Mar Biol (2013) 160:1065–1081
DOI 10.1007/s00227-012-2157-5

ORIGINAL PAPER

Strong population differentiation of softshell clams (*Mya arenaria*) sampled across seven biogeographic marine ecoregions: possible selection and isolation by distance

Philippe St-Onge · Jean-Marie Sévigny · Carly Strasser · Réjean Tremblay

Received: 18 August 2012 / Accepted: 18 December 2012 / Published online: 26 January 2013

Abstract Twenty-two *Mya arenaria* samples spanning seven marine ecoregions mostly situated in the Cold Temperate Northwest Atlantic (CTNA) biogeographic province were collected between 2001 and 2010 and genotyped at seven highly polymorphic microsatellite loci to test for population differentiation. Results showed strong regional differentiation with six genetic clusters: (1) Northern Gulf of St. Lawrence (GSL), (2) Magdalen Archipelago, (3) Southern GSL, (4) Lower Atlantic Canada, (5) US Coasts and (6) Northern Europe. Population structure was supported no matter the statistical approach and generally does not reflect the geographical limits of marine ecoregions. A latitudinal cline in allelic richness provides evidence for a northward post-glacial expansion range for this species. While geographical distance explains the genetic variation detected in southern CTNA, increased heterogeneity observed in northern CTNA can be explained by isolation by distance, marine landscaping and presumable selective processes acting at the *Mar5* locus. Exclusion of *Mar5* from analyses resulted in the detection of three genetic clusters instead of six.

Communicated by T. Reusch.

Electronic supplementary material The online version of this article (doi:10.1007/s00227-012-2157-5) contains supplementary material, which is available to authorized users.

P. St-Onge (✉) · R. Tremblay
Institut des Sciences de la Mer, Université du Québec à Rimouski, 310 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
e-mail: Philippe.St-Onge@uqar.qc.ca

J.-M. Sévigny
Institut Maurice-Lamontagne, Pêches et Océans Canada, 850 Route de la Mer, Mont-Joli, QC G5H 3Z4, Canada

C. Strasser
UC Curation Center, University of California Office of the President, 415 20th Street, Oakland, CA 94612, USA

Introduction

Several marine invertebrate species that remain sedentary during adult stages reproduce through a pelagic larval dispersal phase (Thorson 1950; Mileikovsky 1971). The genetic structure of such species provides important insights on the patterns of gene flow occurring in the marine system and generally reflects the dispersal pathways that facilitate the migration of new recruits amongst subpopulations (e.g. Dupont et al. 2007; Yasuda et al. 2009). Knowledge of the genetic diversity of marine invertebrate settlements provides essential tools for fisheries management, conservation biologists and understanding the population ecology of broadcast-spawning marine organisms. It can also improve the practice of artificial seeding of nursery-reared juveniles in the wild (e.g. Congleton et al. 2003; Redjah et al. 2010) by helping to maintain the genetic integrity of indigenous populations.

The softshell clam (*Mya arenaria*) is an ideal model species for population structure studies of benthic marine invertebrates undergoing a planktotrophic larval phase. It is a commercially important, broadcast-spawning and potentially long-distance dispersing species of bivalve residing in the soft-bottom environments of intertidal zones, estuaries, shallow sand bars and mud flats (e.g. St-Onge and Miron 2007; Shanks 2009). Settlements are found in most temperate ecoregions of the Atlantic and Pacific oceans (MacNeil 1965; Ponurovskii and Kolotukhina 2000), most especially in the Cold Temperate Northwest Atlantic (CTNA) biogeographic province (Spalding et al. 2007)

where the first invasion from now extinct Pacific stocks is believed to have occured during the Pliocene over 2.5 million years ago (MacNeil 1965).

The CTNA is characterized by highly heterogeneous coastlines with hundreds of estuaries, coastal lagoons, shallow and deep bays as well as a wide variety of intertidal regimes (Schmidt et al. 2008). The area can be further fragmented into distinct biogeographic marine ecoregions (Fig. 1) mainly delimited by strong taxonomic, thermal and biogeochemical similarities (Engle and Summers 1999; Spalding et al. 2007). Strong latitudinal gradients in mean summer sea surface temperature (SST) are also observed in the CTNA during summer months, varying between 8 and 12 °C in the St. Lawrence estuary (Vézina et al. 1995) and 20 °C in Chesapeake Bay (Shiah and Ducklow 1994). Such distinct physical and environmental conditions observed amongst ecoregions may (1) represent significant barriers limiting larval exchanges between them, (2) restrict gene flow and (3) influence the population structure through local adaptation (e.g. Wares 2002; Drouin et al. 2002; Domingues et al. 2010).

However, studies evaluating the genetic diversity of *M. arenaria* populations are scarce and show very little population differentiation across its natural distribution. Genotyping of the ITS-1 ribosomal DNA marker in 12 populations sampled along the New England coast resulted in only two haplotypes and no significant genetic heterogeneity (Caporale et al. 1997). For their part, Lasota et al. (2004) genotyped 302 adult clams from seven North European populations with six allozyme loci (*Gpi*, *Idh*, *Lap*, *Mdh*, *Pgd* and *Pgm*). They found low polymorphism (mean number of alleles between 2.0 and 2.7) and no population structure (mean Nei's F_{ST} = 0.0133, non-significant). Finally, Strasser and Barber (2009) widened the range of study and sampled 212 adults from 12 North American and North European populations spanning three different coastlines. Sequencing the cytochrome oxidase (*COI*) mitochondrial gene revealed significant genetic divergence between North American and North European populations (global F_{ST} = 0.159, $P < 0.001$) but no evidence of significant differentiation amongst North American populations due to one haplotype dominating (65–100 %) at all sites sampled.

The ribosomal DNA, allozyme loci and COI mitochondrial gene analysed in the above-mentioned studies could not detect the presence of population structure in *M. arenaria* due to very little genetic variation. Population genetic studies using multi-locus analyses of polymorphic microsatellite markers specific to *M. arenaria* (St-Onge et al. 2011) may thus be more appropriate for detecting finer-scale patterns of population structure in the Northwest Atlantic (Selkoe and Toonen 2006). Lack of such studies tends to lead conservation biologists towards bioregionalization concepts like marine ecoregions to help them delineate marine protected areas and implement resource management policies (Fogarty and Botsford

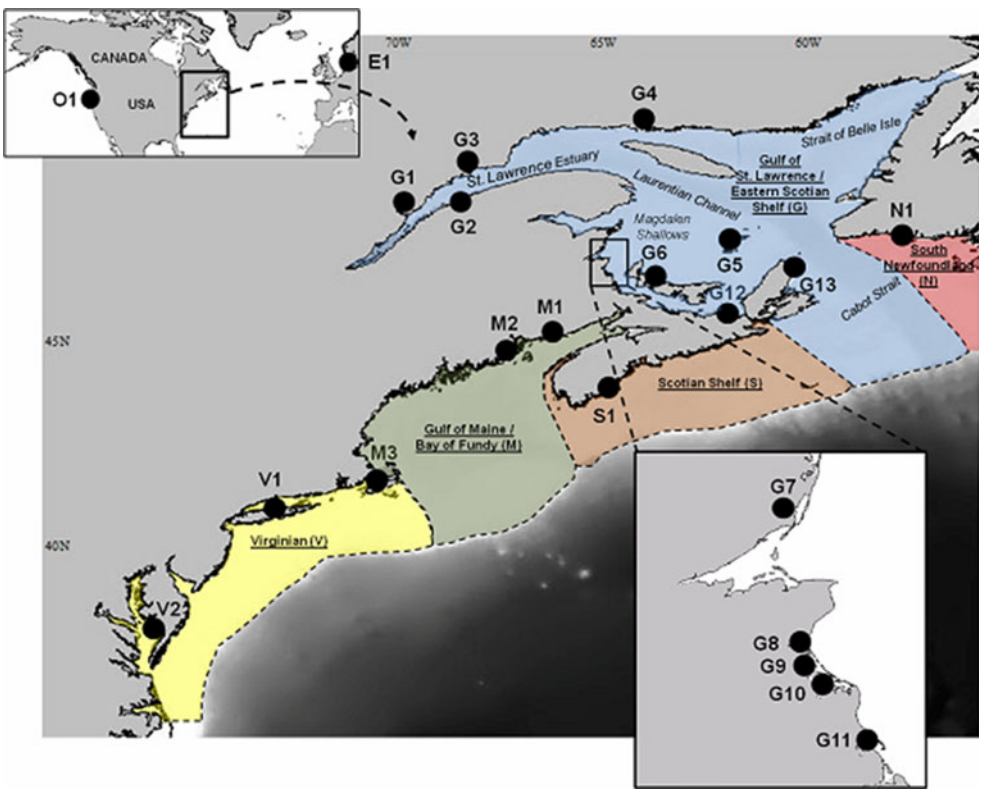

Fig. 1 Map of part of the Cold Temperate Northwest Atlantic (CTNA) biogeographic province showing the location of the 22 sampling sites and the limits of marine ecoregions (Spalding et al. 2007). *Upper inset map* shows the O1 and E1 sampling sites. Names, labels and GPS coordinates of each subpopulation are found in Table 1

2007; Spalding et al. 2007). However, the relevance of marine ecoregion classifications has seldom been tested genetically for marine species naturally distributed over large geographical areas like the CTNA. Results may provide more weight to factors such as geographical distance between settlements, environmental and physical discontinuities or local adaption (e.g. Pineira et al. 2008; Schmidt et al. 2008). Evaluating latitudinal differences in genetic diversity of *M. arenaria* across the CTNA could also give more information on its colonization history during the last deglaciation of the Late Glacial Maximum (between 10 and 13 thousand years ago) when ice retreated from most of northern CTNA and Europe (Dyke and Prest 1987; Maggs et al. 2008). Such knowledge is essential to help separate genetic patterns into distinct categories, mainly the ones resulting from historical colonization and those resulting from contemporary factors affecting gene flow and larval dispersal between settlements.

The main goal of this study will be to use a series of polymorphic microsatellite markers to test the hypothesis of significant genetic differentiation amongst *M. arenaria* settlements sampled across several marine ecoregions of the CTNA. More specifically, the study will investigate whether the population structure of this species (1) corresponds with the geographical limits of major biogeographic marine ecoregions, (2) reflects a post-glacial recolonization genetic signature caused by the last deglaciation event, (3) is correlated with distance between settlements, (4) is caused by important barriers to gene flow and (5) is influenced by locally selective processes.

