

Contemporary nuclear and mitochondrial genetic clines in a north temperate estuarine fish reflect Pleistocene vicariance

I. R. Bradbury^{1,2,*,**}, M. W. Coulson^{2,3,**}, S. E. Campana⁴, I. G. Paterson², P. Bentzen²

¹Fisheries and Oceans Canada, Science Branch, 80 East White Hills Road, PO Box 5667, St. John's, Newfoundland and Labrador A1C 5X1, Canada

²Marine Gene Probe Laboratory, Biology Department, Life Sciences Centre, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

³Rivers and Fisheries Trusts of Scotland, Marine Scotland Freshwater Laboratory Faskally, Pitlochry PH8 0RG, UK

⁴Population Ecology Division, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada

ABSTRACT: Contemporary genetic spatial structure in north temperate marine species is likely the culmination of multiple vicariant and dispersive cycles. Here we evaluate spatial genetic structure in an estuarine fish, rainbow smelt *Osmerus mordax*, from coastal Newfoundland, Canada, using data from both mtDNA (ND5) sequences and nuclear loci (11 microsatellites). Sequence analysis of ND5 identified a previously unrecognized genetic discontinuity between 2 hypothesized glacial clades in southeastern Newfoundland. Microsatellite based tests for directional selection identified a locus (Omo11, $p < 0.001$) that mirrored mtDNA clades in the geographic distribution of its 2 common alleles but did not display elevated differentiation following correction for heterozygosity. Bayesian multilocus clustering of the remaining microsatellite loci supported the presence of 2 predominant groups, for which the spatial distribution was also largely consistent with those of the mtDNA and Omo11 clades. Taken together, the similarity in microsatellite and mtDNA clines supports the hypothesis that contemporary spatial structure in smelt reflects historical landscape isolation maintained by low dispersal and selective processes producing reinforcement between diverging populations. As genetic structure in northern marine and estuarine species may be largely determined by historical glacial cycles of vicariance, contemporary estimates of connectivity should be interpreted in the context of both past and present landscape structure.

KEY WORDS: Dispersal · Connectivity · Microsatellite · mtDNA · Glacial isolation · Phylogeography · Rainbow smelt

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Present-day spatial genetic structure represents the culmination of historical and contemporary influences on gene flow (e.g. Barton & Hewitt 1985, Avise 2000, Hewitt 2000). This interaction between contemporary and historical forces is particularly evident in temperate and high latitude species where cycles of glaciation have caused repeated range contrac-

tions and expansions (Pielou 1991, Hewitt 2000). Both genetic and paleontological data suggest that Pleistocene glacial advances constricted the ranges of many terrestrial and marine temperate species to isolated refugia (Bernatchez & Wilson 1998, Hewitt 2004, Maggs et al. 2008, Provan & Bennett 2008). Following glacial retreat, recolonization of previously glaciated areas occurred. In areas where residents of differing refugia came into contact, transition zones

*Email: ibradbur@me.com

**The first 2 authors contributed equally to this work

formed due to varying combinations of limited dispersal, genome incompatibilities, and reinforcement (e.g. Bernatchez & Wilson 1998, Knowles & Richards 2005, Maggs et al. 2008).

In marine species several studies have identified transition zones or regions of secondary contact (Duvernell et al. 2003, Hickerson & Cunningham 2006, Maggs et al. 2008), suggesting that historical barriers to gene flow may persist, often in association with strong selection (e.g. Sotka et al. 2004, Bradbury et al. 2010a). The relative importance of historical and contemporary factors (i.e. vicariance vs. selection and dispersal) in determining contemporary connectivity remains unexamined in most marine species, despite the fact that the interpretation may dramatically alter conservation priorities and objectives (Waples et al. 2008). Resolving these influences on spatial structure is complicated by the fact that the time scales on which each function may dramatically differ, and as such the comparison of molecular loci that differ in their rate of mutation may likely be most informative.

Rainbow smelt *Osmerus mordax* (Mitchill) is a species of small pelagic fish found in coastal and freshwater systems throughout northeastern North America from New Jersey to Labrador (Nellbring 1989). Anadromous smelt spawn just above the tidal influence in coastal rivers and streams and the larvae develop in downstream estuaries (e.g. Ouellet & Dodson 1985, Bradbury et al. 2004). Phylogeographic studies focusing on mainland eastern North America have revealed the presence of 2 major mtDNA clades, hypothesized to be associated with glacial refugia on the Grand Banks off Newfoundland and along the Atlantic coastal plain off the eastern coast of the United States (Bernatchez 1997). Several studies have documented a region of secondary contact between these clades within the St. Lawrence estuary and Gulf of St. Lawrence (Baby et al. 1991, Taylor & Bentzen 1993, Bernatchez 1997). More recently, extensive sampling throughout Newfoundland and Labrador revealed evidence of moderate genetic structuring and limited dispersal (Bradbury et al. 2006a, 2006b, 2008a), though prior to this study, only 1 glacial clade had been observed in Newfoundland.

Here we evaluate potential influences of historical isolation, selection, and differing mutational dynamics among loci on molecular divergence by examining both mtDNA sequence variation (ND5) and 11 microsatellite loci in an estuarine fish, the rainbow smelt *Osmerus mordax*, from coastal Newfoundland, Canada, an area adjacent to a proposed glacial

refuge. We identify a novel region of secondary contact between 2 refugial lineages and highlight the role that Pleistocene landscape structure may play on contemporary genetic connectivity.

MATERIALS AND METHODS

Study area

Overall, 22 locations encompassing Newfoundland and Labrador were sampled during the period from 2002 to 2006 with a few locations sampled in multiple years (total = 26 samples, Fig. 1, Table 1). Newfoundland's coastline is characterized by numerous large embayments (Fig. 1) of varying size and shape. Sampling locations were distributed around the entire coastline and were typically in small coastal streams generally <5 m in width and <1 m depth (see Bradbury et al. 2006b, 2008b). Adult fish were collected with dip and fyke nets from coastal streams or rivers during the spawning period, and pectoral or caudal fin clips were sampled and immediately placed in 95% ethanol. See Bradbury et al. (2008a) for previous analysis of these samples.

Laboratory methods

DNA was extracted following the protocol of Elphinstone et al. (2003), modified to work with a 96-well filter plate and automated on a robotic liquid handling system (Perkin Elmer). For mitochondrial DNA (mtDNA) analysis, a ~522 bp portion of the ND5 gene was amplified in 20 µl volumes as follows: 2 µl template DNA, 2.0 µl 10× PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl), 2.0 µl 2 mM dNTPs, 0.60 µl 10 µM forward (5'-TCT GAC CCA AAA CGA CAT CA-3' or 5'-CCA CGC CAG TAT CTG CCT A-3') and reverse primers (5'-TAA GGC AAG AAT AAC GGC AA-3') based on deposited ND5 *Osmerus mordax* GenBank sequences (see Pigeon et al. 1998; accession nos. AF034751-AF034752), and 1 U *Taq* DNA polymerase (NEB). PCR products were visualized on 0.8% agarose gels stained with Gel Green (Biotium) and cleaned with the Exo-SAP protocol (USB) prior to sequencing. Sequences were run on an automated DNA sequencer ABI3730XL (ABI). All samples were sequenced bi-directionally, edited in the program SEQUENCHER (Gene Codes) and aligned using CLUSTALX v. 1.83 (Thompson et al. 1997).

