial temperatures of 1.0, 3, 6, and 12 and the "LastChains" replication option. The "Moving Steps" option was used with parameter value 0.2, an epsilon for long chains of infinity was employed, and Gelman's convergence criterion was used to ensure convergence to the maximum likelihood solution.

Genetic Diversity

To identify patterns in genetic diversity, we calculated Nei's gene diversity (called $\hat{h}$; Nei 1987, Equation 8.5, p. 179) using ARLEQUIN. Nei's is a simple measure of genetic diversity
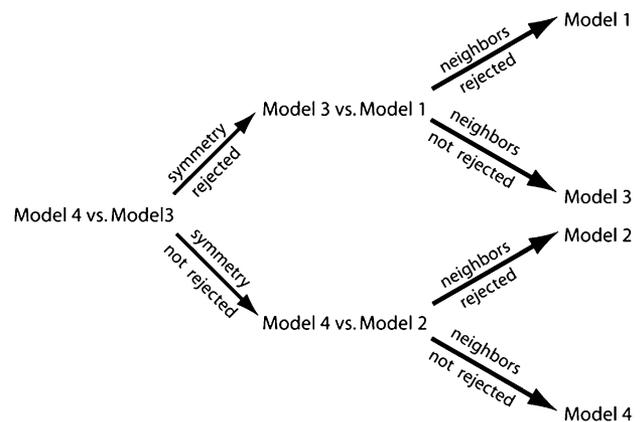
Figure 2. Schematic of hLTRs employed on 4 migration models (Models 1, 2, 3, and 4) using the software MIGRATE. Testing proceeded from left to right following the decision arrows until a model was chosen as the best fit to the data. "Neighbors" refers to migration models allowing only nearest neighbor migration; "symmetry" refers to models constraining migration to be symmetric between population pairs.
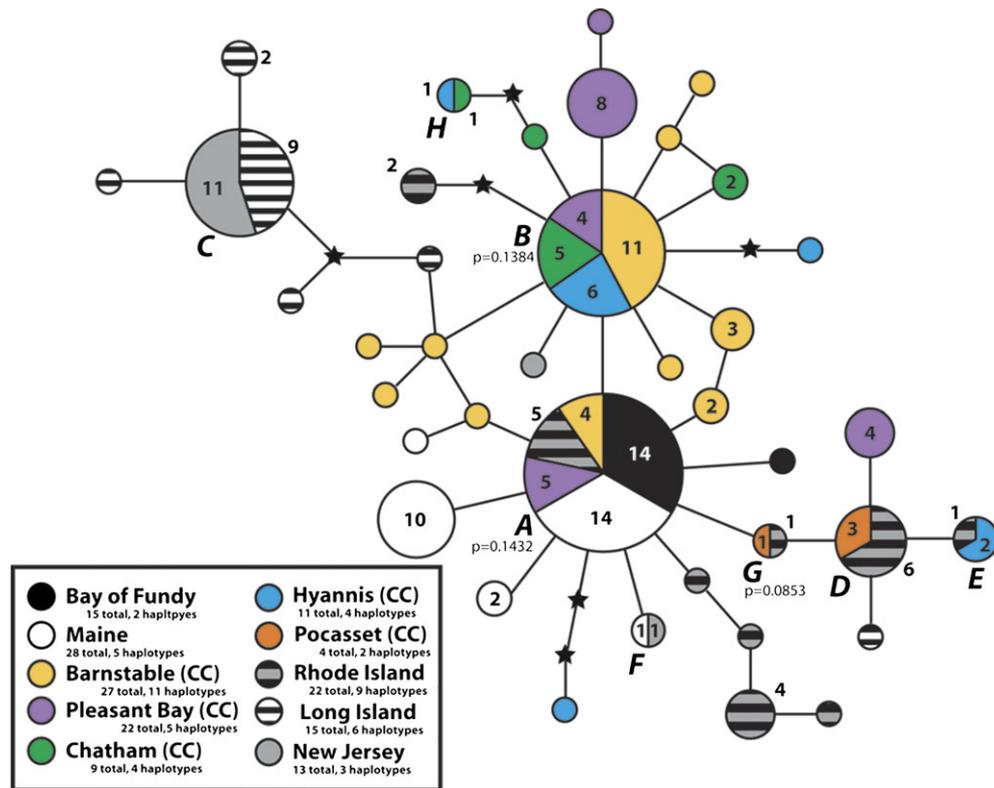
Figure 3. Parsimony network for combined gene sequences. Circles represent the different haplotypes observed, with circle areas proportional to the number of individuals possessing that haplotype. Haplotypes shared among geographic locations are further broken down into pie graphs. Connecting lines represent single base-pair differences, and missing intermediates are represented by stars. *P* values are the probabilities of haplotypes being the ancestral type. Large italic letters next to shared haplotypes correspond to their identities in the haplotype map (Figure 4).