Materials and methods

Sampling and genotyping

Because allele frequencies can vary over time and amongst breeding seasons (Waples 1989), we first established whether temporal structure in *M. arenaria* populations could be detected. Adult clams were collected at the two sampling sites of N1 and G8 (Fig. 1) and separated according to three age classes per site. Age was determined by counting shell annuli (Maximovich and Guerassimova 2003). Thirty-five individuals of each age class were genotyped with seven microsatellite markers (described below). No genetic differences were detected amongst the three age classes after 1,000 allele randomizations at either N1 (global Θ_{ST} = 0.0027, P = 0.109) or G8 (global Θ_{ST} = −0.0004; P = 0.919). Therefore, this study focused on evaluating the spatial genetic structure.

To assess the spatial genetic structure, a total of 604 adult and juvenile softshell clams were collected between 2001 and 2010 at 22 sampling sites spanning seven marine ecoregions (Spalding et al. 2007) (Table 1; Fig. 1). These included 13 sites in the Gulf of St. Lawrence/Eastern Scotian Shelf (G), one site in South Newfoundland (N), one site in the Scotian Shelf (S), three sites in the Gulf of Maine/Bay of Fundy (M), two sites in the Virginian (V), one site in the Oregon, Washington, Vancouver Coast and Shelf (O) and one site in the North Sea (E). Sample size differed for each site, ranging from 11 to 37 individuals. Eight of the 22 samples (G12, M1–M3, V1–V2, O1 and E1) were comprised of individuals analysed in a prior study (Strasser and Barber 2009).

Although some clams were dissected fresh, most collected clams were frozen at −40 °C upon arrival at the laboratory. Once partially thawed, small tissue samples of mantle, foot or siphon were preserved into a vial filled with 95 % ethanol. Genomic DNA of each individual was extracted from approximately 100 mg of preserved tissue with an E.Z.N.A. kit (Omega Bio-Tech Inc.).

Using a fluorescent labelling method, seven microsatellite loci (*Mar1*–GenBank accession number JN191327.1, *Mar3*—JN191329.1, *Mar4*—JN191330.1, *Mar5*—JN191331.1, *Mar6*—JN191332.1, *Mar7*–JN191333.1 and *Mar8*–JN191334.1) were amplified in each individual according to the methods described in St-Onge et al. (2011) (see also Molecular Ecology Resources Primer Development Consortium 2011). Amplified fragments were analysed by capillary electrophoresis on an ABI 3130 genetic analyser using a Genescan LIZ1200 size standard (Applied Biosystems Inc.). Raw allelic scores were assessed with the GENEMAPPER software (Applied BioSystems Inc.) (Chatterji and Pachter 2006). Final allele identity for all genotypes was based on resulting allele intervals generated by the TANDEM software (Matschiner and Salzburger 2009) and corrected by visually inspecting scatter plots of raw scores. Finally, the quality of data for the analysis of population structure (periodicity of markers, probability of null alleles, stuttering patterns and large allele dropouts at each locus of each sample) was evaluated with MICROCHECKER v.2.2.3. (Van Oosterhout et al. 2004).

Sample characteristics

Number of detected alleles, allelic richness, linkage disequilibrium between pairs of loci and conformance to Hardy–Weinberg equilibrium were all assessed with FSTAT v.2.9.3 (Goudet 2001) with the two latter set at a significance level of 5 % and adjusted for multiple comparisons. Private alleles were summarized with CONVERT v.1.31 (Glaubitz 2004), while observed and expected

Table 1 Statistics calculated over seven microsatellite loci for 22 *M. arenaria* sampling sites with GPS coordinates and year of sampling

Label—sampling site (GPS coordinates)	Year sampled	N	$_TN_A$	$_{MWL}AR$ (n = 10)	$_TN_{PA}$	H_o	H_E	F_{IS}
G1—Saguenay, Quebec (048°12′N, 069°54′W)	2005	20	68	7.26	1	0.79	0.78	−0.02
G2—Baie de Métis, Quebec (048°40′N, 068°01′W)	2005	22	67	7.23	1	0.75	0.77	0.04
G3—Pointe-aux-Outardes, Quebec (049°03′N, 068°27′W)	2010	35	87	7.69	0	0.79	0.79	−0.01
G4—Mingan, Quebec (050°18′N, 063°59′W)	2005	35	84	7.68	1	0.77	0.80	0.04
N1—Little Barasway Bay, Newfoundland (047°36′N, 057°39′W)	2005	32	88	7.99	3	0.74	0.78	0.05
G5—Havre-aux-Maisons, Quebec (047°26′N, 061°47′W)	2005	36	99	7.99	2	0.77	0.78	0.01
G6—Malpèque Bay, Prince Edward Island (046°28′N, 063°42′W)	2010	34	83	7.77	0	0.74	0.78	0.06
G7—Tabusintac, New Brunswick (047°20′N, 064°60′W)	2010	32	91	8.13	1	0.75	0.79	0.05
G8—Kouchibouguac, New Brunswick (046°50′N, 064°56′W)	2009	37	89	7.73	0	0.69	0.76	0.09
G9—St.-Louis-de-Kent, New Brunswick (046°45′N, 064°55′W)	2009	32	92	8.05	2	0.79	0.78	−0.03
G10—Richibucto, New Brunswick (046°43′N, 064°53′W)	2009	34	92	7.82	2	0.78	0.77	−0.01
G11—Bouctouche, New Brunswick (046°31′N, 064°41′W)	2010	35	100	8.35	1	0.76	0.78	0.03
G12—Antigonish, Nova Scotia (045°37′N, 061°59′W)	2001–2006	31	96	8.17	3	0.81	0.81	−0.01
G13—Ingonish Harbour, Nova Scotia (046°38′N, 060°25′W)	2010	35	87	7.72	2	0.78	0.78	0.00
S1—Port-Mouton, Nova Scotia (043°52′N, 064°48′W)	2010	33	92	8.65	0	0.79	0.82	0.05
M1—St.-John, New Brunswick (045° 17′N, 066°04′W)	2001–2006	11	60	8.55	1	0.71	0.80	0.11
M2—Pembroke, Maine (044°57′N, 067°10′W)	2001–2006	15	74	9.01	1	0.80	0.85	0.06
M3—Barnstable, Massachusetts (041°42′N, 070°20′W)	2001–2006	12	63	9.40	1	0.74	0.86	**0.14***
V1—Stony Brook, New York (040°54′N, 073°07′W)	2001–2006	13	66	8.66	1	0.83	0.85	0.02
V2—Chesapeake Bay, Maryland (038°47′N, 076°08′W)	2001–2006	35	117	9.98	6	0.81	0.87	**0.08***
O1—Newport, Oregon (044°36′N, 124°03′ W)	2001–2006	20	76	8.05	1	0.76	0.81	0.07
E1—Sylt, Germany (054°55′N, 008°21′E)	2001–2006	15	47	6.29	2	0.74	0.79	0.06

Significant F_{IS} values (Bonferoni-adjusted nominal level of 0.032 % after 3,080 randomizations) are shown in bold font with asterisk

N total number of individuals sampled, $_TN_A$ total number of detected alleles, $_{MWL}AR$ = mean within-loci allelic richness based on 10 diploid individuals, $_TN_{PA}$ total number of detected private alleles, H_O Nei's non-biased expected heterozygosity, H_E Nei's observed heterozygosity, and F_{IS} inbreeding coefficient

heterozygosity values (Nei 1978) were determined with GENETIX v.4.05 (Belkhir et al. 1996).

Spatial genetic structure

Different analytical approaches were used to investigate the population genetic structure of softshell clams sampled across the Cold Temperate Northwest Atlantic (CTNA) biogeographic province. Population differentiation was first assessed with an exact G-test (Goudet et al. 1996) with FSTAT v.2.9.3 (Goudet 2001) using the global Θ_{ST} statistic (Weir and Cockerham 1984), an unbiased estimate of F_{ST} calculated from multi-locus allelic frequencies. FSTAT was also used to calculate pairwise Θ_{ST} values (Weir and Cockerham 1984). Both tests were set at a significance level of 5 % and appropriately adjusted for multiple comparisons if needed. To account for the use of highly polymorphic microsatellite markers, mean pairwise D_{EST} values (Jost 2008) were also calculated with SMOGD (Crawford 2010) and averaged across all loci. Confidence intervals (95 %) were calculated by averaging the mean variances of all individual loci (Sokal and Rohlf 1981).

Population structure was also investigated using a Bayesian model-based MCMC clustering method as implemented in the program STRUCTURE v.2.3.3. (Pritchard et al. 2000; Falush et al. 2003). This method can be used to infer population genetic structure from individual multi-locus genotypes assuming Hardy–Weinberg and linkage equilibrium. Four simulations were carried out for each inferred K ranging between two and 22 and allowed admixture of individuals as well as correlations between allele frequencies (Falush et al. 2003). Each run consisted of a burn-in period of 50,000 steps followed by 50,000 Markov Chain Monte Carlo (MCMC) repetitions. The geographical location of individuals was only used as a prior to assist the clustering process (Hubisz et al. 2009). Resulting estimated log probabilities for each run were then used to compute the ad hoc statistic ΔK (Evanno et al. 2005), representing the most probable number of genetic clusters (K) in the data

set. Individual clams were then probabilistically assigned to one of K distinct genetic clusters.

Isolation by distance

Three independent isolation-by-distance analyses were carried out with the IBDWS program (Jensen et al. 2005) by way of a Mantel test (10,000 randomizations) correlating pairwise Θ_{ST} $(1 - \Theta_{ST})^{-1}$ values (Weir and Cockerham 1984; Rousset 1997) with geographical distance. Distance was based on the main sea surface pathways (Shore et al. 2000; Savenkoff et al. 2001; Kenchington et al. 2006) and estimated with the distance calculator available in the GEBCO Digital Atlas (British Oceanographic Data Centre 2003). Independent analyses were carried out for samples in (1) CTNA ($N = 20$), (2) G and N only ($N = 14$), and (3) S, M and V only ($N = 6$). The same three sets of samples were also used to analyse isolation by distance with pairwise D_{EST} values (Jost 2008). These analyses were carried out with the Isolde program implemented in the Genepop v.4.0.10 on the web (Raymond and Rousset 1995; Rousset 2008) and using a Spearman rank correlation test ($\alpha = 5$ %) (Sokal and Rohlf 1981).

Barriers to gene flow

The software BARRIER v.2.2 (Manni et al. 2004) was used to build a Delaunay triangulated connectivity network with Voronoi tessellations based on the geographical location of each sample. The five most significant geographical barriers to gene flow were then detected using Monmonier's algorithm (Monmonier 1973) from pairwise genetic distances (Nei 1978) between linked samples in the network.

