Molecular Ecology (2007) **16**, 5030–5043 doi: 10.1111/j.1365-294X.2007.03559.x

Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*)

JULIAN J. DODSON,* SUZY TREMBLAY,* FRANÇOISE COLOMBANI,* JAMES E. CARSCADDEN† and FREDERIC LECOMTE‡

**Département de Biologie, Pavillon Vachon, Université Laval, Québec, Canada G1K 7P4, †Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, PO Box 5667, St John's, NL, Canada A1C 5X1, ‡Département Biologie Intégrative, Institut des Sciences de l'Évolution (UMR 5554 Université Montpellier 2), Station Méditerranéenne de l'Environnement Littoral, F3200 Sète, France*

Abstract

Trans-Arctic dispersals and population and range expansions during the Pleistocene enhanced opportunities for evolutionary diversification and contributed to the process of speciation within the capelin, a northern marine-fish complex exhibiting a circumpolar distribution. Capelin is composed of four highly divergent and geographically discrete mitochondrial DNA (mtDNA) clades (609 bp; cytochrome *b*). Two clades occur in the North Atlantic, one associated with Canadian Atlantic waters, including Hudson Bay, and the second distributed from West Greenland to the Barents Sea. Two additional clades occur in the Arctic and northeast Pacific Oceans, representing the most recent divergence within the capelin phylogenetic tree. Judged from mtDNA diversity, capelin populations comprising all clades experienced at least one demographic and spatial reduction–expansion episode during recent Pleistocene glaciations that imprinted their molecular architecture. The large contemporary populations in the northeast Pacific and Arctic Oceans exhibited significant genetic structure whereas no such structure was detected in the equally extensive North Atlantic clades. All clades are characterized by one or two prevalent mtDNA haplotypes distributed over the entire range of the clade. Assuming a Pacific ancestor for capelin, we infer that capelin dispersed on two separate occasions to the North Atlantic. A more recent event resulted in the isolation of eastern Pacific and Arctic clades, with the Arctic clade positioned for a potential third Atlantic invasion, as revealed by the presence of this clade in the Labrador Sea. The Labrador Sea is a potential contact zone for three of the four capelin clades.

Keywords: cytochrome *b*, demographic and spatial expansion, marine phylogeography, mtDNA, Pacific–Atlantic interchange, Pleistocene glaciations

Received 8 April 2007; revision received 17 August 2007; accepted 20 August 2007

Introduction

The contemporary distributions and genetic structures of high-latitude species have been strongly influenced by climatic oscillations during the Pliocene and Pleistocene epochs (Taylor & Dodson 1994; Avise *et al*. 1998; Wares & Cunningham 2001). Such climate oscillations may have a diversity of impacts on gene pools and evolution. On one hand, contraction into multiple isolated refuges during glacial advances may promote evolutionary diversification and allopatric speciation, if the partitioning is retained across several glacial cycles (Hewitt 1996). On the other hand, even if the intervals of isolation are sufficiently long to allow genetic differentiation to occur, the remixing of populations during periods of glacial retreat could obliterate any accumulated change (Coope 1979). Climate oscillations are generally viewed as impeding opportunities for genetic diversification at high latitudes as repeated disturbances and founder-flush cycles caused by repetitive cycles of glacial advance and retreat would eliminate intraspecific diversity at higher latitudes. (Bernatchez & Wilson 1998; Jansson & Dynesius 2002).

One of the more spectacular examples of biotic range extensions driven by climate oscillations occurred during the Late Pliocene when the opening of the Bering Strait,

Correspondance: J. J. Dodson, Fax: 418-656-2043; E-mail: julian.dodson@bio.ulaval.ca

coupled with a warming event, allowed temperate marine species in the Pacific and the Atlantic to interchange and gain access to the Arctic basin (Briggs 1970, 1995; Vermeij 1991). Based on fossil records, particularly those of marine molluscs, this trans-Arctic interchange involved hundreds of marine species with eight times as many Pacific species migrating to the Atlantic than species of Atlantic origin migrating to the Pacific (Vermeij 1991). In contrast to the generally held view that climate oscillations impede opportunities for evolutionary radiations at high latitudes, successive closings and openings of the Bering Strait may have provided a mechanism to promote intraspecific cladogenesis and the evolution of sister taxa (Taylor & Dodson 1994; Väinölä 2003).

It is increasingly evident that the so-called trans-Arctic interchange involved a far more complicated series of events than previously surmised. First, the Bering Strait opened on multiple occasions; at least four openings are now documented. Estimated ages of the principal trans-Arctic interchange vary between 3.1 and 4.1 million years ago (Ma) (Marincovich & Gladenkov 2001) and more precisely around 3.5 Ma at a time when several northern marine species reached the opposite ocean (Vermeij 1991; Briggs 1995). However, earlier Pacific–Arctic connections also occurred. Diatom fossil records from North Chukotka suggested two prior openings, one during the Early Middle Miocene ~16–17 Ma, and another at the end of the Middle Miocene ~11–12 Ma (Sher 1999). Other fossil-calibrated dates indicate an opening between the Late Miocene and Early Pliocene epochs ~4.8–7.4 Ma (Marincovich & Gladenkov 1999), most likely near the end of the Miocene ~5.3 Ma (Gladenkov *et al*. 2002; Gladenkov & Gladenkov 2004). Phylogenetic analyses of several taxa have revealed Pacific–Atlantic divergence estimates varying between 6 and 12 Ma (Collins *et al*. 1996; Harrison & Crespi 1999). In some cases, multiple Pacific–Atlantic exchanges within a genus have been identified, e.g. in the fish genus *Osmerus* (Taylor & Dodson 1994) and in the clam *Macoma balthica* complex (Väinölä 2003; Nikula *et al*. 2007). Evidence also exists of recent or ongoing interoceanic gene flow (e.g. *Strongylocentrotus*, Palumbi & Kessing 1991). Over millions of years, marine species have had manifold opportunities to disperse through the Arctic. Dispersal and colonization patterns are thus complex, covering long periods of time and distinct Bering Strait openings.