For microsatellites, the following 13 loci were used: Omo1, Omo2, Omo3, Omo4, Omo5, Omo6a,

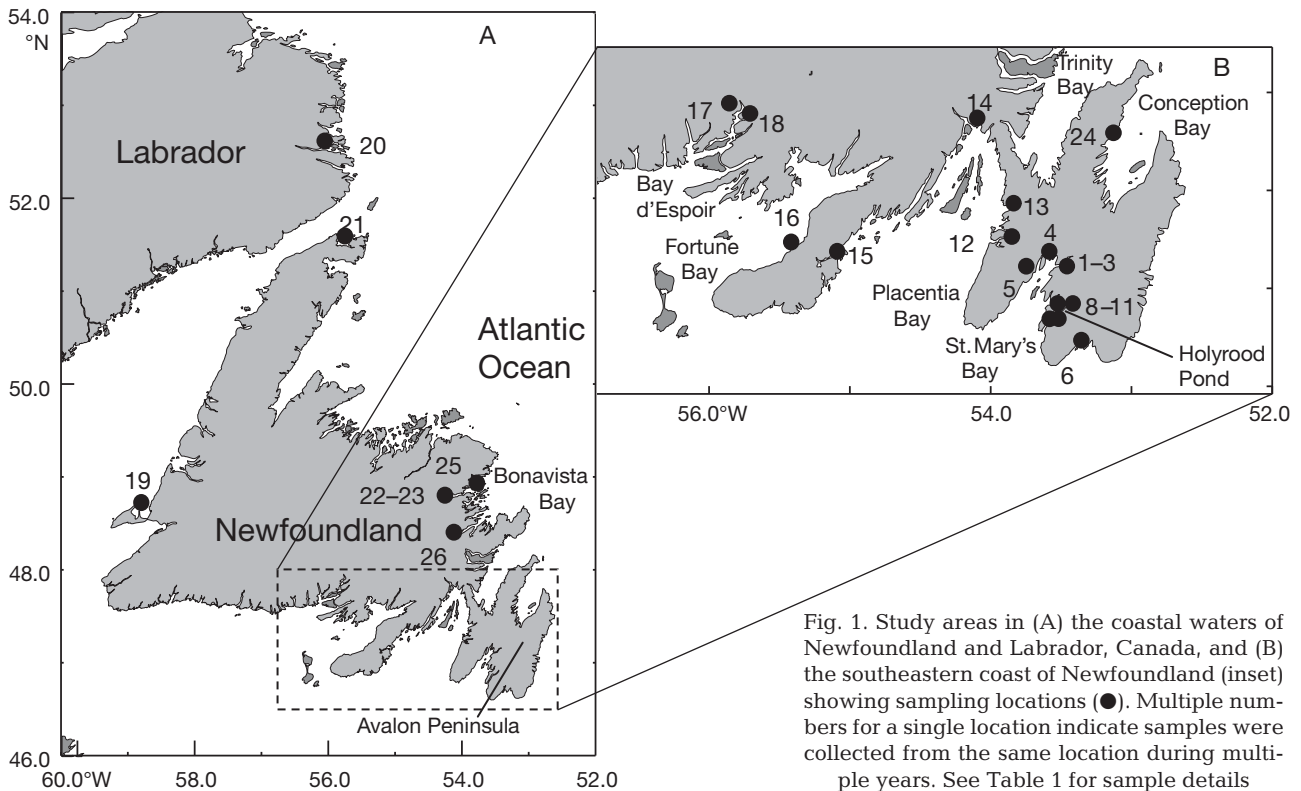


Fig. 1. Study areas in (A) the coastal waters of Newfoundland and Labrador, Canada, and (B) the southeastern coast of Newfoundland (inset) showing sampling locations (●). Multiple numbers for a single location indicate samples were collected from the same location during multiple years. See Table 1 for sample details

Omo9, Omo11, Omo13, Omo14, Omo15, and Omo16 (Coulson et al. 2006). Primers for Omo6a amplified a second locus, Omo6b, which was also scored and examined. Individuals were genotyped using PCR conditions of 5 or 10 μ l volumes containing 20 to 100 ng DNA, 1.5 mM $MgCl_2$, 80 μ M each dNTP, 0.5 U *Taq* DNA polymerase (New England Biolabs), 0.3 μ M of each primer (forward primers were end-labeled with HEX or ROX dye), and 10 \times PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl). Two temperature profiles were used for touchdown to allow for the possibility of multiplex PCRs. Touchdown PCR conditions were as follows: 94°C for 2 min, followed by 4 to 5 cycles of 94°C for 30 s, program specific touchdown annealing temperatures (T_a) minus 1°C per cycle for 30 s, 72°C for 30 s, followed by 25 to 26 cycles where the T_a was held constant at 4°C below the starting temperature. A final extension was held at 72°C for 5 min. Reactions were run on Eppendorf thermocyclers and imaged on an FMBioII system (Hitachi Genetic Systems). See Coulson et al. (2006) for further details regarding PCR conditions. This work builds on a previous microsatellite study which examined variation in 8 of these loci (Omo1, Omo2, Omo3, Omo4, Omo5, Omo9, Omo15, and Omo16).

Data analysis

ND5 was sequenced in 170 individuals from 18 locations (Table 1), producing a sequence of 522 bp spanning 3 of the diagnostic sites for the 'A' and 'B' lineages (see Bernatchez 1997). A reference individual from each of the 'A' and 'B' clades (provided by J. Dodson, Université Laval) was also sequenced. Unweighted maximum-likelihood distances were used to derive a median-joining network (MJN) (Bandelt et al. 1999) using the program NETWORK v. 4.0.5.1 (www.fluxus-engineering.com). Differentiation statistics were calculated using ARLEQUIN v. 3.5.1.2 (<http://cmpg.unibe.ch/software/arlequin35/>) as were analyses of molecular variance (AMOVA) using the identified mtDNA clades or microsatellite based STRUCTURE clusters (see below) as groups. Associations between mtDNA and microsatellite data were examined using Mantel tests implemented in PASSAGE v. 1 (Rosenberg & Anderson 2011).

Microsatellites were examined for the presence of null alleles and large allele dropout using MICRO-CHECKER (van Oosterhout et al. 2004). Tests for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were performed using FSTAT (v. 2.9.3.3, Goudet 1995) and GENEPOP (Raymond &

Table 1. Locations for rainbow smelt samples used for microsatellite and mtDNA analysis from coastal Newfoundland and Labrador collected from 2002 to 2006. H_o : observed heterozygosity; H_e : expected heterozygosity; ns: not sequenced as spatial or temporal replicates. ^aLocation found within St. Mary's Bay