with a straightforward biological interpretation—it is the probability that 2 randomly chosen individuals in a population bear the same haplotype. We then regressed gene diversity on degrees north latitude of sampling location using linear least squares, in the program SYSTAT 10 (SPSS, Inc., Chicago, IL).

Test of Neutrality

To test our assumption that these phylogenetic analyses were conducted on selectively neutral genes, we calculated Tajima's *D* (Tajima 1989, 1996) for each population separately in ARLEQUIN and for all populations treated as a single population. Significance values for neutrality tests were calculated from 1000 bootstrap replicates.

Results

From the 10 populations, we succeeded in amplifying a total of 189 and 176 individuals for *atp6* and *nad4*, respectively (Table 1); 23 individuals in the *atp6* data set did not amplify for *nad4* and 10 individuals in the *nad4* data set did not amplify for *atp6*, for a total of 166 individuals from which we obtained both gene sequences.

Population Structure

A total of 387 equally parsimonious trees, each 44 steps long and containing 38 haplotypes, was recovered from the heuristic PAUP* search. When condensed and pruned by CombineTree, the resultant best network estimate (Figure 3) contained only 3 cycles (groups of haplotypes with multiple connection paths) and was otherwise treelike. One cycle involved 3 haplotypes, and 2 involved 4. The 2 most common haplotypes, which together represented 41% of all individuals in the data set, occupied central positions, with less common haplotypes radiating from them. These 2 most common haplotypes were shared by 5 and 4 populations, 6 additional haplotypes were found in pairs of populations, and the remaining haplotypes were found in single populations only. We designated shared haplotypes with letters A through H in order of descending frequency in the total data set. According to the algorithm of Castelloe and Templeton (1994), haplotype A was the most likely root haplotype with a probability of 0.1432, followed by haplotype B with a probability of 0.1384; the root probability for any other haplotype was <0.0860. Six intermediate haplotypes not found in any sample were required to completely connect the network. A geographic representation of these data (the haplotype map, Figure 4) showed that the most
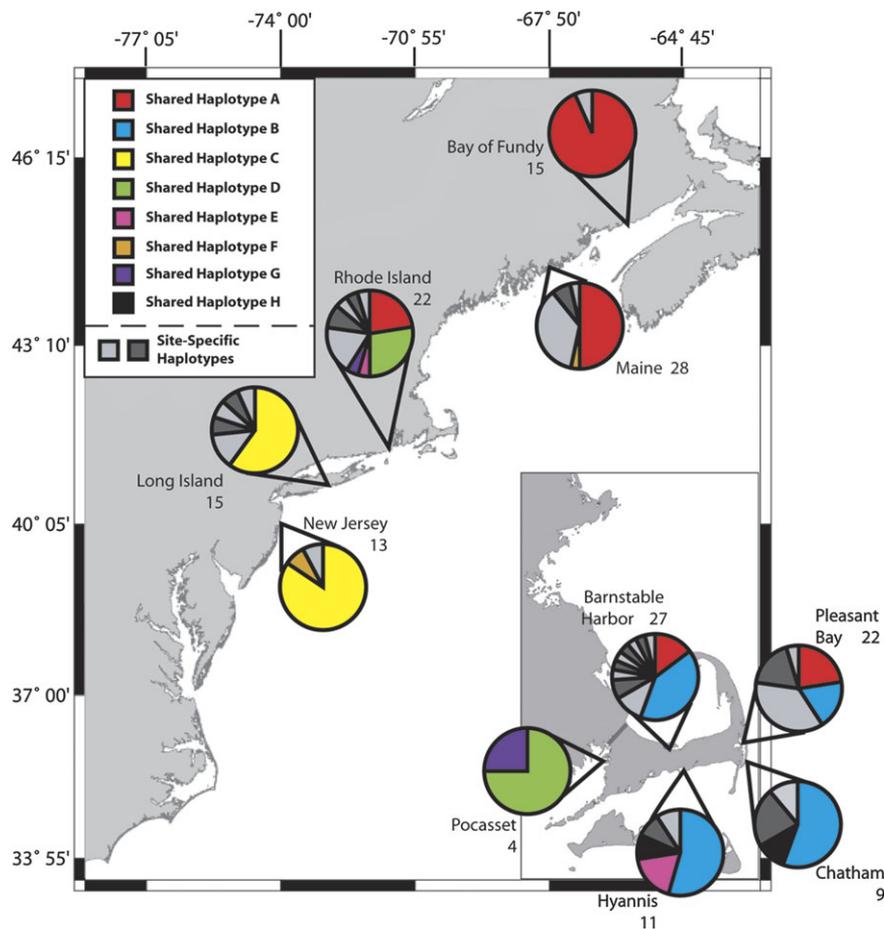
Figure 4. Haplotype map for combined gene sequences. Pie graphs indicate the haplotypic composition at each location; halftone and patterned haplotypes are shared among multiple locations, and white indicates site-specific haplotypes found only in a single location. Shared haplotypes are identified as in Figure 3.