Positive selection

Finally, locus-specific values of Θ_{ST} (Fig. 2) were calculated for each marker using FSTAT v.2.9.3 (Goudet 2001). Results showed that *Mar5* had a Θ_{ST} value of 0.058, amounting to nearly twice that of the global Θ_{ST} value of 0.031 calculated across all seven loci. The multi-loci neutrality test procedure described in Beaumont and Nichols (1996) arrived to a similar conclusion by placing *Mar5* outside and over the 95 % confidence limit of the expected *β-hat* distribution (Cockerham and Weir 1993) calculated under the island (Wright 1931, 1943) and infinite alleles mutational models (Kimura and Crow 1964). Such results suggest that *Mar5* could be a candidate marker for positive selection. Pairwise Θ_{ST}, isolation-by-distance (with pairwise Θ_{ST} and D_{EST}) as well as a clustering analysis with the STRUCTURE software were thus repeated without *Mar5* as described above. Slopes for the isolation-by-distance correlations with and without *Mar5* were compared using a one-sided paired t test ($\alpha = 0.05$) (Zar 1999). However, some parameters for the clustering analysis differed from those described above in that only three runs were carried out for K values ranging from 2 to 10 with a burn-in period of 25,000 steps and 25 000 MCMC repetitions.

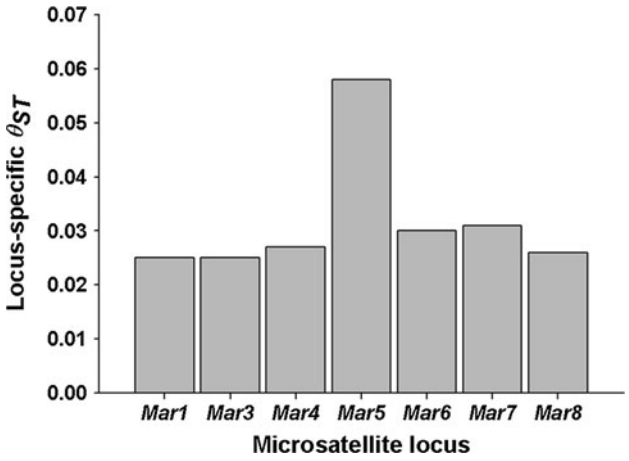

Fig. 2 Locus-specific values of Θ_{ST} for each microsatellite marker

Results

Sample characteristics

Analyses with MICROCHECKER showed that evidence for null alleles (homozygote excess) was found in only 12 out of 154 (7.8 %) combinations of loci and samples, including five for *Mar1* (G5, G6, G8, M2 and M3) and three for *Mar3* (M1, M2 and V2). However, there was no evidence of technical problems associated with stuttering patterns or large allele dropouts. Data were considered acceptable for the analysis of population structure.

No linkage disequilibrium was found between any pair of loci with P values ranging between 0.02 (*Mar4–Mar6*) and 0.94 (*Mar1–Mar6*). Significance for this test was Bonferroni-adjusted to 0.002 % after 420 randomizations. Each individual locus was thus considered statistically independent. High genetic variability was found in *M. arenaria* across all seven loci and 22 populations with a mean allelic richness of 8.1 (based on a minimum sample of 10 individuals; see *Mar8* in the sample M3) as well as high observed and expected heterozygosity values of 0.77 and 0.80, respectively (Tables 1, 2 and Online Resource 1). When calculated across samples (Table 2), total number of alleles varied between seven (*Mar6*) and 50 (*Mar4*), mean allelic richness ($N = 10$) ranged from 3.2 (*Mar6*) to 11.2 (*Mar1*), while observed and expected within-sample heterozygosity values varied from 0.50 (*Mar6*) to 0.86 (*Mar8*) and from 0.52 (*Mar6*) to 0.89 (*Mar1*), respectively. When calculated across loci (Table 1), mean

Table 2 Global descriptive statistics calculated over all samples of *M. arenaria* individuals genotyped at seven microsatellite loci

	Mar1	*Mar3*	*Mar4*	*Mar5*	*Mar6*	*Mar7*	*Mar8*
	All 22 samples						
${}_{T}N$	601	604	601	601	604	604	599
${}_{T}N_{A}$	30	28	50	28	7	24	33
${}_{M}N_{A}$	14.00	11.82	20.05	8.27	3.68	11.91	14.36
${}_{T}N_{PA}$	5	3	3	9	1	5	8
${}_{MWS}AR$	11.20	8.54	9.80	5.84	3.19	8.45	9.66
${}_{SEWS}AR$	0.42	0.30	0.31	0.40	0.11	0.27	0.28
${}_{MWS}H_{O}$	0.83	0.81	0.85	0.71	0.50	0.83	0.86
${}_{SEWS}H_{O}$	0.02	0.02	0.02	0.03	0.02	0.03	0.02
${}_{MWS}H_{E}$	0.89	0.86	0.88	0.72	0.52	0.85	0.88
${}_{SEWS}H_{E}$	0.01	0.01	0.02	0.02	0.02	0.01	0.01
${}_{WS}F_{IS}$	**0.07***	**0.04***	0.02	0.03	0.01	0.03	**0.03***

Within-sample inbreeding coefficients (${}_{WS}F_{IS}$) showing significant deficiencies in heterozygotes are highlighted in bold font with asterisk (after 1,000 allele randomizations)

${}_{T}N$ total sample size, ${}_{T}N_{A}$ total number of detected alleles, ${}_{M}N_{A}$ mean number of detected alleles per sampled site, ${}_{T}N_{PA}$ total number of private alleles, ${}_{MWS}AR$ mean within-sample allelic richness as calculated from a minimum sample size of 10 diploid individuals, ${}_{MWS}AR$ standard error of within-sample mean allelic richness, ${}_{MWS}H_{O}$ mean within-sample observed heterozygosity, ${}_{SEWS}H_{O}$ standard error of within-sample observed heterozygosity, ${}_{MWS}H_{E}$ Nei's mean non-biased within-sample expected heterozygosity, ${}_{SEWS}H_{E}$ standard error of within-sample expected heterozygosity, ${}_{WS}F_{IS}$ within-sample inbreeding coefficient

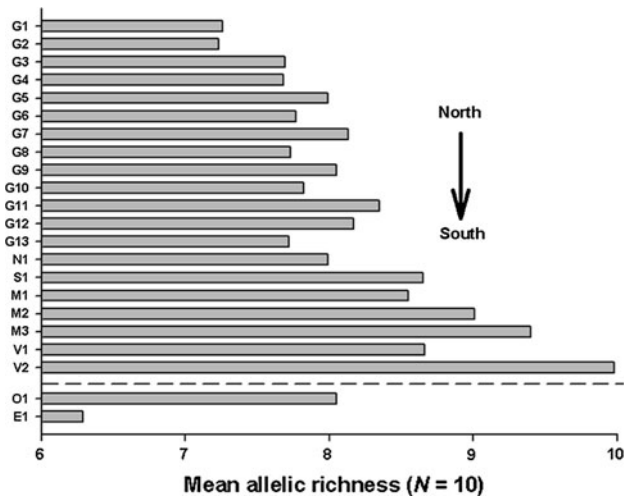

Fig. 3 Allelic richness (N = 10) based on the genotyping of 22 *M. arenaria* sampling sites from seven microsatellite markers

allelic richness (N = 10) decreased with increasing latitude (Fig. 3) and was the lowest in the E1 sample (6.3) and the highest in the V2 and M3 samples (10.0 and 9.4, respectively). Observed heterozygosity ranged from 0.70 (G8) to 0.83 (V1), while expected heterozygosity varied between 0.76 (G8) and 0.87 (V2). Global statistics (Tables 1 and 2) revealed deviations from Hardy–Weinberg equilibrium at *Mar1*, *Mar3* and *Mar8* ($P < 0.001$, 0.003 and 0.005, respectively) and in V2 and M3 (both P = 0.0003). In total, 32 private alleles were observed in 18 out of 22 samples with six of them (18.8 %) exclusive to the V2 sample.

Spatial genetic structure

The global multi-locus Θ_{ST} revealed significant population differentiation (Θ_{ST} = 0.031, $P < 0.001$). Pairwise Θ_{ST} values varied between −0.003 (G2 and G3) and 0.099 (G5 and E1) (Table 3). Out of 231 pairwise Θ_{ST} values, 161 (69.7 %) showed significant differences in allelic frequencies (see Online Resources 2 to 8 for allelic frequency tables per locus and sample). Most of the 70 pairwise tests that did not show significant differences were between sites sampled within the same geographical region. None of the 10 pairwise tests generated with G1-G4 and N1 samples (Fig. 1) showed significantly different allelic frequencies. Only six out of 36 pairwise tests (16.7 %) generated with the remaining nine G samples showed significantly different allelic frequencies with five of them implicating G5. Ten out of 21 pairwise Θ_{ST} tests (47.6 %) generated with the seven samples originating from both American coasts showed significantly different allelic frequencies. Six of these differences occurred between the northern S1, M1 and M2 and the southern V2 and O1 samples (Fig. 1). Finally, the E1 sample had allelic frequencies significantly different from all other samples.

Mean pairwise D_{EST} values (Jost 2008) across all loci and samples amounted to 0.180, which is considerably higher than the global Θ_{ST} value of 0.031. Structure patterns based on the 95 % confidence interval were more conservative than those detected with the pairwise Θ_{ST} method and showed fewer significant pairwise differences (105 out of 231 or 45.5 %), especially between samples

Table 3 Pairwise Θ_{ST} values (above diagonal) and Jost's (2008) mean pairwise D_{EST} across loci (below diagonal) between all 22 *M. arenaria* samples