Sample no. and name (year)	Location	Microsatellites				mtDNA			
		N	No. of alleles	H_o	H_e	N	Haplotype diversity	Haplotype A	Haplotype B
1. Salmonier River (02)	St. Mary's Bay	80	9.077	0.642	0.647	17	0.221	2	15
2. Salmonier River (03)	St. Mary's Bay	94	8.538	0.642	0.634	ns	ns	ns	Ns
3. Salmonier River (06)	St. Mary's Bay	93	9.385	0.672	0.657	ns	ns	ns	Ns
4. Colinet River (03)	St. Mary's Bay	94	9.308	0.657	0.646	10	0.222	0	10
5. North Harbour River (04)	St. Mary's Bay	94	8.923	0.650	0.634	11	0.182	0	11
6. Biscay Bay River (03)	St. Mary's Bay	94	8.692	0.639	0.648	13	0	0	13
7. Biscay Bay River (06)	St. Mary's Bay	93	8.231	0.600	0.628	ns	ns	ns	Ns
8. Holyrood Pond Brook (05)	Holyrood Pond	94	9.615	0.598	0.658	16	0	0	16
9. Holyrood Pond Park(04)	Holyrood Pond ^a	94	8.846	0.634	0.639	ns	ns	ns	Ns
10. Deer Pond Brook (04)	Holyrood Pond ^a	41	7.538	0.606	0.601	ns	ns	ns	Ns
11. Pathend Brook (05)	Holyrood Pond ^a	94	8.846	0.646	0.648	ns	ns	ns	Ns
12. Southeast Placentia (05)	Placentia Bay	94	8.462	0.608	0.617	13	0.385	0	13
13. Long Harbour Brook (04)	Placentia Bay	79	7.077	0.596	0.592	11	0	0	11
14. North Harbour River (03)	Placentia Bay	80	7.462	0.602	0.639	12	0	0	12
15. Salt Pond Brook (03)	Placentia Bay	92	8.769	0.631	0.687	15	0	14	1
16. Garnish River (04)	Fortune Bay	74	10.154	0.719	0.775	9	0	9	0
17. Conne River (04)	Bay d'Espoir	94	10.385	0.599	0.618	8	0	7	1
18. Little River (04)	Bay d'Espoir	94	9.385	0.618	0.625	9	0.389	7	2
19. Point Amal (04)	Port aux Port Peninsula	80	12.846	0.784	0.794	12	0.709	11	1
20. Mary's Harbour (04)	Labrador	55	7.231	0.583	0.604	12	0.167	12	0
21. St. Anthony (03)	Northern Peninsula	76	8.692	0.646	0.704	3	0.667	2	1
22. Gambo River (03)	Bonavista Bay	80	9.538	0.585	0.658	10	0.378	0	10
23. Gambo River (05)	Bonavista Bay	80	9.308	0.564	0.664	ns	ns	ns	ns
24. Salmon Cove River (05)	Conception Bay	94	9.385	0.583	0.618	18	0.399	1	17
25. Chuff Brook (06)	Bonavista Bay	93	10.385	0.632	0.686	ns	ns	ns	ns
26. Traverse Pond (06)	Bonavista Bay	93	9.615	0.611	0.660	7	0.286	1	6

Rousset 1995). F -statistics and significance were calculated using FSTAT and ARLEQUIN, and the calculated G -statistics (G_{ST}) were corrected for heterozygosity (G'_{ST}) following Hedrick (2005) using MICROSATELLITE ANALYSER (MSA; Dieringer and Schlötterer 2003). In addition, Rho_{ST} was calculated using GENEPOP (Rousset 2008), and Jost's D was estimated using Eq. 11 of Jost (2008) to allow comparison with G_{ST} and G'_{ST} . As complex hierarchical spatial structure is likely present in this species, we followed the hierarchical island model approach outlined by Excoffier et al. (2009) to determine the potential influence of selection on our microsatellite loci, as implemented in ARLEQUIN v. 3.5.1.2 (Excoffier & Lischer 2010). This software calculates F_{ST} and heterozygosity from the data, which are compared with null predictions based on coalescent simulations assuming a hierarchical island model. The hierarchical island model is preferred over the island model approach (Beaumont & Nichols 1996) because it significantly reduces the number of false positives present due to hierarchical population structure. The simulations were run using recommended values

(Excoffier & Lischer 2010) of 100 demes, 50 groups, and 20 000 simulations. Various numbers of groups were examined (2, 5, 25) with no change in results. Given the low number of loci examined for tests of this sort, the potential for false positives remains relatively high and the results are therefore tentative.

Bayesian clustering of the microsatellite data (excluding Omo11) was done using STRUCTURE v. 2.0 (Pritchard et al. 2000) to examine whether the predictions of 2 predominant groups present in mtDNA and Omo11 were consistent with results from remaining microsatellite loci. This approach uses assumptions of HWE and linkage equilibrium between loci, introduces population structure and assigns populations that are not in linkage equilibrium using a Markov chain Monte Carlo (MCMC) algorithm to estimate the number of populations (K). The algorithm was run 3 times with $K = 1$ to 22 to ensure convergence of values and with a burn in of 100 000 reps, 500 000 reps after burn in. The value of K that represents the actual number of populations was estimated using ΔK values calculated following Evanno et al. (2005).

RESULTS

mtDNA analysis

The haplotype network based on the 522 bp region of the ND5 gene revealed the presence of the 2 major mitochondrial clades (Fig. 2; Table 1) previously identified for rainbow smelt (Baby et al. 1991, Taylor & Bentzen 1993, Bernatchez 1997). Among the 170 individuals included in the sequence alignment, 13 haplotypes were found. Based on the known individuals, the 'B' haplotypes predominated in SE Newfoundland (Avalon Peninsula to Bonavista Bay), while the 'A' haplotypes predominated along the south shore and northward to the northern Peninsula and Labrador. The highest haplotype diversity was found in Point Amal on the SW coast of the island (Table 2). AMOVA results (Table 3) indicated the largest component of the variation (79%) was explained when individuals are grouped by putative glacial race, followed by 20% of the variance present within each sample.

Microsatellite analysis

Average microsatellite heterozygosities within populations were variable ranging from 0.58 to 0.78

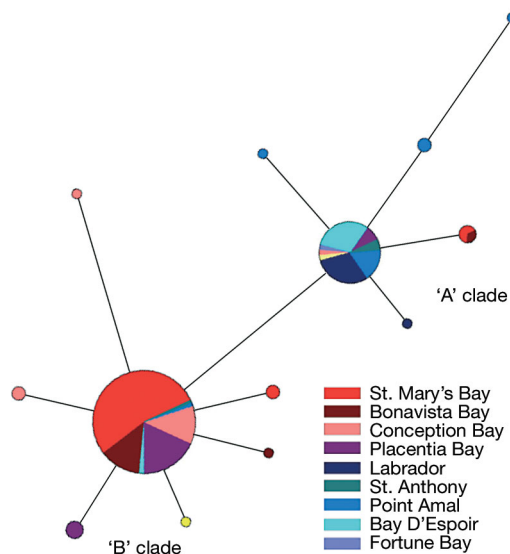


Fig. 2. Median-joining haplotype network based on 522 bp of ND5 mitochondrial sequence. The node size is proportional to the frequency of each haplotype and color codes represent broad regions of interest. For each of the groups of haplotypes ('A' vs. 'B' clade), yellow colors represent known positive 'A' or 'B' individuals (contributed by J. Dodson, Université Laval)

(Table 1; Table S1 in the supplement available at www.int-res.com/articles/suppl/m438p207_supp.pdf). Deviations from HWE were sporadic and not associated with locations or loci with the exception of Omo14, which was significantly out of HWE in >90% of comparisons (Table S2 in the supplement available at www.int-res.com/articles/suppl/m438p207_supp.pdf). Evidence of linkage disequilibrium was detected between Omo6a and Omo6b in ~90% of comparisons. Omo16 and Omo3 displayed some evidence of linkage in <50% of comparisons. Instances of null alleles estimated using MICRO-CHECKER (van Oosterhout et al. 2004) were rare and not consistently associated with a specific locus or population. Nulls were estimated to be present in ~5% of all comparisons (loci \times location) and when present the estimated frequency was usually <5%. In light of these results, the analysis was conducted on a dataset containing 10 microsatellite loci, including Omo11 but excluding Omo14, Omo6b, and Omo16.