common shared haplotype (A) was dominant in the north, declined in frequency around Cape Cod, and was absent south of Rhode Island. Haplotype B was encountered solely on Cape Cod, D and G were found solely in Rhode Island and Pocasset, and C was found only in Long Island and New Jersey. The Bay of Fundy sample exhibited markedly lower diversity than others of comparable size (Pocasset possessed low diversity, but its sample size was small).

The AMOVA analysis revealed highly significant differences at all hierarchical groupings tested (all *P* values ≤0.003, Table 2). Differences within populations explained the largest amount of total variance for both genes, followed by differences between the regional groups defined by geographic proximity. When gene flow was further analyzed between all population pairs, 36 of 45 F_{ST} values were significantly greater than zero after sequential Bonferroni correction, indicating that most population pairs exchanged few migrants (Table 3, below diagonal). Of the 9 values that were not significantly different from zero, 7 involved the small sample from Pocasset. Hyannis was not significantly differentiated from Pleasant Bay or Rhode Island.

The LRT comparing MIGRATE Model 4 with Model 3 allowed us to reject the symmetric model in favor of the asymmetric ($2\Delta \ln L = 354.73$, degrees of freedom [df] = 18, $P < 1 \times 10^{-6}$). The LRT comparing Model 3 to Model 1 similarly allowed us to reject the neighbor-only model in favor of the full, asymmetric model ($2\Delta \ln L = 880.899$, df = 72, $P < 1 \times 10^{-6}$). The final sample of parameters from all 3 tested models was tabulated to confirm convergence of the search algorithm (Supplementary Table 1). Although this extensive testing of migration models probably strains the limit of resolution for single-locus mitochondrial data (Beerli 2006), in general, the runs converged quite well, beyond a computational difficulty of MIGRATE in employing true zero as an estimate (extremely small values are instead used, often prohibiting convergence of the percentiles). The Model 1 estimates of migrants between population pairs (Table 3, above diagonal) were similar to the symmetric rates estimated in ARLEQUIN. Most migration was oriented downcoast, although some upcoast migration was inferred from Maine to the Bay of Fundy, Pleasant Bay to Maine, Pocasset to Pleasant Bay, and Pocasset to Hyannis.

Table 2. AMOVA results and significance from 10 000 bootstrap replicates

Source of variation	df	SS	Variance components	Percentage of variation	Significance
Among regional groups	2	124.10	1.10	24.72	$P = 0.003$
Among populations					
within regional groups	7	58.02	0.34	7.68	$P \leq 0.000005$
Within populations	169	468.69	3.00	67.60	$P \leq 0.000005$
Total	178	650.81	4.44		

SS, sum of squares.

These pairs involve nearest neighbors or Cape sites with more variable currents. Even though nearest neighbor asymmetric migration (Model 3) was rejected in favor of the full asymmetric model, estimated numbers of migrants were zero for the majority of nonneighbor pairs (64 of 72 possible pairs).

The Mantel test revealed a significant negative correlation between the estimated number of migrants exchanged between populations and the shortest overwater distance between them (correlation coefficient $= -0.5408$, $r^2 = 0.2925$, $P = 0.0280$). Thus, a given population is much less likely to receive migrants from distant populations than from neighboring populations, as expected given the biological constraints on larvae of *C. torquata*. An alternative Mantel test comparing the same estimated number of migrants, but this time computing distances without going through the Cape Cod Canal, resulted in a slightly smaller but still significant correlation (correlation coefficient $= -0.5221$, $r^2 = 0.2726$, $P = 0.0282$).