	G1	G2	G3	G4	N1	G5	G6	G7	G8	G9	G10	G11	G12	G13	S1	M1	M2	M3	V1	V2	O1	N1
G1		0.01	0.01	0.02	0.03	**0.05**	**0.05**	**0.04**	**0.05**	**0.05**	**0.06**	0.05	**0.04**	**0.04**	**0.04**	**0.05**	**0.03**	**0.04**	**0.04**	**0.06**	**0.05**	**0.08**
G2	0.03		−0.00	−0.00	0.01	**0.05**	**0.03**	**0.03**	0.02	0.03	0.03	**0.03**	0.01	0.02	**0.02**	**0.04**	**0.02**	0.04	**0.04**	**0.05**	**0.05**	**0.08**
G3	0.05	−0.01		−0.00	0.00	**0.05**	**0.03**	**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	**0.01**	**0.02**	**0.02**	0.02	**0.02**	**0.03**	**0.04**	**0.05**	**0.05**	**0.07**
G4	0.10	0.00	−0.01		0.00	**0.05**	**0.02**	**0.02**	**0.02**	**0.02**	**0.02**	0.02	**0.01**	0.02	**0.01**	0.03	0.01	**0.03**	**0.03**	**0.04**	**0.04**	**0.06**
N1	0.13	0.06	0.04	0.04		**0.06**	**0.04**	**0.04**	**0.03**	**0.04**	**0.04**	**0.04**	**0.02**	**0.02**	0.01	0.01	0.01	0.03	**0.04**	**0.06**	**0.06**	**0.07**
G5	**0.19**	**0.20**	**0.21**	**0.24**	**0.29**		**0.02**	0.01	**0.02**	0.02	**0.03**	0.01	**0.03**	**0.02**	**0.04**	**0.09**	**0.04**	**0.06**	**0.08**	**0.07**	**0.07**	**0.10**
G6	**0.19**	0.12	0.14	0.14	**0.21**	0.13		0.00	0.01	0.01	0.01	−0.00	−0.00	0.01	**0.02**	**0.07**	**0.04**	**0.05**	**0.06**	**0.06**	**0.06**	**0.08**
G7	0.16	0.12	0.10	0.11	0.16	0.11	0.03		0.00	−0.00	−0.00	−0.01	0.00	0.01	**0.02**	**0.07**	**0.02**	**0.04**	**0.06**	**0.05**	**0.06**	**0.08**
G8	0.16	0.07	0.09	0.09	0.12	0.14	0.05	0.03		−0.00	0.00	−0.00	0.01	0.01	**0.02**	**0.07**	**0.03**	**0.06**	**0.07**	**0.07**	**0.07**	**0.10**
G9	0.17	0.09	0.09	0.09	0.16	0.10	0.04	−0.03	−0.00		0.00	−0.00	0.01	**0.01**	**0.02**	**0.06**	**0.03**	**0.05**	**0.07**	**0.06**	**0.07**	**0.09**
G10	**0.19**	0.09	0.09	0.08	0.16	**0.18**	0.07	0.01	0.02	0.00		0.00	0.01	0.01	**0.02**	**0.07**	**0.03**	**0.05**	**0.07**	**0.06**	**0.07**	**0.09**
G11	0.17	0.12	0.11	0.09	**0.18**	0.09	0.01	−0.05	0.02	−0.02	0.02		0.00	0.01	**0.02**	**0.07**	**0.03**	**0.05**	**0.07**	**0.06**	**0.06**	**0.09**
G12	0.16	0.07	0.08	0.06	0.12	**0.19**	−0.00	0.03	0.02	0.03	0.02	0.02		0.00	0.01	0.04	**0.02**	**0.04**	**0.04**	**0.05**	**0.05**	**0.07**
G13	0.15	0.08	0.07	0.09	0.10	0.14	0.06	0.06	0.06	0.07	0.05	0.05	0.01		0.01	**0.05**	**0.02**	**0.05**	**0.05**	**0.06**	**0.06**	**0.08**
S1	**0.18**	0.13	0.09	0.08	0.05	**0.22**	0.13	0.10	0.11	0.10	0.12	0.12	0.07	0.06		0.01	−0.00	**0.02**	**0.03**	**0.04**	**0.04**	**0.06**
M1	**0.26**	**0.19**	0.13	0.09	0.07	**0.40**	**0.31**	**0.28**	**0.26**	**0.23**	**0.23**	**0.26**	0.16	**0.19**	0.08		0.00	0.01	0.03	**0.05**	**0.05**	**0.05**
M2	0.17	0.14	0.09	0.06	0.02	**0.22**	**0.23**	0.14	0.17	0.13	0.16	0.13	0.12	0.08	0.01	0.01		0.00	0.00	**0.02**	**0.03**	**0.03**
M3	**0.23**	0.16	0.14	0.13	0.14	**0.33**	**0.27**	0.14	**0.23**	**0.21**	**0.21**	**0.20**	**0.22**	**0.21**	0.16	0.09	0.02		0.01	0.02	0.01	**0.04**
V1	**0.26**	**0.22**	**0.22**	**0.20**	**0.20**	**0.41**	**0.33**	**0.33**	**0.33**	**0.33**	**0.35**	**0.34**	**0.24**	**0.24**	0.17	**0.18**	0.02	0.13		**0.02**	0.01	**0.06**
V2	**0.38**	**0.34**	**0.35**	**0.32**	**0.35**	**0.43**	**0.39**	**0.32**	**0.43**	**0.38**	**0.40**	**0.34**	**0.35**	**0.37**	**0.32**	**0.32**	**0.19**	0.17	0.17		**0.02**	**0.06**
O1	**0.30**	**0.26**	**0.28**	**0.23**	**0.31**	**0.36**	**0.32**	**0.29**	**0.33**	**0.32**	**0.34**	**0.30**	**0.29**	**0.29**	**0.27**	**0.26**	**0.19**	0.06	0.10	**0.21**		**0.07**
E1	**0.36**	**0.39**	**0.35**	**0.33**	**0.34**	**0.47**	**0.38**	**0.37**	**0.44**	**0.42**	**0.41**	**0.39**	**0.38**	**0.38**	**0.33**	**0.28**	**0.22**	**0.25**	**0.34**	**0.41**	**0.36**	

Significant values are indicated in bold font. See Table 1 for information on population labels

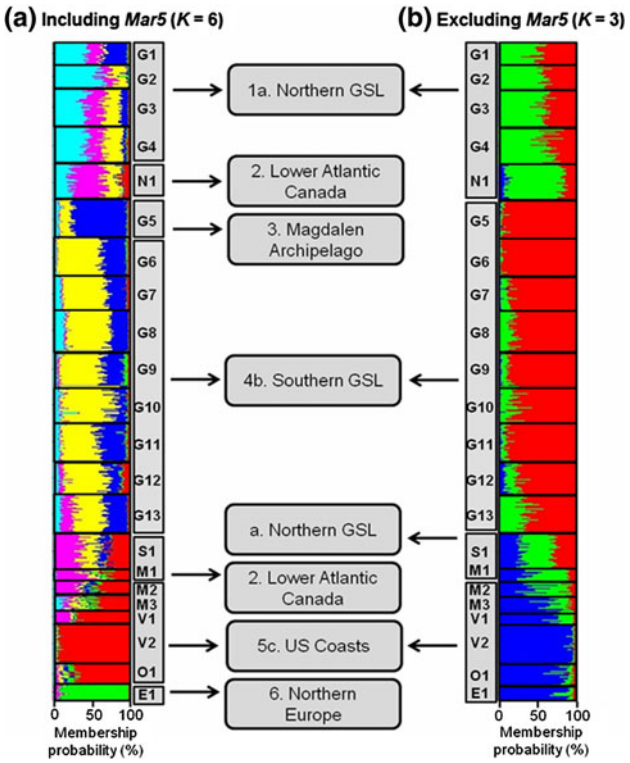

Fig. 4 Summary plots of the 604 *M. arenaria* individual coefficients of membership to each inferred cluster obtained with the STRUCTURE software performed with **a** six clusters ($K = 6$) and all seven loci (clusters 1–6 in Table 4) and with **b** three clusters ($K = 3$) and six microsatellite markers (all but *Mar5*). *Shade colours* within each *plot* represent distinct clusters. *Sites* are labelled as in Table 1, while *letters* and *numbers* in cluster *name boxes* refer to membership probabilities in Table 4

from the northern range of the CTNA (Table 3). Most values outlying the 95 % confidence interval were observed between samples south and north of M1.

Six genetic clusters ($K = 6$) were detected with STRUCTURE (Figs. 4a and Online Resource 9): (1) four northern G samples (G1–G4); (2) G5; (3) all remaining G samples (G6–G13); (4) N1, S1 and M1; (5) V and O samples including M2 and M3; and (6) E1 (see Table 4 for membership probabilities to each cluster).

Isolation by distance

Pairwise $\Theta_{ST}(1 - \Theta_{ST})^{-1}$ and mean pairwise D_{EST} values obtained between the 20 CTNA samples showed a significant positive correlation with distance ($r^2 = 0.463$ and $r_S^2 = 0.567$, respectively, both $P < 0.0001$) (Fig. 5a, b). The same results were observed when the data set was partitioned according to only G and N samples ($r^2 = 0.307$, $P = 0.0006$ and $r_S^2 = 0.303$, $P < 0.0001$, respectively) (Fig. 5cd) and only S, M and V samples ($r^2 = 0.686$, $P = 0.0011$ and $r_S^2 = 0.716$, $P < 0.0001$, respectively) (Fig. 5e, f).

Barriers to gene flow

The implementation of Monmonier's algorithm from the Delaunay triangulated connectivity network built with BARRIER v.2.2 showed that most probable barrier to gene flow was detected around the G5 sample (Fig. 6a), while the second, third and fourth most probable barriers isolated the three southernmost samples, that is, M3, V1 and V2 (Fig. 6b, c, d). Finally, the fifth barrier to gene flow isolated the four northernmost GSL samples (G1–G4) (Fig. 6e).

Positive selection

Removal of the *Mar5* locus from the analyses revealed its strong influence on the genetic population structure of *M. arenaria* in the CTNA. The clustering analysis carried out with STRUCTURE without *Mar5* genotypes resulted in only three genetic clusters ($K = 3$) instead of six (Table 4; Figs. 4b and Online Resource 10). In this scenario, the N1, S1 and M1 samples were clustered with G1–G4 in the northern Gulf of St. Lawrence, the G5 sample was clustered with G6–G13 in the southern Gulf of St. Lawrence, while the sample E1 was clustered with M2–M3, V1–V2 and O1.

Furthermore, the pairwise Θ_{ST} analysis performed without *Mar5* revealed that only 133 pairwise tests out of 231 (57.6 %) were significantly different (Online Resource 11) compared to 161 (69.7 %) when *Mar5* was included (Table 3). A proportion of 41.9 % of altered significant differences caused by the removal of *Mar5* occurred between northern and southern G samples. Exclusion of *Mar5* genotypes also significantly reduced the influence of distance on the isolation of samples in the northern CTNA (Paired t test, $t_{178} = 3.17$, $P < 0.002$) (Fig. 5c), while it significantly increased in the whole CTNA range (Paired t test, $t_{376} = 2.99$, $P < 0.005$) (Fig. 5a). Although genetic distance as a function of geographic distance increased in the southern range without *Mar5* (Fig. 5e), slopes were not significantly different (Paired t test, $t_{26} = 0.39$, $P > 0.50$). Allelic distribution for *Mar5* showed a dominance of the 182-bp allele in the northern part of the G ecoregion (G1–G4) and in the N ecoregion (N1) and a dominance of the 170-bp allele in the southern part of the G ecoregion (G5–G13) (Fig. 7). The frequencies of these alleles were more evenly distributed in the remaining sampled sites.