Overall genetic differentiation (F_{ST}) was 0.124 ± 0.04 (mean \pm SD) using all loci. With the exclusion of Omo11, overall differentiation dropped slightly to 0.113 ± 0.03 . Average microsatellite pairwise F_{ST} among samples from each of the mtDNA clades was 0.087 ± 0.023 compared with 0.032 ± 0.021 within each of the groups. Single-locus estimates of G_{ST} were 0.060 to 0.150 for the 10 non-outlier loci (see below) but were higher (0.224) for Omo11 (Fig. 3a). Locus-specific G_{ST} was negatively related to heterozygosity (Fig. 3a, $p < 0.001$, $R^2 = 0.60$). The ranking changed upon standardization for heterozygosity (Fig. 3b) and there was a positive relationship between heterozygosity and G'_{ST} or D (Fig. 3b,c). Locus specific Rho_{ST} was not associated with heterozygosity and identified Omo3, Omo5, and Omo15 as the most divergent among the samples, though none were clearly elevated or outliers (Table S1). Hierarchical tests for selection yielded results suggestive of selection acting on 2 loci (Omo11 and Omo4; Fig. 4a). Omo11 displayed significantly elevated levels of differentiation consistent with directional selection and Omo4 displayed reduced divergence consistent with the possibility of balancing selection (Fig. 4a). These were significant at $\alpha = 0.01$, but again given small numbers of loci the potential for false positives is quite high. Examination of allele frequencies at Omo11 revealed 2 predominant alleles with relatively discrete spatial distributions (Fig. 4b,c). One allele, 190 bp, was at high frequency in St. Mary's Bay (>70%) and declined in frequency with distance from the Bay both to the north and west (Fig. 4c). These declines coincided with an increase

Table 2. Estimates of genetic differentiation using mtDNA for rainbow smelt samples from coastal Newfoundland showing F_{ST} values above the diagonal and significance values below. **Bold** values indicate significant differences at $p < 0.05$. Samples collected from the same location in subsequent years were not included. See Table 1 and Fig. 1 for sample details and locations, respectively

	Sample location																									
	1	4	5	6	8	12	13	15	16	17	18	19	20	21	22	24	26									
1	0.260	0.019	0.029	0.040	0.057	0.096	0.026	0.712	0.648	0.727	0.541	0.602	0.771	0.391	0.028	0.004	-0.109									
4	0.297	0.999	-0.110	0.043	0.068	0.100	0.023	0.917	0.895	0.925	0.700	0.693	0.913	0.644	-0.003	-0.012	0.021									
5	0.493	0.406	0.454	0.016	0.036	0.112	0.000	0.929	0.913	0.934	0.724	0.715	0.920	0.686	0.004	-0.004	0.041									
6	0.510	0.341	0.400	0.999	0.000	0.167	0.000	1.000	1.000	1.000	0.791	0.763	0.962	0.817	0.027	-0.002	0.095									
8	0.063	0.128	0.165	0.226	0.092	0.195	0.000	1.000	1.000	0.867	0.691	0.701	0.966	0.845	0.050	0.012	0.128									
12	0.494	0.471	0.999	0.999	0.999	0.207	0.145	0.858	0.828	1.000	0.773	0.744	0.958	0.616	0.090	0.072	0.092									
13	0.999	0.009	0.004	0.001	0.002	0.001	0.001	1.000	1.000	1.000	0.000	0.000	0.000	0.791	-0.039	-0.013	0.069									
15	0.999	0.009	0.001	0.000	0.000	0.000	0.001	0.999	0.999	0.000	0.000	0.000	0.000	0.000	0.856	0.743	0.675									
16	0.004	0.001	0.001	0.000	0.000	0.001	0.001	0.999	0.999	0.000	0.000	0.000	0.000	-1.000	0.818	0.690	0.538									
17	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.999	0.999	0.000	0.000	0.000	0.000	-1.000	0.868	0.756	0.707									
18	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.999	0.999	0.548	0.306	0.025	0.132	-0.241	0.669	0.592	0.471									
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.999	0.685	0.175	0.017	0.084	-0.039	0.678	0.641	0.526									
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.999	0.999	0.426	0.586	0.170	0.312	0.881	0.794	0.779									
21	0.099	0.064	0.043	0.028	0.018	0.016	0.021	0.999	0.999	0.426	0.001	0.586	0.000	0.040	0.580	0.463	0.256									
22	0.166	0.872	0.449	0.178	0.159	0.105	0.593	0.001	0.993	0.001	0.001	0.000	0.000	0.045	0.472	0.001	0.018									
24	0.306	0.674	0.516	0.481	0.473	0.031	0.579	0.001	0.126	0.002	0.000	0.000	0.000	0.045	0.472	0.001	-0.023									
26	0.999	0.450	0.371	0.340	0.322	0.137	0.431	0.038	0.258	0.010	0.012	0.001	0.001	0.209	0.404	0.500										

in the frequency of the second common allele (166 bp), which was widely distributed but the common allele on the south and northeast coasts. In contrast to these uniform patterns, the west coast showed a mixture of alleles, and the sample from Labrador lacked both of the 2 common island alleles.

Comparison of spatial trends

STRUCTURE analysis, excluding Omo11, using $K = 1$ to 22 indicated strong support for 2 major clusters (Fig. 5) evident in the largest ΔK (Evanno et al. 2005). The majority of individual Q-values were associated with these 2 main groups and few intermediate individuals were observed (Fig. S1 in the supplement available at www.int-res.com/articles/suppl/m438p207_supp.pdf). However, the likelihood values plateaued at approximately $K = 15$ or 16, suggesting that microsatellites (excluding Omo11) resolved finer-scale differentiation (Fig. 5). AMOVA of the microsatellite data suggested the major 2 groupings were significant and explained 3.67% ($p < 0.001$) of the total variance (Table 3).

Multidimensional scaling of F_{ST} values revealed a clear tendency for the Avalon Peninsula locations to cluster together based on the first axis, though the exact composition of clusters varied with the marker (Fig. 6). The spatial distribution of the common mtDNA clades was consistent with a transition zone along the south and northeast coasts (Fig. 7a). Similar clines, though not identical, were observed in the frequencies of the common Omo11 alleles (Fig. 7b) and the distribution of the 2 STRUCTURE clusters (see 'Data analysis', Fig. 7c). Mantel tests examined associations among the pairwise differentiation statistics and revealed significant associations between nuclear and mtDNA loci. Differentiation observed among mtDNA haplotypes was significantly correlated with differentiation at Omo11 ($r = 0.26$, $t = 3.33$, $p < 0.001$) and neutral loci though the correlation was weak ($r = 0.16$, $t = 2.14$, $p = 0.032$).