Genetic Diversity

A scatter plot of gene diversity against latitude of sampling location showed that the Bay of Fundy population exhibited much lower genetic diversity than the other populations (not shown). Although a linear least squares regression for all populations was significant (slope $= -0.070$, $r^2 = 0.4553$, $P = 0.032$), the Bay of Fundy was identified as a statistical outlier (Studentized residual $= -3.161$). When this data point was removed from regression analysis, no significant linear relationship remained between gene diversity and sampling latitude (slope $= -0.020$, $r^2 = 0.085$, $P = 0.4460$).

Test of Neutrality

No convincing evidence of a departure from selective neutrality was found (Table 4). Tajima's D was negative in 8 populations (Bay of Fundy, Maine, Barnstable Harbor, Chatham, Hyannis, Pocasset, Long Island, and New Jersey) and positive in 2 populations (Pleasant Bay and Rhode

Table 3. Estimated number of migrants per generation between population pairs

	BF	ME	BH	PB	CH	HY	PO	RI	LI	NJ
BF	BF	166.5618								
			4.0829	0.3547						
ME	0.35	ME		0.5399						
								0.6388		
BH	0.51	1.14	BH							
PB	0.5	1.15	2.47	PB			1.5832			
					1.1469	1.0184			1.4493	
CH	0.26	0.64	1.12	1.12	CH					
						4.0697				
HY	0.46	1.23	3.13		1.22	HY	3.0403			
								0.6385		
PO	0.3	0.96					PO			
								5.1008		
RI	0.54	1.32	3.27	3.37	1.31			RI		
LI	0.37	0.85	1.59	1.6	0.8	1.85		1.96	LI	
										1.7865
NJ	0.49	1.17	2.62	2.69	1.15	3.54		3.72	1.67	NJ

BF, Bay of Fundy; ME, Maine; BH, Barnstable Harbor; PB, Pleasant Bay; CH, Chatham; HY, Hyannis; PO, Pocasset; RI, Rhode Island; LI, Long Island; NJ, New Jersey. Below-diagonal values are estimates from ARLEQUIN (symmetric). Above-diagonal values are asymmetric estimates from MIGRATE; upper values in each box represent migrants moving upcoast, and lower values represent migrants moving downcoast. In the whole table, only statistically significant values are shown.

Table 4. Test of neutrality for each population and significance estimated from 1000 replicates

Location	Tajima's D	Significance
Bay of Fundy	−1.16	0.146
Maine	−1.21	0.131
Barnstable Harbor	−0.96	0.186
Pleasant Bay	0.81	0.812
Chatham	−0.69	0.292
Hyannis	−1.36	0.088
Pocasset	−0.61	0.378
Rhode Island	0.58	0.751
Long Island	−1.80	0.021
New Jersey	−1.34	0.099
Total population	−1.44	0.051
"North"	−1.76	0.007
"South"	−1.66	0.031

Island). Only the Long Island value was statistically significant ($P = 0.021$). When all populations were combined into one total population, Tajima's D was negative and not significant. Given the observed strong change in haplotype frequencies between Rhode Island and Long Island, populations were split into 2 groups: 1) the 2 populations south of this location (South) and 2) the 8 populations to the north (North). In this configuration, Tajima's D was significantly negative for both groups (Table 4).

Discussion

Bamboo Worm Populations Reveal a Genetic Break

The most striking feature of the parsimony network and haplotype map is an abrupt shift in haplotype frequencies between Rhode Island and Long Island, indicating a strong barrier to gene flow across the region south of Cape Cod. New Jersey and Long Island share no haplotypes with any site to the north, apart from a single haplotype encountered in one individual from New Jersey and one from Maine (haplotype F in Figure 3). Given that dispersal over such a long distance (~700 miles over water) is extremely unlikely for larvae that are only mobile for 3 days, homoplasy must be considered as a possible explanation for this haplotype's distribution. Indeed, most other New Jersey haplotypes are restricted to one branch of the network, setoff from other haplotypes; the same is true of shared Rhode Island haplotypes.