Discussion

Spatial genetic structure

Regardless of which statistical approach was used and contrary to previous studies (Caporale et al. 1997; Lasota et al., 2004; Strasser and Barber 2009), the use of polymorphic

Table 4 Probabilities of membership as determined with STRUCTURE to each of the six clusters ($K = 6$) when the locus *Mar5* is included and to each of the three clusters ($K = 3$) when *Mar5* is excluded from analyses

Label—sampling site	Probabilities of membership to each inferred cluster								
	Including *Mar5* ($K = 6$)						Excluding *Mar5* ($K = 3$)		
	1	2	3	4	5	6	a	b	c
G1—Saguenay, Quebec	**0.44**	0.19	0.30	0.04	0.01	0.02	**0.50**	0.48	0.01
G2—Baie de Métis, Quebec	**0.63**	0.14	0.03	0.19	0.00	0.02	**0.58**	0.41	0.01
G3—Pointe-aux-Outardes, Quebec	**0.42**	0.25	0.10	0.22	0.01	0.01	**0.59**	0.40	0.02
G4—Mingan, Quebec	**0.45**	0.21	0.05	0.25	0.03	0.02	**0.65**	0.34	0.01
N1—Little Barasway Bay, Newfoundland	0.25	**0.42**	0.03	0.21	0.09	0.01	**0.74**	0.16	0.10
G5—Havre-aux-Maisons, Quebec	0.08	0.02	**0.71**	0.16	0.02	0.02	0.04	**0.94**	0.02
G6—Malpèque Bay, Prince Edward Island	0.04	0.01	0.32	**0.58**	0.02	0.02	0.03	**0.96**	0.02
G7—Tabusintac, New Brunswick	0.11	0.02	0.26	**0.57**	0.03	0.01	0.11	**0.86**	0.04
G8—Kouchibouguac, New Brunswick	0.15	0.03	0.23	**0.57**	0.01	0.01	0.18	**0.81**	0.01
G9—St.-Louis-de-Kent, New Brunswick	0.06	0.03	0.28	**0.56**	0.02	0.05	0.09	**0.87**	0.04
G10—Richibucto, New Brunswick	0.10	0.03	0.17	**0.66**	0.02	0.02	0.189	**0.79**	0.02
G11—Bouctouche, New Brunswick	0.09	0.02	0.35	**0.49**	0.03	0.03	0.07	**0.89**	0.04
G12—Antigonish, Nova Scotia	0.08	0.07	0.18	**0.55**	0.09	0.03	0.16	**0.74**	0.10
G13—Ingonish Harbour, Nova Scotia	0.11	0.17	0.30	**0.40**	0.02	0.01	0.35	**0.64**	0.01
S1—Port-Mouton, Nova Scotia	0.02	**0.35**	0.13	0.23	0.27	0.01	**0.41**	0.29	0.30
M1—St.-John, New Brunswick	0.04	**0.43**	0.03	0.12	0.26	0.12	**0.52**	0.10	0.38
M2—Pembroke, Maine	0.03	0.32	0.10	0.06	**0.40**	0.09	0.45	0.09	**0.46**
M3—Barnstable, Massachusetts	0.14	0.10	0.03	0.14	**0.44**	0.15	0.25	0.08	**0.67**
V1—Stony Brook, New York	0.03	0.21	0.01	0.03	**0.68**	0.03	0.23	0.03	**0.74**
V2—Chesapeake Bay, Maryland	0.03	0.01	0.01	0.02	**0.93**	0.01	0.03	0.01	**0.96**
O1—Newport, Oregon	0.04	0.04	0.05	0.06	**0.72**	0.09	0.11	0.04	**0.85**
E1—Sylt, Germany	0.01	0.07	0.02	0.02	0.01	**0.87**	0.12	0.03	**0.85**

The most probable cluster for each population is indicated in bold font for both analyses

microsatellite markers in this study clearly showed the existence of six spatially distinct genetic clusters of *M. arenaria* subpopulations, five of which were situated exclusively within the Cold Temperate Northwest Atlantic (CTNA) province (Fig. 4a). Results show a much greater genetic structuring of *M. arenaria* than what was previously reported in the CTNA for other marine bivalve species, such as the Arctic surfclam *Mactromeris polynyma* (Cassista and Hart 2007) or to a lesser extent *Placopectenmagellanicus* (Kenchington et al. 2006). Clusters detected with the STRUCTURE software will hereafter be named (1) Northern GSL (G1–G4), (2) Magdalen Archipelago (G5), (3) Southern GSL (G6–G13), (4) Lower Atlantic Canada (N1, S1 and M1), (5) US Coasts (M2–M3, V1–V2 and O1) and (6) Northern Europe (E1). The validity of these clusters was independently supported by pairwise Θ_{ST} (Table 3) and landscape genetic analyses (Fig. 6).

Genetic relevance of marine ecoregions

The geographical patterns of the above-mentioned genetic clusters are not synonymous with those delimiting biogeographic marine ecoregions (Spalding et al. 2007) given the fact that the Lower Atlantic Canada and US Coasts clusters each cover three distinct ecoregions (Fig. 1). Another important discrepancy was observed at the level of the G ecoregion which is comprised of three distinct genetic clusters: Northern GSL, Southern GSL and Magdalen Archipelago. Similar conclusions of discrepancies between genetic structure and marine ecoregions could be made with other mollusc species, for example, the brooding gastropod *Littorina saxatilis* (Panova et al. 2011). However, our results are in agreement with the biodiversity assemblages described for this ecoregion which follow the same geographical patterns as described above, especially for the northern and southern regions of G (Chabot et al. 2007; DFO 2009; Archambault et al. 2010).

Since the difference between age classes (see "Materials and methods") and the marine ecoregion classifications were both ruled out as potential factors explaining the genetic variation in our data set, four other major factors can be invoked to explain the genetic population differentiation of *M. arenaria* settlements in the CTNA. These

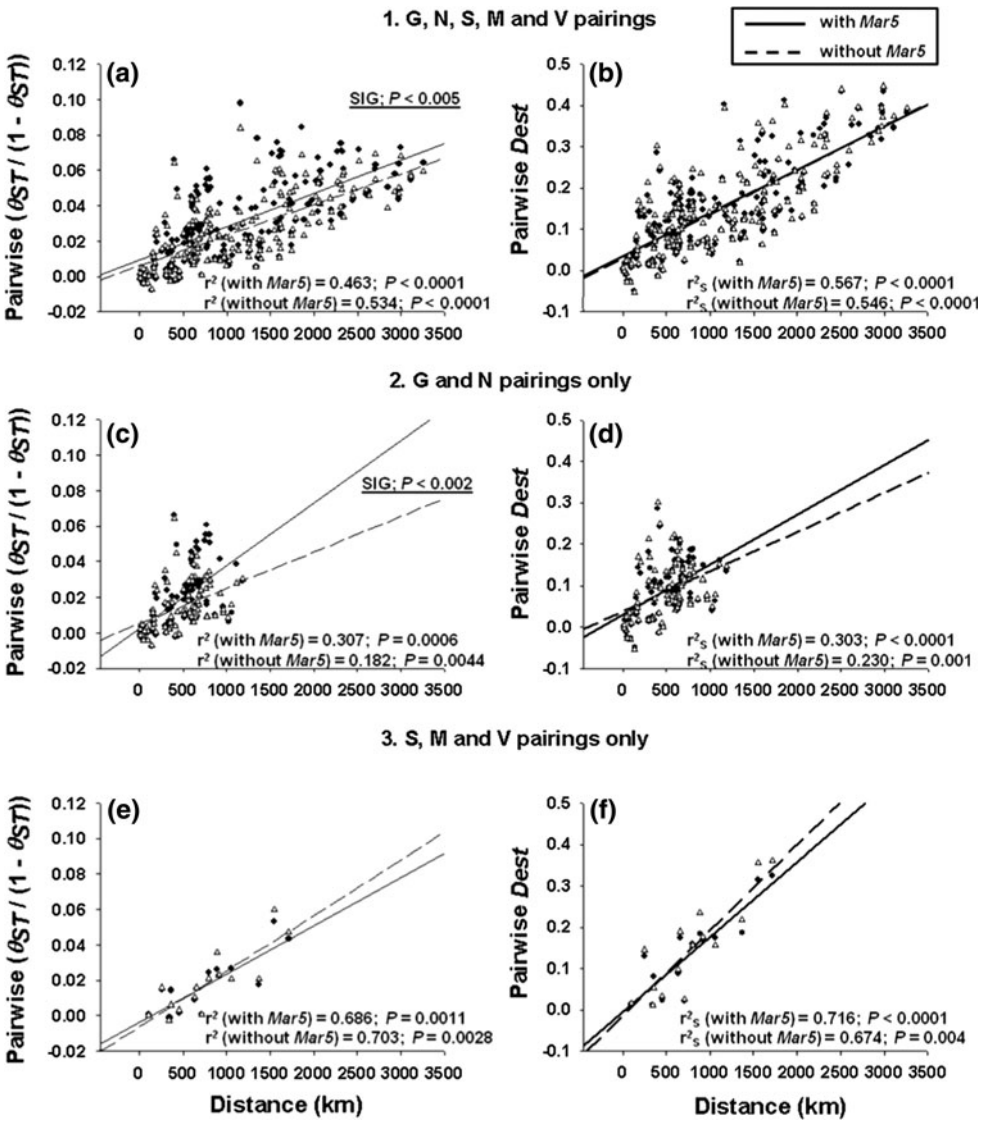

Fig. 5 Isolation-by-distance analyses performed with and without *Mar5* according to pairwise Θ_{ST} $(1 - \Theta_{ST})^{-1}$ (Rousset 1997) and D_{EST} (Jost 2008) for three groups of samples: (*1*) G, N, S, M and V samples ($N = 20$); (*2*) G and N samples ($N = 14$); and (*3*) S, M and V samples ($N = 6$). Significant differences between slopes are marked with the *underlined inscription* "SIG" presented with the associated *P* value

are enumerated here in what seems to be the most likely chronological order of their historical impact on the current state of population structure: (1) northward post-glacial expansion range, (2) differentiation of sampling sites due to the geographical distance, (3) marine landscaping limiting or facilitating larval dispersal and gene flow between settlements, and (4) variable selection pressures and subsequent maintenance of spatially distinct genotypes potentially due to environmental heterogeneity within the study area.