DISCUSSION

Interpreting the relative roles of historical and contemporary processes in determining genetic structure of natural populations remains a recurring challenge in marine population genetics and

Table 3. AMOVA results comparing variation in (A) 2 mtDNA clades shown in Fig. 2 and (B) 2 microsatellite loci clusters identified with the STRUCTURE analysis in smelt populations sampled in Newfoundland coastal waters. See Table 1 and Fig. 1 for sample details and locations, respectively. All F -statistics were significant at $p < 0.0001$

Grouping	Source of variation	df	Sum of squares	Variance component	Percent of total	F_{CT}	F_{SC}	F_{ST}
A) mtDNA clades								
	Among groups	1	49.3	0.7656	78.6	0.7861		
	Among populations within groups	15	3.97	0.0064	0.66		0.0310	
	Among populations	151	30.5	0.2020	20.7			0.7926
	Total	167	83.7	0.9741	100			
B) Microsatellite loci excluding Omo11								
	Among groups	1	306.37	0.1172	3.67	0.0367		
	Among populations within groups	24	1085.55	0.2467	7.71		0.0800	
	Among populations	4460	12627.34	2.8312	88.63			0.1131
	Total	4485	14019.26	3.1946	100			

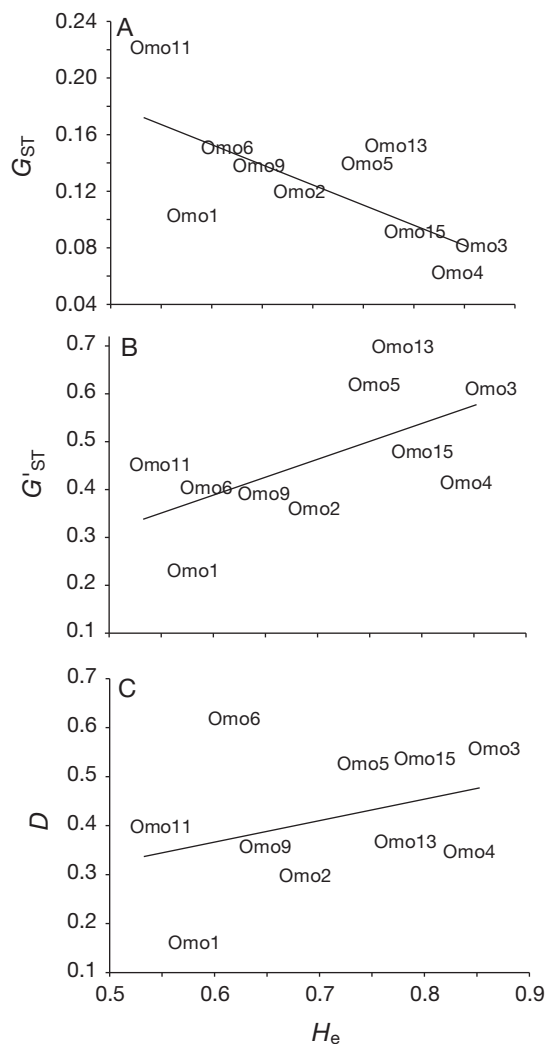


Fig. 3. Relationships between (A) G'_{ST} , (B) standardized G'_{ST} (Hedrick 2005), (C) Jost's D (Jost 2008) and expected heterozygosity (H_e) based on 11 tetranucleotide microsatellite loci for anadromous rainbow smelt from coastal Newfoundland. See 'Data analysis' for further details of the method

conservation (Waples et al. 2008). Increasingly, historical vicariant processes are being implicated in the formation and maintenance of contemporary patterns of spatial connectivity (e.g. Duvernell et al. 2003, Orsini et al. 2008, McCusker & Bentzen 2010). Our results suggest that broad-scale contemporary genetic differentiation of anadromous smelt largely reflects historical glacial isolation and subsequent recolonization. We demonstrate the presence of repeated transitional zones in mtDNA and microsatellite clades in smelt inhabiting Newfoundland and Labrador in the absence of any contemporary physical or hydrographic barriers to dispersal. The isolation of the southern Avalon Peninsula is consistent with a hypothesis of a Pleistocene Grand Banks refugium and secondary contact along the south and north coasts of Newfoundland. Overall, our results support the growing consensus that contemporary genetic structure in northern species largely reflects the influences of Pleistocene glaciation and subsequent range expansions.

Awareness is increasing that isolation in Pleistocene refugia and postglacial recolonization may be a principle determinant of contemporary genetic connectivity in many northern species (Bernatchez & Wilson 1998, Wilson & Veraguth 2010). During periods of glaciation, lower sea temperatures and a reduction of coastal and shelf habitat restricted many species to mid latitude refugia (e.g. Bigg et al. 2008). Many continental shelf species, such as Atlantic cod *Gadus morhua*, display large genetic discontinuities between the eastern and western Atlantic that are indicative of a lack of suitable northern habitat during periods of glaciation (Verspoor 2005, Bigg et al. 2008, Bradbury et al. 2010b). Following glacial retreat, range expansions reestablished contact and several

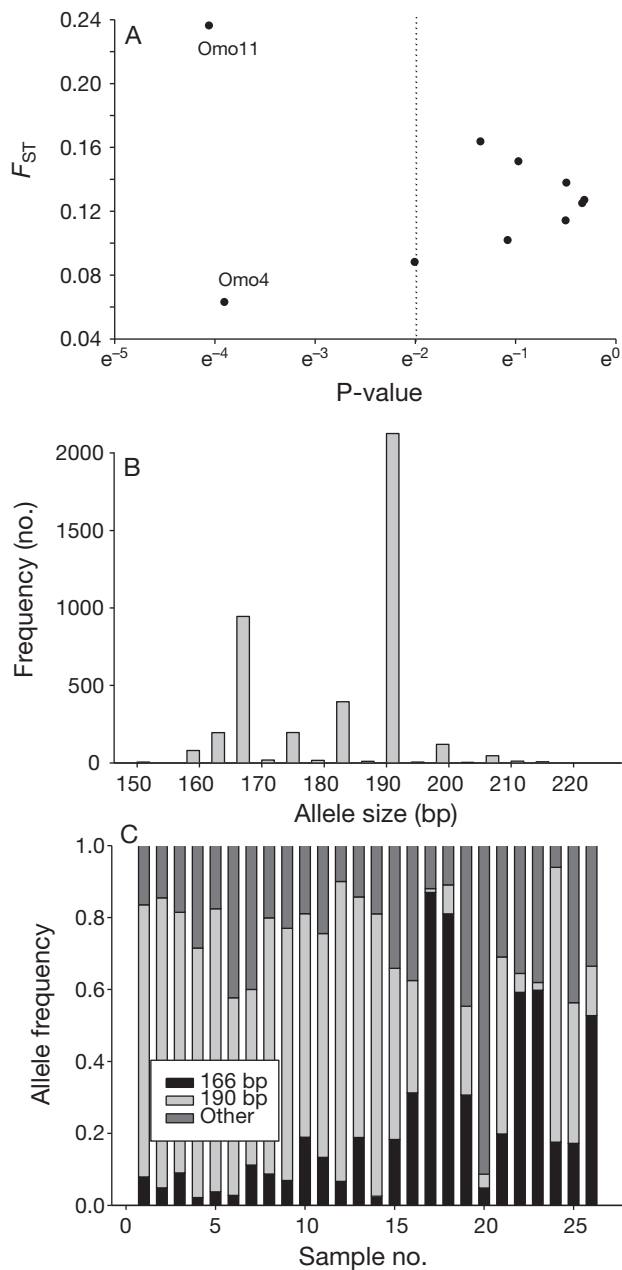


Fig. 4. (A) Test for selection using the hierarchical island model as implemented in ARLEQUIN v. 3.5.1.2 (Excoffier & Lischer 2010). (B) Frequency distribution of alleles at Omo11 overall populations. (C) The proportion of common alleles in each sample. The dotted line in (A) represents the threshold for significance

instances have been reported where secondary contact of these eastern and western Atlantic races occurs in the western Atlantic off Newfoundland (e.g. Atlantic salmon *Salmo salar*, Verspoor 2005). However, the impact of Pleistocene glaciation seems to vary dramatically among coastal species (e.g. Wares & Cunningham 2001, Maggs et al. 2008) and