A similar but less pronounced shift in gene frequencies can be observed in the Rhode Island and Pocasset samples as compared with the others. These populations also share 2 haplotypes not found in any other population, although Rhode Island did exhibit the most common haplotype. This distribution offers the first implication that the Buzzard's Bay and Narragansett Bay regions may also be isolated, although the small sample size of Pocasset cautions against drawing a firm conclusion. Together, these facts imply that northern lineages have diverged from Long Island/New

Jersey lineages and that there has been little to no recent gene flow across this division. Both of the marked changes in haplotype frequencies occur south of Cape Cod, making it unlikely that the physical presence of the peninsula is the underlying cause. Although the extremely short larval life of *C. torquata* implies that haplotypes detected in this range should be minimally affected by populations in the rest of the species' range, the patchiness of haplotype distributions raises the possibility that New Jersey and Long Island haplotypes are aberrant, and haplotypes similar to Cape Cod and northern populations might reappear farther south. Although this question can only be answered by more extended sampling, it does not negate the observation of an abrupt break in haplotypes within the sampled range.

Possible Causes of the Genetic Break

Although previous data do not suggest a simple, uniform explanation for genetic discontinuities near Cape Cod, the data presented here do shed light on hypotheses put forward by others. There is no clear indication that the LGM had any effect on population genetics of *C. torquata* in the studied region because gene diversity was nearly constant over the sampled range. Diversity was low only in the Bay of Fundy, which is known for its large tidal range and highly dynamic environment; these factors may reduce survival of *C. torquata*, resulting in fewer haplotypes. Present-day populations of *C. torquata* are generally found in the intertidal and shallow subtidal on Cape Cod, RMJ, (personal observation) and to the north (Mangum 1964) but are more abundant subtidally to the south. Given that this distribution implies preference for cold waters, the LGM might not have been detrimental to *C. torquata* if populations were able to shift their depth distribution in response to altered water temperatures. Finally, the dynamics of warming after the LGM were more complicated than a simple reemergence of the coast. In New Jersey, the ice sheet remained on land and was never closer than 150 km to the shoreline (Duncan et al. 2000), in contrast to northern sites where ice cover extended over coastal waters. Ice is thought to have retreated from southern New England rapidly, exposing sites as far north as Maine by about 17 000 years ago (Smith and Hunter 1989). Because most sites sampled in this study were ice free very soon after glacial melting began (some 2000–3000 after the LGM), the process of coastline evolution as sea level rose may have been more relevant to *C. torquata* than successive exposure of ice-covered shores.

Although little gene flow was detected in sampled populations of *C. torquata*, coastal currents do appear to influence its magnitude and direction. The low gene flow inferred for *C. torquata* does appear to occur along coastal current vectors; however, it is unclear how currents could create the genetic discontinuity near Rhode Island. If anything, divergence of currents in this area, which flow northeast to Buzzard's Bay and southwest to Long Island Sound, should increase genetic similarity by distributing any transported northern haplotypes to multiple locations in the south. The significance of regional genetic differentiation in

the AMOVA (Table 2) and absence of haplotype A south of Rhode Island (Figure 4) argue that this is not the case. Since the mid-1930s, the Cape Cod Canal has provided a potential alternative path between sites on either side of Cape Cod. Water in the Canal oscillates on tidal cycles, but there is little net water flow through the Canal (Anraku 1964). Although this study does not explicitly examine the effect the Canal has on population genetics, the lack of a substantial increase in the Mantel test statistic when migration through the Canal is allowed argues against its being a significant dispersal pathway, at least for *C. torquata*.

In contrast to the 2 hypotheses above, the phylogeographic patterns exhibited by *C. torquata* fit well with the water mass hypothesis. Although this boundary was established from species compositions, the genetic boundary exhibited by *C. torquata* shares similar features. Both are displaced south of the southern shore of Cape Cod. Similar to the asymmetric pattern of species' ranges, there is no evidence that southern haplotypes persist in the north whereas the dominant northern haplotype (A in Figures 3 and 4) does persist at low frequency in the south. Further, more haplotypes are shared between Cape Cod and northern sites than southern sites, arguing for the closer affinity of Cape Cod to the characteristics of northern waters. Therefore, the greater similarity of water temperature in the Gulf of Maine and all around Cape Cod may facilitate gene flow among these populations while prohibiting southern haplotypes from surviving in the north even if they are transported there. Genetic differences among these geographically defined groups were highly significant (AMOVA $P \leq 0.0001$; Table 2). In addition, although no departure from neutrality was detected for individual populations or for a pooled population, a departure from neutrality was strongly implied for the 2 groupings north and south of this position. Seasonal variation in temperature could also play a role, especially because the reproductive cue in *C. torquata* is expected to propagate northward through the sampled region as water temperature warms. If this scenario is true, however, it is unclear which aspect of the smaller yearly temperature amplitude of northern sites prohibits individuals from southern populations from persisting in the north. A similar genetic break is seen in the deep sea (Weinberg et al. 2003), and a species shift in deep demersal fishes is found in the same region (J. Moore, personal communication); the greatly reduced seasonal variations of these habitats argue that average temperature is more relevant than seasonal variation in temperature.