Post-glacial expansion range

One could argue that the sample showing the highest number of private alleles (V2) should be considered older than the others and thus a source population from which neighbouring settlements likely originated through a source-sink post-glacial colonization scenario (Maggs et al. 2008; Vuilleumier et al. 2010). In these scenarios, geographic regions characterized with relatively high allelic richness are associated with ancient glacial refugia where surviving populations were maintained outside of the ice sheet cover areas (Maggs et al. 2008; Kenchington et al. 2009). Since V2 also represents the most southerly site included in the US Coasts cluster, the colonization was probably achieved incrementally in a northward fashion following the disappearance of the Laurentide ice sheet present over Atlantic Canada and most of the northern United States during the end of the Pleistocene (Dyke and Prest 1987). This would explain the observed northward latitudinal decrease in allelic richness (Fig. 3) and suggest a southern refugia during the Last Glacial Maximum, which is also in agreement with preliminary identification reports suggesting the relative absence of *M. arenaria* shells in Holocene shell deposits along the north shore of the St. Lawrence estuary (Bernatchez et al. 1999; A. Martel,

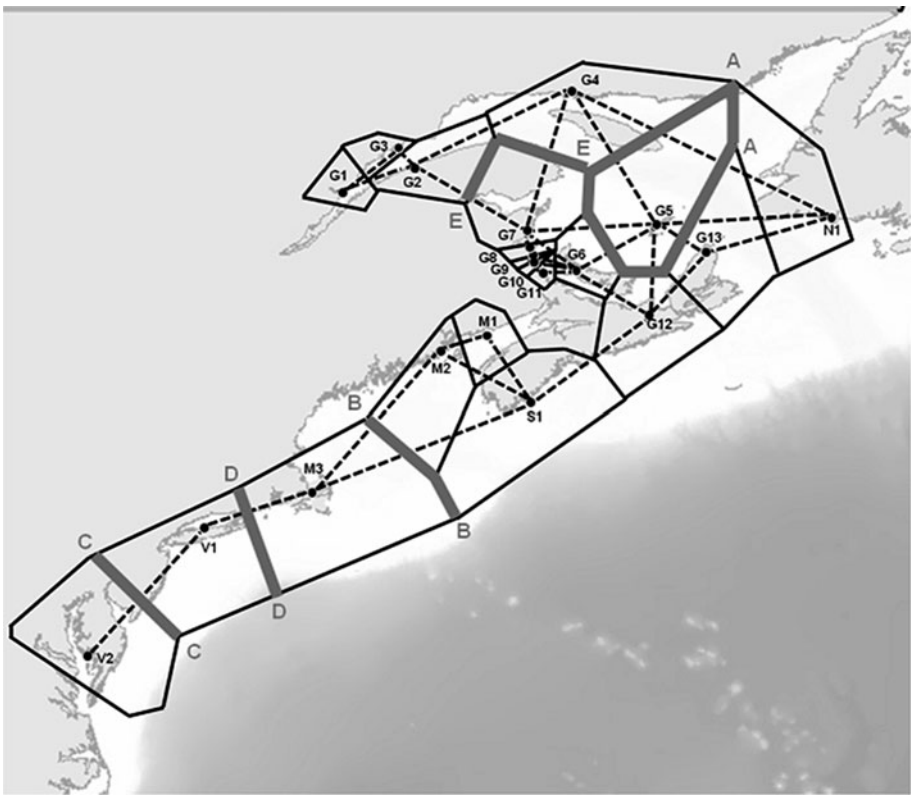

Fig. 6 Delaunay triangulated connectivity network of the 20 CTNA samples with Voronoi tessellations built with the BARRIER v.2.2 software. *Grey lines* represent the five most probable barriers to gene flow (decreasing probability from **a**–**e**) as detected with Monmonier's maximum difference algorithm from pairwise genetic distances (Nei 1978)

personal communication). A northward expansion trend was also observed in other marine organisms (Wares and Cunningham 2001; Young et al. 2002; Zuccarello and West 2003; Baker et al. 2008; Kenchington et al. 2009; Santos et al. 2012). However, the opposite southward trend was observed for both the ocean quahog *Arctica islandica* (Dahlgren et al. 2000) and the cockle *Cerastoderma edule* (Krakau et al. 2012), suggesting a rather northern refugia for these species during the Last Glacial Maximum.

Since results suggest that the northern distribution of *M. arenaria* was the most recently colonized, one could find it surprising to observe up to four distinct genetic clusters in Atlantic Canada alone. However, it seems that the Gulf of St. Lawrence region was characterized by the presence of local ice caps around 18,000 years ago which may have contributed to *M. arenaria* invading different locations at separate occasions (Dyke and Prest 1987). Additionally, when the candidate marker to positive selection *Mar5* is excluded from analyses, only two genetic clusters remain. The extent at which *M. arenaria* populations are structured at neutral markers in the north-western Atlantic is thus similar to that of other marine bivalves that readily expanded in the northern Atlantic during the Late Glacial Maximum (e.g. Santos et al. 2012).

Sea surface currents that modulated larval dispersal during this geological phenomenon differed from those prevailing today (Paul and Schäfer-Neth 2003; Kenchington et al. 2009) and might then have been favourable to a northward range expansion for *M. arenaria*. However, a southward direction of post-glacial colonization was suggested for *A. islandica* despite the evidence for dispersal occurring predominantly from the western Atlantic to the eastern Atlantic (Dahlgren et al. 2000). One should thus be careful when solely using contemporary current patterns to explain the population structure of broadcast-spawning bivalves and should attribute a great importance to the colonization history of the species through geological time (Wares 2002). In this context, contemporary currents are rather associated as important barriers to gene flow, useful in explaining how the genetic integrity of clusters is maintained across generations.

Isolation by distance

Geographic distance between *M. arenaria* samples influences the degree of genetic differentiation at various spatial scales, including both northern and southern ranges of the CTNA as well as the CTNA as a whole (Fig. 5). Distance separating the S, M and V samples explains between 69 and 72 % of the genetic variation (Fig. 5e, f). This strong isolation along the US Coasts cluster is most likely responsible for driving the isolation observed in the CTNA. Interestingly, the US Coasts is the only cluster found within a very wide geographical range of approximately 1,450 km

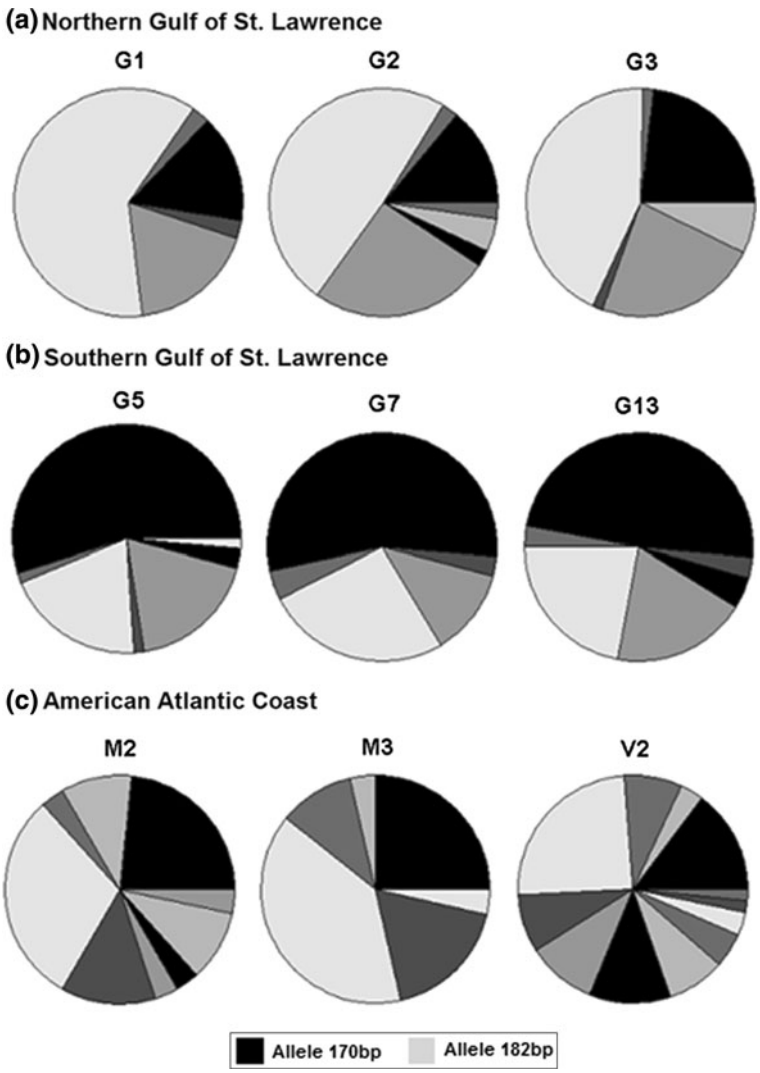

Fig. 7 Relative frequencies of the 182 (*light grey*) and 170 bp (*black*) *Mar5* alleles found in three random sampling sites of the **a** Northern GSL, **b** Southern GSL/Magdalen Archipelago, and **c** American Coasts clusters. Relative frequencies of remaining *Mar5* alleles are represented by the other incrementing *grey colours*

(distance between M2 and V2; Fig. 1) and includes both M and V ecoregions. An increasing southward cline in membership probabilities (difference of 53 %) to the cluster is, however, clearly observed between samples at both geographical extremes of the cluster, that is, M2 and V2 (Table 4), which also show significantly different allelic frequencies (Pairwise $\Theta_{ST} = 0.0173$; $P = 0.0002$) (Table 3).

These patterns of isolation by distance imply that gene flow between the more geographically separated samples of the US Coasts cluster was accomplished incrementally over several generations through accumulated colonization events. Such results are also consistent with a mean dispersal range of 35 km for *M. arenaria* larvae (as reviewed by Shanks 2009) and the observed barriers to gene flow (Fig. 6). The O1 sample is closely related to all other US Coasts samples (Table 3; Fig. 4a) despite the significant distance separating them, which is consistent with the fact that clams were introduced onto the western American coast sometime during the last 150–400 years from Northwestern Atlantic stocks (Powers et al. 2006; Strasser and Barber 2009).

On the other hand, the northern part of the CTNA sampling range (G and N ecoregions) was scattered across a smaller area with a maximal geographical distance of approximately 1100 km between G1 and G12 (Fig. 1). Despite being smaller, this area is comprised of two genetic clusters when the potentially selected *Mar5* is not considered (Fig. 4b) and up to four clusters when it is indeed considered (Fig. 4a). Distance between settlements was significantly correlated with transformed pairwise Θ_{ST} values but only explains between 30 and 31 % of the genetic variation (Fig. 5c, d). This implies that the *M. arenaria* population structure in northern CTNA and in the G ecoregion is more genetically heterogeneous than in the south. Clustering of the more northerly samples should thus be the consequence of differential spatial distribution

of a smaller number of alleles (relatively to southern CTNA) maintained by both selective effects and possibly lower rates of gene flow between settlements (Fig. 6a, e).