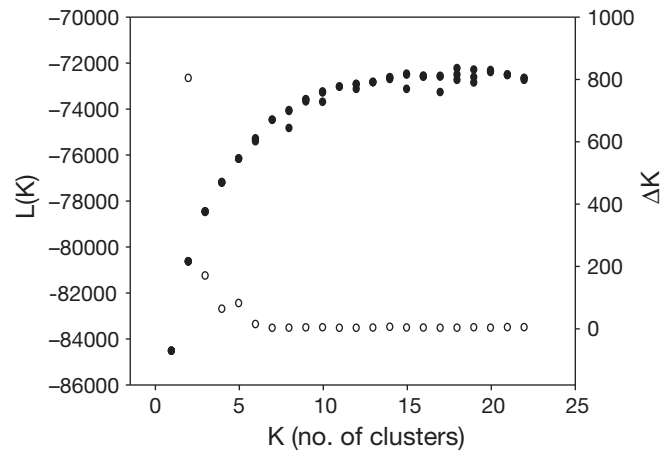


Fig. 5. Results of STRUCTURE analysis on non-outlier loci (i.e. Omo11 excluded). (●) $L(K)$ with values of K from 1 to 22 from 3 replicates run with 100 000 iterations burn in and 500 000 iterations following burn in. (○) ΔK values calculated following Evanno et al. (2005)

likely depends on specific thermal tolerances and habitat requirements. In smelt, Bernatchez (1997) hypothesized the existence of 2 glacial refugia along the Atlantic coast of North America that diverged ~700 000 yr before present (YBP); one on the Atlantic coastal plain thought to be associated with the 'B' mitochondrial clade, and the other on the Grand Banks off Newfoundland, assumed to be the source of the 'A' clade. Contrary to previous studies (Baby et al. 1991, Taylor & Bentzen 1993, Bernatchez 1997), the mtDNA results obtained in this study indicate that both mitochondrial lineages are present in Newfoundland. As previous work had only observed the 'A' clade in Newfoundland (Bernatchez 1997), this clade was thought associated with a refugium located on the exundated Grand Banks of Newfoundland. The documentation of the 'B' clade in eastern Newfoundland directly adjacent to the proposed Grand Banks refugium suggests that a revision of the current recolonization hypothesis is required.

Both microsatellite and mtDNA analysis support a hypothesis of limited introgression among the 2 groups observed. Mechanistically this could result from limited dispersal, strong barriers to introgression, reinforcement, or some combination of the above. As dispersal in Newfoundland smelt has been shown to be very limited during both the larval (Bradbury et al. 2006b) and adult stages (Bradbury et al. 2008a), low levels of gene flow may reduce the potential for introgression. Previous work (Bradbury et al. 2008a) observed limited gene flow and strong genetic isolation in Newfoundland smelt, but failed to resolve a historical signature because the number

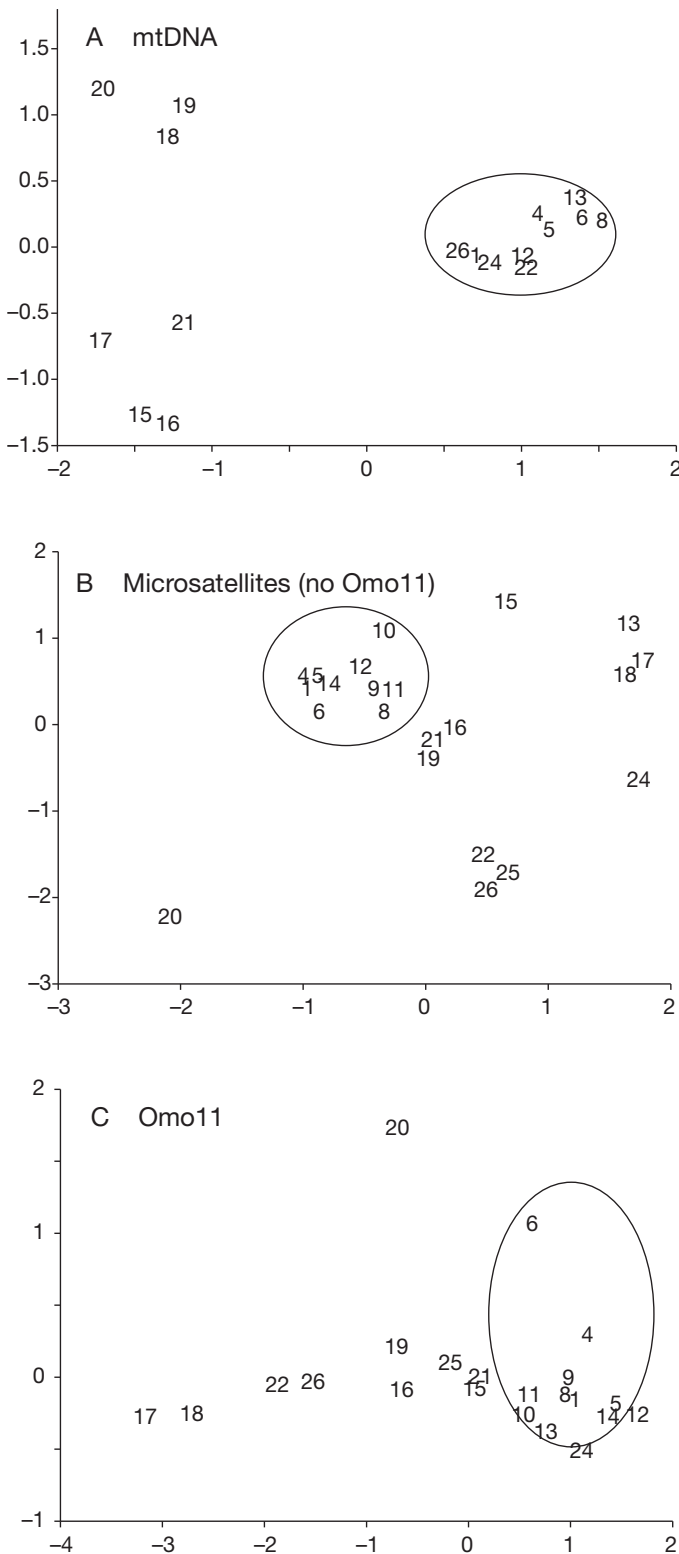


Fig. 6. Two-dimensional multidimensional scaling (MDS) plot of differentiation present at (A) mtDNA, (B) microsatellites not including Omo11, and (C) Omo11. Numbers correspond to sample number shown in Table 1. Ellipses represent approximate extent of distinct Avalon Peninsula clade

of loci was low ($n = 8$) and did not include Omo11 or mtDNA sequence data. The general observation of limited dispersal among smelt populations and the absence of any physical or hydrographic barriers to explain the observed clines support the hypothesis that these are historical in nature.