In conclusion, the population genetic patterns of *C. torquata* are most consistent with differential tolerance of water masses north and south of Rhode Island. Overall, gene flow is low, with most exchanges involving neighboring populations in a direction congruent with local currents. Although testing of multiple migration models using MIGRATE allowed a more thorough exploration of the nature of migration for this species, multilocus data would provide greater resolution. In contrast to other species, particularly those living on hard substrates, *C. torquata*

exhibits no sign of a latitudinal gradient consistent with eradication by Pleistocene glaciation. Sampling of closely spaced populations around Cape Cod does, however, indicate that genetic discontinuities in this region are closely associated with a change in water body characteristics south of Cape Cod and are not caused by the obstruction of the peninsula itself.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

References

Anraku M. 1964. Influence of the Cape Cod Canal on the hydrography and on the copepods in Buzzards Bay and Cape Cod Bay, Massachusetts. I. Hydrography and distribution of copepods. Limnol Oceanogr. 9:46–60.

Bastrop R, Juerss K, Sturmbauer C. 1998. Cryptic species in a marine polychaete and their independent introduction from North America to Europe. Mol Biol Evol. 15:97–103.

Beerli P. 2002. MIGRATE: documentation and program, part of LAMARC. Version 1.6. [Internet]. Available from: URL http://evolution.genetics.washington.edu/lamarc.html. Revised August 23, 2002. Distributed over the Internet. [Downloaded: Version 2.0.3, 12/19/2004].

Beerli P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics. 22:341–345.

Bernatchez L, Wilson CC. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Mol Ecol. 7:431–452.

Cassens I, Mardulyn P, Milinkovitch MC. 2005. Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? Syst Biol. 54:363–372.

Castelloe J, Templeton AR. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol Phylogenet Evol. 3:102–113.

Caudill CC, Bucklin A. 2004. Molecular phylogeography and evolutionary history of the estuarine copepod, *Acartia tonsa*, on the Northwest Atlantic coast. Hydrobiologia. 511:91–102.

Churchill JH. 1985. Properties of flow within the coastal boundary layer off Long Island, New York. J Phys Oceanogr. 15:898–916.

Collin R. 2001. The effects of mode of development on phylogeography and population structure of North Atlantic *Crepidula* (Gastropoda: Caplytraeidae). Mol Ecol. 10:2249–2262.

Cunningham CW, Collins TM. 1998. Beyond area relationships: extinction and recolonization in molecular marine biogeography. In: DeSalle R, Schierwater B, editors. Molecular approaches to ecology and evolution. Basel (Switzerland): Birkhäuser Verlag. p. 297–321.

Dahlgren TG, Weinberg JR, Halanych KM. 2000. Phylogeography of the ocean quahog (Arctica islandica): influences of paleoclimate on genetic diversity and species range. Mar Biol. 137:487–495.

Dillon RT, Manzi JJ. 1992. Population genetics of the hard clam, *Mercenaria mercenaria*, at the northern limit of its range. Can J Fish Aquac Sci. 49:2574–2578.

Duncan CS, Goff JA, Austin JA, Fulthorpe CS. 2000. Tracking the last sea-level cycle: seafloor morphology and shallow stratigraphy of the latest Quaternary New Jersey middle continental shelf. Mar Geol. 170:395–421.

Engle VD, Summers JK. 1999. Latitudinal gradients in benthic community composition in Western Atlantic estuaries. J Biogeogr. 26:1007–1023.

Franz DR, Worley EK, Merrill AS. 1981. Distribution patterns of common seastars of the middle Atlantic continental shelf of the northwest Atlantic (Gulf of Maine to Cape Hatteras). Biol Bull. 160:394–418.