Both pairwise Θ_{ST} and D_{EST} statistics produced similar results as dependent variables for isolation-by-distance analyses (Fig. 5). As expected, pairwise D_{EST} values were greater than pairwise Θ_{ST} values (Table 3). Even if Θ_{ST} is usually considered valuable to infer population genetic structure, it can underestimate population differentiation due to the high mutation rate associated with highly polymorphic microsatellite loci which tend to increase heterozygosities and lower Θ_{ST} values (Hedrick 2005; Jost 2008). However, structure patterns based on the 95 % confidence intervals of the pairwise D_{EST} distribution were more conservative than those observed using the pairwise Θ_{ST} approach (Table 3) (see Table 2 in Schunter et al. 2011 for similar findings). This may be due to the fact that D_{EST} is not a statistic meant to test for the null hypothesis of no significant differentiation between two samples, rather representing an actual relative measure of genetic differentiation between two samples independent from Nei's (1978) gene diversity statistics.

Barriers to gene flow

Three out of the five most probable barriers to gene flow detected from Monmonier's algorithm (Fig. 6b, c, d) were observed between samples included in the US Coasts cluster, suggesting that gene flow between settlements in southern CTNA most likely follows a linear stepping-stone model (Kimura and Weiss 1964). A break in surface currents around Cape Cod, Massachusetts (e.g. Wares 2002; Strasser and Barber 2009), is likely to explain the presence of Barrier B. Site-dependent physical factors such as major sea surface or local currents, coastal topography and the presence of density fronts impeding water masses from mixing can significantly modulate gene flow by shaping dispersal pathways and allowing or prohibiting the transport of larvae from one site to another. A good example is the coastal sea surface current that primarily flows southward from the Grand Banks of Newfoundland to Cape Hatteras, North Carolina, where it is cut-off by the Gulf Stream (e.g. Pietrafesa et al. 1994; Shearman and Lentz 2010). These southward currents may have contributed to the historical maintenance of some of the more southerly clam settlements and may also have limited the historical northward spread of the species (Wares 2002).

In our data set, V2 had the highest number of private alleles and the highest mean allelic richness based on 10 genotyped individuals (Table 1; Figs. 3, 6c). The area covered by the Chesapeake Bay is semi-sheltered from all other US Coasts settlements which could explain the presence of Barrier C favourable to larval retention processes and the fixation of rare alleles arising from mutations (Slatkin 1985).

In the vicinity of the G ecoregion, the barriers to gene flow shaping the Northern GSL, Magdalen Archipelago and Southern GSL clusters are not as straightforward as predicted above for US Coasts, mainly because of the more complex landscape and physical characteristics of the Gulf of St. Lawrence estuarine system. Surface water circulation in the St. Lawrence Estuary (SLE) is mainly characterized by an outflow of fresh St. Lawrence River water along the southern shore which generates the Gaspe Current (Sheng 2001; Gan et al. 2004). While a small part of the latter runs into the Magdalen Shallows, a more important volume is flushed directly into the Atlantic Ocean through the Cabot Strait following the Laurentian Channel (Sheng 2001; Gan et al. 2004), partly explaining the presence of Barrier E (Fig. 6e). Limited gene flow occurring between northern and southern G settlements is thus more likely to be achieved in a southward fashion as most water being transported into the SLE is situated very deep below the zone where primary production is optimal and where developing larvae thrive (i.e. photic zone).

Terrestrial boundaries and barriers (e.g. Newfoundland, Prince Edward Island and Nova Scotia) also limit gene flow between most of the Canadian and American settlements. Several samples included in the Southern GSL cluster were sampled in enclosed bays and lagoons protected by prominent sand dunes or terrestrial peninsulas with restricted (e.g. G7–G9) and semi-restricted openings (e.g. G6 and G10–G11) to the open sea. These might significantly diminish the ability of pelagic larvae to travel across clusters. Barrier effects to gene flow can often be caused by the presence of fjords (e.g. Suneetha and Naevdal 2001; Skold et al. 2003). However, *M. arenaria* individuals sampled in the Saguenay Fjord (G1) were not different from those sampled in the neighbouring estuary (G2–G4). This lack of differentiation was also observed for other species, including bottom fishes and crustaceans analysed with microsatellite and allozyme loci (Sevigny et al. 2009). Interestingly, the Magdalen Archipelago sample (G5) was significantly differentiated from all samples included in the Southern GSL cluster and represent the most significant barrier to gene flow in the analysed data set (Fig. 6a). However, G5 is clustered with the rest of the Southern GSL cluster when *Mar5* is not taken into account, which suggests that differentiation of this cluster is due to selection rather than geographical seclusion.

Positive selection

Exclusion of *Mar5* genotypes from analyses clearly shows the influence of this locus on the genetic structuring of

M. arenaria settlements throughout the CTNA. This is especially true for the clustering analysis that fuses the Magdalen Archipelago and Southern GSL clusters and fuses the Lower Atlantic Canada and Northern GSL clusters (Fig. 4b). Removal of *Mar5* also resulted in 28 fewer significant pairwise Θ_{ST} tests, mostly between northern and Southern GSL samples (Online Reource 11). *Mar5* also influenced isolation by distance which significantly decreased in the G and N ecoregions (although still significantly different) (Fig. 5c) but significantly increased in the CTNA as a whole (Fig. 5a). While the latter shows the importance of geographical isolation in the southern range of the CTNA (see above), the former suggests that selective effects may potentially be responsible for a portion of the genetic differentiation of northern settlements.

Region-specific allelic distributions show that the 182-bp allele dominated in the Northern GSL cluster while the 170-bp allele dominated in both Magdalen Archipelago and Southern GSL clusters (Fig. 7). The major discrepancy between both group locations is the sea surface temperature (SST). Summer SST within the GSL ecoregion averages 8–12 °C in the St. Lawrence Estuary (Vézina et al. 1995) and over 25 °C in the shallow bays and lagoons of the southern Gulf of St. Lawrence (Lachance et al. 2008). Strong salinity gradients in the G ecoregion are also present since the St. Lawrence Estuary represents one of the largest discharge of fresh water in the world (Dai and Trenberth 2002).

Differentiation originating from environmental variables causing selection can sometimes be detected if microsatellites or other presumably neutral loci are linked to the locus on which selection is acting (e.g. Larsson et al. 2007; Westgaard and Fevolden 2007). Allelic diversity in several species of marine invertebrates has been shown to covary with environmental factors, including latitudinal SST gradients (e.g. Elderkin and Klerks 2001; Kenchington et al. 2006), salinity gradients (e.g. Michinina and Rebordinos 1997; Camilli et al. 2001) and even levels of intertidal desiccation (e.g. Schmidt et al. 2000; Schmidt and Rand 2001).

The environmental factors described above are likely to influence larval and juvenile clams during two critical periods. The first of these periods is during their pelagic larval development, that is, immediately before reaching competency. Lutz and Jablonski (1978) showed that size at metamorphosis of developing *M. arenaria* pediveligers in the water column was inversely correlated with water temperature, most likely due to high metabolism rates at higher temperatures. The second critical period of development susceptible of being influenced by selective processes is during the competency period which includes habitat exploration, delayed metamorphosis, settlement and other various post-settlement events. For example, several studies investigating the role of selection in the recruitment processes of the acorn barnacle (*Semibalanus balanoides*) found that survival rates of different genotypes were contingent upon the habitat encountered in the few days immediately following settlement (Schmidt et al. 2000; Schmidt and Rand 2001; Drouin et al. 2002; Schmidt et al. 2008). Larvae should have optimal survival rates at their birth location, whereas incoming larvae from broodstock adapted to a different set of environmental conditions are likely to be at a disadvantage, following a resident preemption dynamic (Vuilleumier et al. 2010). These selection processes can lead to the isolation of favoured genotypes and increase population structuring.

Conclusion

This study showed the presence of six spatially defined genetic clusters of *M. arenaria* settlements across seven distinct marine ecoregions. Selection processes and geographical barriers to gene flow are suggested to play a more important role in the northern range of the CTNA, whereas geographical distance mainly dictates the intensity of genetic differentiation in the southern range of the CTNA. It is thus important to consider both processes (selection and gene flow) together with the post-glacial colonization history when examining the genetic structure of *M. arenaria*. The examination of multi-locus patterns of genetic variation at large geographical scales carried out in this study will provide an interesting framework for future smaller-scale connectivity and recruitment studies.

Acknowledgments Authors wish to thank the following people for their much appreciated help in the field and in the laboratory: Éric Parent, Éric Tremblay, Léophane Leblanc, Firmin Leblanc, Gilles Miron, Julie Quimper, Alexandra Valentin, Philippe Galipeau, Chantale Daigle, Carole Degrâce, Andre Siah, Sylvie Brulotte, Patrice Pelletier, the Kouchibouguac, Kejimkujik and Cape Breton Highlands National Parks and the Atlantic Veterinary College. Authors would also like to thank all anonymous reviewers who commented on earlier drafts of the manuscript. This study was funded by the Aquaculture Collaborative Research and Development Program (ACRDP), the National Sciences and Engineering Research Council of Canada (NSERC) and the Réseau Aquaculture du Québec (RAQ) grants to P. St-Onge, J.M. Sévigny and R. Tremblay.

References

Archambault P, Snelgrove PVR, Fisher JAD et al (2010) From sea to sea: Canada's three oceans of biodiversity. PLoS ONE 5:e12182

Baker P, Austin JD, Bowen BW, Baker SM (2008) Range-wide population structure and history of the northern quahog (*Mercenaria mercenaria*) inferred from mitochondrial DNA sequence data. ICES J Mar Sci 65:155–163

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc B Biol Sci 263:1619–1626

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows^{TM} pour la génétique des populations. Université de Montpellier II, Montpellier

Bernatchez P, Dubois JMM, Dionne JC (1999) Holocene shell beds of Baie-Comeau along the North shore of the St. Lawrence estuary (Quebec). Can J Earth Sci 36:519–531

British Oceanographic Data Centre (2003) Centenary edition of the GEBCO Digital Atlas. Published on behalf of the International Hydrographic Organization (IHO) and the International Oceanographic Commission (IOC) of UNESCO, Liverpool, UK

Camilli L, Castelli A, Lardicci C, Maltagliati F (2001) Evidence for high levels of genetic divergence between populations of the bivalve *Mytilaster minimus* from a brackish environment and two adjacent marine sites. J Molluscan Stud 67:506–510

Caporale DA, Beal BF, Roxby R, VanBeneden RJ (1997) Population structure of *Mya arenaria* along the New England coastline. Mol Mar Biol Biotech 6:33–39

Cassista MC, Hart MW (2007) Spatial and temporal genetic homogeneity in the Arctic surfclam (*Mactromeris polynyma*). Mar Biol 152:569–579

Chabot D, Rondeau A, Sainte-Marie B, Savard L, Surette T, Archambault P (2007) Distribution of benthic invertebrates in the Estuary and Gulf of St. Lawrence. Can Sci Advis Secr 2007/018

Chatterji S, Pachter L (2006) Reference based annotation with GeneMapper. Genome Biol 7:R29