Nonetheless, the possibility of selection associated with an unidentified environmental gradient remains, and the role of selection and adaptation in structuring marine populations is increasingly receiving attention across a suite of taxa (e.g. Pogson 2001, Gaggiotti et al. 2009, Bradbury et al. 2010a,b). In Atlantic cod, parallel clines in multiple temperature associated genes have recently been reported in the eastern and western Atlantic consistent with temperature associated adaptation over fine spatial scales (Bradbury et al. 2010b). Our application of the hierarchical test for elevated differentiation and selection suggests that Omo11 may be experiencing directional selection and Omo4 stabilizing selection. Interestingly, similar results were obtained by Gaggiotti et al. (2009) in a study of Atlantic herring *Clupea harengus* where 2 loci of 8 were associated with selection. In that case, clear associations were observed between outlier allele frequencies and salinity and feeding migrations which supported the hypothesis of selection. In the absence of such associations here, it remains unclear what role the presence of selective barriers to introgression play in the structuring the observed clines. Furthermore, the observation that standardizing for heterozygosity (e.g. G'_{ST} value or D) removed the outlier status of Omo11 supports a neutral mechanism. Nonetheless, selective barriers and reinforcement have been implicated in the isolation of postglacially colonized sympatric smelt populations. These examples include evidence of reduced fertilization success in hybrid crosses and low survival of immigrating individuals (Bradbury et al. 2008b, 2010a), both of which support a role for selection in reinforcing isolation in contemporary populations.

Estimates of the strength of selection required to maintain the observed clines (mtDNA and Omo11) could represent upper bounds on the magnitude of selection required in the absence of barriers to gene flow. The spatial pattern at Omo11 can be treated as 2 separate clines, one to the north and another to the west originating at the center of St. Mary's Bay, and as such the strength of contemporary selection required to maintain the observed clines can be estimated. At equilibrium the balance between gene flow and selection should be approximated by $w^2 = \sigma^2 s^{-1}$ where w is the clinal width, σ the standard devi-

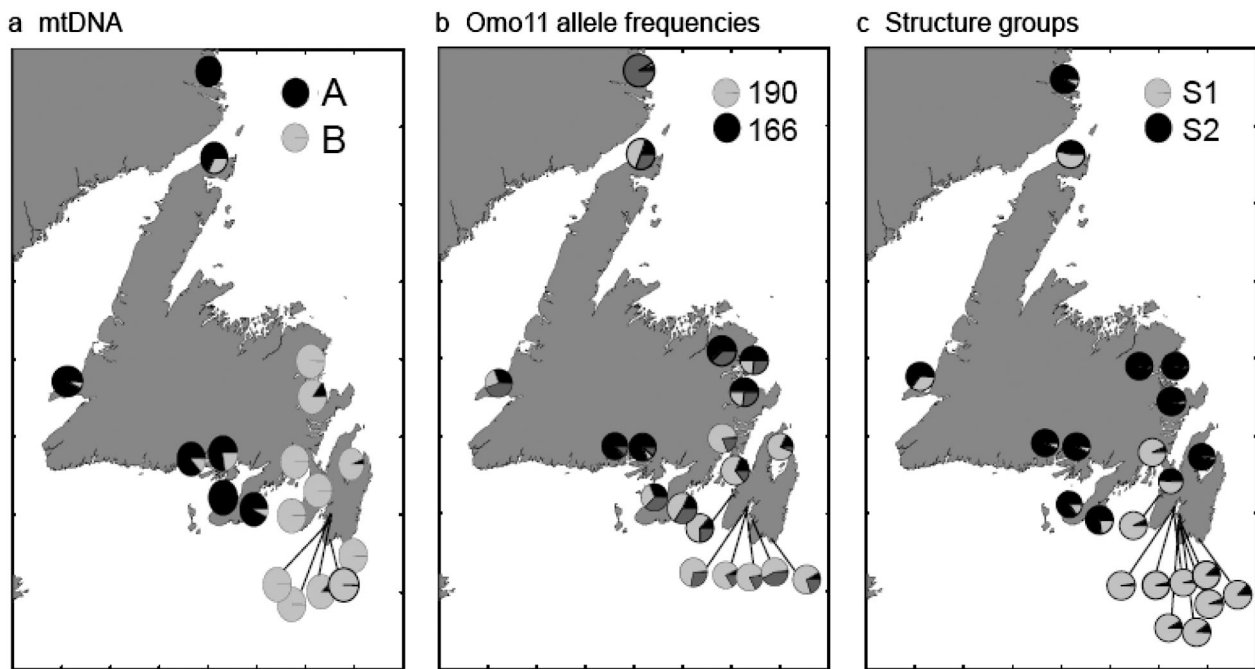


Fig. 7. Spatial distribution of the (a) 2 predominant mtDNA haplotypes, (b) common Omo11 allele frequencies, and (c) 2 groups identified using the neutral loci (excluding Omo11) and STRUCTURE (S) clusters with $K = 2$. (●, ●) common alleles or groups, (●) pooled frequency of other alleles present (B only)

ation in dispersal distance from parents to offspring, and s the strength of selection (Barton & Hewitt 1985). Based on estimates of dispersal for smelt encompassing all life stages in these habitats (1 to 10 km, Bradbury et al. 2006b, 2008a,b), we expect selection values of <0.01 would be necessary to maintain the observed gradients at Omo11. As values of selection in this range are commonly reported in natural systems (Endler 1977, Kingsolver et al. 2001, Sotka et al. 2004), the values estimated here certainly seem reasonable. Nonetheless, in the absence of environmental data or obvious gradients and given the clear influence of mutation rate on divergence, such inferences regarding the role of selection should be made cautiously.

Although similar, the geographic patterns identified by the marker types (Omo11, non-outlier microsatellites and mtDNA) were not identical. This was most evident along the northeast coast, where both Omo11 and the non-outlier microsatellites suggested a secondary contact zone in the Conception Bay to Bonavista Bay area, whereas the mtDNA data suggested a contact zone somewhere to the north of this area in a region not sampled in this study. This disagreement may reflect the differential introgression of nuclear and mitochondrial markers. The differences in the relative ability of the various markers and loci to resolve signatures

of historical isolation or in contemporary dispersal patterns may in part be due to differences in mutation rate. Given the expected differences in mutation rate between microsatellites and mtDNA (e.g. 10^{-2} to 10^{-6} for microsatellite loci, Schug et al. 1997, Weber & Wong 1993; $\sim 10^{-8}$ for mtDNA or $\sim 5.2 \times 10^{-6}$ when correcting for sequence length, Haag-Liautard et al. 2008), the assumption that differentiation revealed by microsatellite loci may reflect contemporary trends in gene flow to a greater degree is reasonable. Moreover, among microsatellites, mutation rates have been shown to vary with variation in mutation rate usually associated with microsatellite length or the presence of interruptions in a repeat sequence (Sainudiin et al. 2004). Omo11 contains an interruption, which may explain the observed difference with the remaining microsatellites. Attempts to correct for mutation rate support a mutational basis for differences among microsatellite loci (e.g. O'Reilly et al. 2004) because Omo11 did not display elevated divergence upon correction using either G'_{ST} value or D . As expected, loci characterized by high heterozygosity are likely to be more responsive to contemporary gene flow and a complete picture of historic and contemporary structuring forces may only be achieved when a variety of markers or loci representing a range of mutation rates are compared.