Gardner JPA, Palmer NL. 1998. Size-dependent, spatial and temporal genetic variation at a leucine aminopeptidase (LAP) locus among blue mussel (*Mytilus galloprovincialis*) populations along a salinity gradient. Mar Biol. 132:275–281.

Govindarajan AF, Halanych KM, Cunningham CW. 2005. Mitochondrial evolution and phylogeography in the hydrozoan *Obelia geniculata* (Cnidaria). Mar Biol. 146:213–222.

Hare MP, Weinberg JR. 2005. Phylogeography of surfclams, *Spisula solidissima*, in the western North Atlantic based on mitochondrial and nuclear DNA sequences. Mar Biol. 146:707–716.

Hayden BP, Dolan R. 1976. Coastal marine fauna and marine climates of the Americas. J Biogeogr. 3:71–81.

Holder K, Montgomerie R, Friesen VL. 1999. A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). Evolution. 53:1936–1950.

Hutchins LW. 1947. The bases for temperature zonation in geographical distribution. Ecol Monogr. 17:325–335.

Jennings RM, Halanych KM. 2005. Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): evidence for conserved gene order in Annelida. Mol Biol Evol. 22:210–222.

Kelly DW, MacIsaac HJ, Heath DD. 2006. Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. Evolution. 60:257–267.

Koehn RK, Milkman R, Mitton JB. 1976. Population genetics of marine pelecypods. 4. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. Evolution. 30:2–32.

Lee CE. 1999. Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. Evolution. 53:1423–1434.

Lermusiaux PFJ. 2001. Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay. J Mar Syst. 29:385–422.

Lynch DR, Holboke MJ, Naimie CE. 1997. The Maine coastal current: spring climatological circulation. Cont Shelf Res. 17:605–634.

Lynch DR, Ip JTC, Naimie CE, Werner FE. 1996. Comprehensive coastal circulation model with application to the Gulf of Maine. Cont Shelf Res. 16:875–906.

Maddison DR, Maddison WP. 2000. MacClade. Sunderland (MA): Sinauer Associates, Inc.

Mangum CP. 1962. Studies on speciation in the Maldanid polychaetes of the North American Atlantic Coast. I. A taxonomic revision of three species of the subfamily Euclymeninae. Yale Peabody Mus Postilla. 65, p. 12.

Mangum CP. 1964. Studies on speciation in Maldanid polychaetes of the North American Atlantic Coast. II. Distribution and competitive interaction of five sympatric species. Limnol Oceanogr. 9:12–26.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Newell GE. 1951. The life-history of *Clymenella torquata* (Leidy). Proc Zool Soc Lond. 121:561–590.

Pielou EC. 1991. After the ice age. Chicago: University of Chicago Press.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution. 43:223–225.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin: a software for population genetics data analysis. Geneva (Switzerland): Genetics and Biometry Laboratory, University of Geneva.

Smith GW, Hunter LE. 1989. Late Wisconsinan deglaciation of coastal Maine. In: Tucker RD, Marvinney RG, editors. Studies in Maine geology, volume 6: quaternary geology. Augusta (ME): Maine Geological Survey. p. 13–32.

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool. 35:627–632.

Spaulding ML, Gordon RB. 1982. A nested NUMERICAL tidal model of the Southern New England Bight. Ocean Eng. 9:107–126.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates, Inc.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Tajima F. 1996. The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. Genetics. 143:1457–1465.

Valentine JW. 1966. Numerical analysis of marine molluscan ranges on the extratropical Northeastern Pacific shelf. Limnol Oceanogr. 11:198–211.

Vieira MEC. 2000. The long-term residual circulation in Long Island Sound. Estuaries. 23:199–207.

Vogler AP, DeSalle R. 1993. Phylogeographic patterns in coastal North American tiger beetles (*Cicindela dorsalis*) inferred from mitochondrial DNA sequences. Evolution. 47:1192–1202.

Wares JP. 2002. Community genetics in the northwestern Atlantic intertidal. Mol Ecol. 11:1131–1144.

Wares JP, Cunningham CW. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. Evolution. 55:2455–2469.

Weinberg JR, Dahlgren TG, Trowbridge N, Halanych KM. 2003. Genetic differences within and between species of deep-sea crabs (Chaceon) from the North Atlantic Ocean. Biol Bull. 204:318–326.

Young AM, Torres C, Mack JE, Cunningham CW. 2002. Morphological and genetic evidence for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab *Pagurus longicarpus*. Mar Biol. 140:1059–1066.