Cockerham CC, Weir BS (1993) Estimation of gene flow from F-statistics. Evolution 47:855–863

Congleton WR, Pearce BR, Parker MR, Causey RC (2003) Mariculture siting—tidal currents and growth of *Myaarenaria*. J Shellfish Res 22:75–83

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10:556–557

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

Dai A, Trenberth KE (2002) Estimates of freshwater discharge from continents: latitudinal and seasonal variations. J Hydrometeorol 3:660–687

Department of Fisheries and Oceans Canada (2009) Development of a framework and principles for the biogeographic classification of Canadian Marine Areas. Can Sci Advis Secr 2009/056

Domingues CP, Creer S, Taylor MI, Queiroga H, Carvalho GR (2010) Genetic structure of *Carcinus maenas* within its native range: larval dispersal and oceanographic variability. Mar Ecol Prog Ser 410:111–123

Drouin C-A, Bourget E, Tremblay R (2002) Larval transport processes of barnacle larvae in the vicinity of the interface between two genetically different populations of *Semibalanus balanoides*. Mar Ecol Prog Ser 229:165–172

Dupont L, Ellien C, Viard F (2007) Limits to gene flow in the slipper limpet *Crepidula forniacata* as revealed by microsatellite data and a larval dispersal model. Mar Ecol Prog Ser 349:125–138

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide Ice Sheet. Géogr Phys Quat 41:237–263

Elderkin CL, Klerks PL (2001) Shifts in allele and genotype frequencies in zebra mussels, *Dreissena polymorpha*, along the latitudinal gradient formed by the Mississippi River. J N Am Benthol Soc 20:595–605

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

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fogarty MJ, Botsford LW (2007) Population connectivity and spatial management of marine fisheries. Oceanography 20:112–123

Gan J, Ingram RG, Greatbatch RJ, van der Baaren T (2004) Variability of circulation induced by the separation of Gaspe Current in Baie des Chaleurs (Canada): observational studies. Estuar Coast Shelf Sci 61:393–402

Glaubitz JC (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes 4:309–310

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3. http://www2.unil.ch/popgen/softwares/fstat.htm

Goudet J, Raymond M, de-Meeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1933–1940

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:13–18

Jost L (2008) G(ST) and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). Mol Ecol 15:1781–1796

Kenchington EL, Harding GC, Jones MW, Prodöhl PA (2009) Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*. Mol Ecol 18:1654–1667

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kimura M, Weiss GH (1964) Stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

Krakau M, Jacobsen S, Jensen KT, Reise K (2012) The cockle *Cerastoderma edule* at Northeast Atlantic shores: genetic signatures of glacial refugia. Mar Biol 159:221–230

Lachance AA, Myrand B, Tremblay R, Koutitonsky V, Carrington E (2008) Biotic and abiotic factors influencing attachment strength of blue mussels *Mytilus edulis* in suspended culture. Aquatic Biol 2:119–129

Larsson LC, Laikre L, Palm S, Andre C, Carvalho GR, Ryman N (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. Mol Ecol 16:1135–1147

Lasota R, Hummel H, Wolowicz M (2004) Genetic diversity of European populations of the invasive soft-shell clam *Mya arenaria* (Bivalvia). J Mar Biol Assoc UK 84:1051–1056

Lutz RA, Jablonski D (1978) Larval bivalve shell morphometry: a new paleoclimatic tool? Science 202:51–53

MacNeil FS (1965) Evolution and distribution of the genus *Mya*, and Tertiary migrations of Mollusca. Geol Surv Prof Pap 483-G

Maggs CA, Castilho R, Foltz D et al (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. Ecology 89:S108–S122

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum Biol 76:173–190

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982–1983

Maximovich NV, Guerassimova AV (2003) Life history characteristics of the clam *Mya arenaria* in the White Sea. Helgoland Mar Res 57:91–99

Michinina SR, Rebordinos L (1997) Genetic differentiation in marine and estuarine natural populations of *Crassostrea angulata*. Mar Ecol Prog Ser 154:167–174

Mileikovsky SA (1971) Types of larval development in marine bottom invertebrates, their distribution and ecological significance: a re-evaluation. Mar Biol 10:193–213

Molecular Ecology Resources Primer Development Consortium (2011) Permanent genetic resources added to molecular ecology resources database 1(June), pp. 2011–31, July 2011. Mol Ecol Resour 11:1124–1126

Monmonier M (1973) Maximum-difference barriers: an alternative numerical regionalization method. Geogr Anal 3:245–261

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Panova M, Blakeslee AMH, Miller AW, Mäkinen T, Ruiz GM, Johannesson K, André C (2011) Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. PLoS ONE 6:e17511

Paul A, Schäfer-Neth C (2003) Modelling the water masses of the Atlantic Ocean at the last glacial maximum. Paleoceanography 18:1058–1084

Pietrafesa LJ, Morrison JM, McCann MP, Churchill J, Bohm E, Houghton RW (1994) Water mass linkages between the Middle and South Atlantic bights. Deep Sea Res II 41:365–389

Pineira J, Quesada H, Rolan-Alvarez E, Caballero A (2008) Genetic discontinuity associated with an environmentally induced barrier to gene exchange in the marine snail *Littorina saxatilis*. Mar Ecol Prog Ser 357:175–184

Ponurovskii SK, Kolotukhina NK (2000) Larval settlement of the bivalves *Mya arenaria* and *M. uzenensis* from scallop collectors in Vostok Bay, Sea of Japan. Russian J Mar Biol 26:330–337

Powers SP, Bishop MA, Grabowski JH, Peterson CH (2006) Distribution of the invasive bivalve *Mya arenaria* L. on intertidal flats of southcentral Alaska. J Sea Res 55:207–216

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

Raymond M, Rousset F (1995) GENEPOP (version1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Redjah I, Olivier F, Tremblay R, Myrand B, Pernet F, Neumeier U, Chevarie L (2010) The importance of turbulent kinetic energy on transport of juvenile clams (*Mya arenaria*). Aquaculture. 307:20–28

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Rousset F (2008) GENEPOP' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Santos S, Cruzeiro C, Olsen JL, van der Veer HW, Luttikhuizen PC (2012) Isolation by distance and low connectivity in the peppery furrow shell *Scrobicularia plana* (Bivalvia). Mar Ecol Prog Ser 462:111–124

Savenkoff C, Vézina AF, Smith PC, Han G (2001) Summer transports of nutrients in the Gulf of St. Lawrence estimated by inverse modelling. Estuar Coast Shelf Sci 52:565–587

Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat- and life-stage-specific survivorship of Mpi genotypes. Evolution 55:1336–1344

Schmidt PS, Bertness MD, Rand DM (2000) Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. Proc R Soc B Biol Sci 267:379–384

Schmidt PS, Serrão EA, Pearson GA et al (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. Ecology 89:S91–S107

Schunter C, Carreras-Carbonell J, Macpherson E et al (2011) Matching genetics with oceanography: directional gene flow in a Mediterranean fish species. Mol Ecol 20:5167–5181

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615–629

Sevigny JM, Valentin A, Talbot A, Menard N (2009) Connectivité entre les populations du fjord du Saguenay et celles du Golfe du Saint-Laurent. Rev Sci Eau 22:315–339

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. Biol Bull 216:373–385

Shearman RK, Lentz SJ (2010) Long-term sea surface temperature variability along the US East Coast. J Phys Oceanogr 40:1004–1017

Sheng J (2001) Dynamics of a buoyancy-driven coastal jet: the Gaspé Current. J Phys Oceanogr 31:3146–3162

Shiah F-K, Ducklow HW (1994) Temperature regulation of heterotrophic bacterioplankton abundance, production, and specific growth rate in Chesapeake Bay. Limnol Oceanogr 39:1243–1258

Shore JA, Hannah CG, Loder JW (2000) Drift pathways on the western Scotian Shelf and its environs. Can J Fish Aquatic Sci 57:2488–2505

Skold M, Wing SR, Mladenov PV (2003) Genetic subdivision of a sea star with high dispersal capability in relation to physical barriers in a fjordic seascape. Mar Ecol Prog Ser 250:163–174

Slatkin M (1985) Rare alleles as indicators of gene flow. Evolution 39:53–65

Sokal RR, Rohlf FJ (1981) Biometry, 2nd edn. W.H Freeman and Company, New York, NY

Spalding MD, Fox HE, Halpern BS et al (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. Bioscience 57:573–583

St-Onge P, Miron G (2007) Effects of current speed, shell length and type of sediment on the erosion and transport of juvenile softshell clams (*Mya arenaria*). J Exp Mar Biol Ecol 349:12–26

St-Onge P, Parent E, Sevigny JM, Tremblay R, Rioux-Pare R (2011) Isolation and characterization of eight novel microsatellite markers for the softshell clam (Mya arenaria). Mol Ecol Resour Primer Database mer-11-0122

Strasser CA, Barber PH (2009) Limited genetic variation and structure in softshell clams (*Mya arenaria*) across their native and introduced range. Conserv Genet 10:803–814

Suneetha KB, Naevdal G (2001) Genetic and morphological stock structure of the pearlside, *Maurolicus muelleri* (Pisces, Sternoptychidae), among Norwegian fjords and offshore area. Sarsia 86:191–201

Thorson G (1950) Reproductive and larval ecology of marine bottom invertebrates. Biol Rev 25:1–45

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Vézina AF, Gratton Y, Vinet P (1995) Mesoscale physical—biological variability during a summer phytoplankton bloom in the lower St. Lawrence Estuary. Estuar Coast Shelf Sci 41:393–411

Vuilleumier S, Goudet J, Perrin N (2010) Evolution in heterogeneous populations: from migration models to fixation probabilities. Theor Popul Biol 78:250–258

Waples RS (1989) Temporal variation in allele frequencies—testing the right hypothesis. Evolution 43:1236–1251

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 American intertidal. Evolution 55:2455–2469

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Westgaard JI, Fevolden SE (2007) Atlantic cod (*Gadus morhua* L.) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. Fish Res 85:306–315

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Wright S (1943) Isolation by distance. Genetics 28:114–138

Yasuda N, Nagai S, Hamaguchi M, Okaji K, Gerard K, Nadaoka K (2009) Gene flow of *Acanthaster planci* (L.) in relation to ocean currents revealed by microsatellite analysis. Mol Ecol 18:1574–1590

Young AMC, 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

Zar JH (1999) Biostatistical analysis, 4th edn. Prentice Hall, Upper Saddle, NJ

Zuccarello GC, West JA (2003) Multiple cryptic species: molecular diversity and reproductive isolation in the *Bostrychia radicans*/*B. moritziana* complex (Rhodomelaceae, Rhodophyta) with focus on North American isolates. J Phycol 39:948–959