CONCLUSIONS

The interplay of spatial structuring forces in natural populations of non-model organisms remains notoriously difficult to disentangle. We suggest that present genetic spatial structure in northern aquatic species may be primarily due to range contractions and expansions associated with historical glaciation. The differing ability of individual loci and markers to resolve the various spatial structuring influences implicates mutation rate as an explanation. Thus, a complete picture of the forces that influence spatial genetic connectivity seems to be apparent only when a variety of markers representing a range of mutation rates are compared, and single approach/marker studies risk failing to resolve cryptic structure. Finally, in light of the documented secondary point of contact within Newfoundland, a re-evaluation of the colonization history of smelt in eastern North American will be required as previous hypotheses seem unlikely.

Acknowledgements. The authors thank all the summer students who assisted with this work and all members of the Marine Gene Probe Lab. L. Barrett advised on analysis and interpretation. We thank D. Heath, J. Hutchings and 3 anonymous reviewers for comments on earlier drafts. We are grateful to the local residents throughout coastal Newfoundland who assisted in locating and sampling smelt runs, in particular D. and T. Power of St. Catherine's. Funding was provided by the National Sciences and Engineering Research Council of Canada (NSERC), an NSERC Discovery grant to P.B., an NSERC Strategic Grant, the Canadian Department of Fisheries and Oceans, and the Killam Trust.

LITERATURE CITED

- Avice JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA
- Baby MC, Bernatchez L, Dodson JJ (1991) Genetic structure and relationships among anadromous and landlocked populations of rainbow smelt, *Osmerus mordax*, Mitchell, as revealed by mtDNA restriction analysis. *J Fish Biol* 39: 61–68
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu Rev Ecol Syst* 16:113–148
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B* 263:1619–1626
- Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St Lawrence River estuary (Quebec, Canada). *Mol Ecol* 6:73–83
- Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Mol Ecol* 7:431–452
- Bigg GR, Cunningham CW, Ottersen G, Pogson G, Wadley MR, Williamson P (2008). Ice age survival of Atlantic cod: agreement between palaeo-ecology models and genetics. *Proc R Soc Lond B* 275:163–172
- Bradbury IR, Campana SE, Bentzen P, Snelgrove PVR (2004) Synchronized hatch and its ecological significance in rainbow smelt *Osmerus mordax* in St. Mary's Bay, Newfoundland. *Limnol Oceanogr* 49:2310–2315
- Bradbury IR, Coulson MW, Campana SE, Bentzen P (2006a) Morphological and genetic differentiation in anadromous rainbow smelt *Osmerus mordax* (Mitchell): disentangling the effects of geography and morphology on gene flow. *J Fish Biol* 69:95–114
- Bradbury IR, Gardiner K, Snelgrove PVR, Campana SE, Bentzen P, Guan L (2006b) Larval transport, vertical distribution, and localized recruitment in anadromous rainbow smelt (*Osmerus mordax*). *Can J Fish Aquat Sci* 63: 2822–2836
- Bradbury IR, Campana SE, Bentzen P (2008a) Low genetic connectivity in an estuarine fish with pelagic larvae. *Can J Fish Aquat Sci* 65:147–158
- Bradbury IR, Campana SE, Bentzen P (2008b) Otolith elemental composition and adult tagging reveal spawning site fidelity and estuarine dependency in rainbow smelt. *Mar Ecol Prog Ser* 368:255–268
- Bradbury IR, Coulson MW, Cook AM, Bentzen P (2010a) Evidence for divergence and adaptive isolation in post-glacially derived bimodal allopatric and sympatric smelt populations. *Biol J Linn Soc* 101:583–594
- Bradbury IR, Hubert S, Higgins B, Borza T and others (2010b) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc Roy Soc Lond B* 277:3725–3734
- Coulson MW, Paterson IG, Green A, Kepkay R, Bentzen P (2006) Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). *Mol Ecol Notes* 6:942–944
- Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* 3: 167–169
- Duvernell DD, Schmidt PS, Eanes WF (2003) Clines and adaptive evolution in the *methuselah* gene region in *Drosophila melanogaster*. *Mol Ecol* 12:1277–1285
- Elphinstone MS, Hinten GN, Anderson MJ, Nock CJ (2003) An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Mol Ecol Notes* 3:317–320
- Endler JA (1977) Geographic variation, speciation, and clines. Princeton University Press, Princeton, NJ
- 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
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resources* 10: 564–567
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298
- Gaggiotti O, Bekkevold D, Jorgensen H, Foll M, Carvalho G, Andre C, Ruzzante DE (2009) Disentangling effects of evolutionary, demographic & environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution* 63:2939–2951

- Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* 86:485–486
- Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol* 6:e204
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59:1633–1638
- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B* 359: 183–195
- Hickerson MJ, Cunningham CW (2006) Contrasting Quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. *Evolution* 59:344–360
- Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Mol Ecol* 17:4015–4026
- Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D and others (2001) The strength of phenotypic selection in natural populations. *Am Nat* 157:245–261
- Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Mol Ecol* 14:4023–4032
- Maggs CA, Castilho R, Foltz D, Henzler C and others (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* 89:S108–S122
- McCusker MR, Bentzen P (2010) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolfish (*Anarhichas lupus*), across the North Atlantic Ocean. *Mol Ecol* 19:4228–4241
- Nellbring S (1989) The ecology of smelts (genus *Osmerus*): a literature review. *Nord J Freshw Res* 65:116–145
- O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* 13:1799–1814
- Orsini L, Corander J, Alasentie A, Hanski I (2008) Genetic structure in a butterfly metapopulation correlates better with past than present demographic structure. *Mol Ecol* 17:2629–2642
- Ouellet P, Dodson JJ (1985) Tidal exchange of anadromous rainbow smelt (*Osmerus mordax*) larvae between a shallow spawning tributary and the St. Lawrence Estuary. *Can J Fish Aquat Sci* 42:1352–1358
- Pielou EC (1991). After the ice age: the return of life to glaciated North America. University of Chicago Press, Chicago, IL
- Pigeon D, Dodson JJ, Bernatchez L (1998) A mtDNA analysis of spatiotemporal distribution of two sympatric larval populations of rainbow smelt (*Osmerus mordax*) in the St. Lawrence River estuary, Quebec, Canada. *Can J Fish Aquat Sci* 55:1739–1747
- Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (Pan I) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics* 157:317–330
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol Evol* 23:564–571
- Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249
- Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods Ecol Evol* 2:229–232. doi: 10.1111/j.2041-210X.2010.00081.x. Available at <http://www.passagesoftware.net/>
- Rousset F (2008) GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resources* 8:103–106
- Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. *Genetics* 168:383–395
- Schug MD, Mackay TFC, Aquadro CF (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nat Genet* 15:99–102
- Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Mol Ecol* 13:2143–2156
- Taylor EB, Bentzen P (1993) Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. *Evolution* 47: 813–838
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- van Oosterhout C, Hutchison WF, Wills DMP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Verspoor E (2005) Regional differentiation of North American Atlantic salmon at allozyme loci. *Fish Biol* 67:80–103
- Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: How can we do it better? *Fish Fish* 9:423–449
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2:1123–1128
- Wilson AB, Veraguth EI (2010) The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Mol Ecol* 19:4535–4553