Given the complexity of trans-Arctic dispersals, the biogeographical history of northern marine fishes with high vagility capable of rapidly responding to temperature variations may provide new insights into the evolutionary consequences of the opening and closing of the Bering Strait and the importance of Pliocene/Pleistocene climate oscillations for macroevolution. The capelin, a small, cold-water marine pelagic fish with a circumpolar distribution (Fig. 1; see Vilhjálmsson 1994 for a description of Northern Hemisphere capelin distribution) is such a species. It is a member of the Osmeridae (northern smelts; Nelson 1994). Capelin are short-lived and generally do not survive the first spawning event that occurs at about 4 years of age (Vilhjálmsson 1994; Burton & Flynn 1998). Capelin

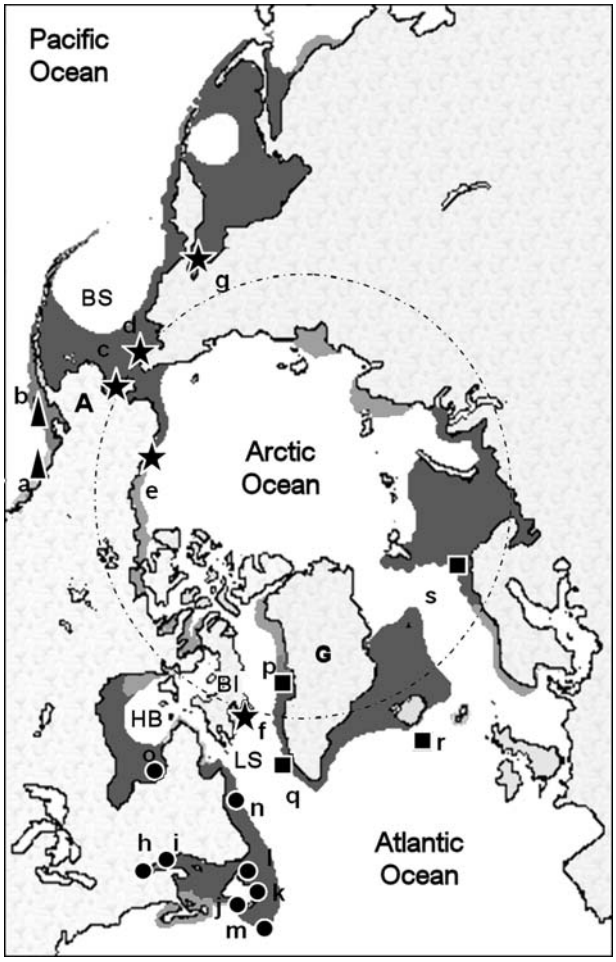

Fig. 1 Capelin world distribution (modified from Vilhjálmsson 1994). Dark grey designates the common occurrence of the species and the light grey areas designate rare or sporadic occurrence. The broken line identifies the Arctic Circle. A, Alaska; BS, Bering Sea; BI, Baffin Island; HB, Hudson Bay; G, Greenland; LS, Labrador Sea. Symbols identify the four capelin mtDNA clades; triangles designate the location of the northeast Pacific (PAC) clade; stars, the Arctic (ARC) clade; circles, the northwest Atlantic (NWA) clade; and squares, the northeast-central Atlantic (NECA) clade. Sampling sites are identified with small letters (also see Table 1): a, Glacier Bay; b, Kodiak Island; c, Bering Sea [US Exclusive Economic Zone (EEZ)]; d, Bering Sea (Russian EEZ); e, Beaufort Sea; f, Pagnirtung, Baffin Island; g, Okhostk Sea; h, St Flavie, Québec; i, St Irénée, Québec; j, Placentia Bay, Newfoundland; k, Conception Bay, Newfoundland; l, Notre Dame Bay, Newfoundland; m, Southeast Shoal; n, Nain, Labrador; o, Hudson Bay; p, Disko Island, Greenland; q, South Greenland; r, Iceland; s, Barents Sea.

populations are highly sensitive to water temperature changes. Evidence from the Canadian Atlantic maritime stocks (Frank *et al*. 1996; Carscadden *et al*. 2001), the Icelandic stock (Vilhjálmsson 1994), and the Barents Sea stock (Gjøsæter 1998) all indicate that this species' distribution was modified within few generations in response to changes in temperature regimes.

The time since the divergence of capelin from the genus *Osmerus* was estimated at ~5.4 Ma (Taylor & Dodson 1994). McAllister (1963) hypothesized that capelin evolved in the northeast Pacific since it is the only location where all genera of the osmerid family occur today, compared to the North Atlantic with only two smelt species (*Osmerus mordax* in North America and *Osmerus eperlanus* in Europe) and the capelin.

Contemporary populations of capelin in the North Atlantic comprise at least two genetically distinct allopatric clades: the northwest Atlantic (NWA) and northeast-central Atlantic (NECA) clades, isolated for approximately 2 million years (Myr), based on a restriction fragment length polymorphism (RFLP) analysis of the entire mitochondrial DNA (mtDNA) molecule (Dodson *et al*. 1991) and sequence analysis of the mitochondrial cytochrome *b* gene (Birt *et al*. 1995). Capelin of these two clades exhibit some morphological differences, but no single character is diagnostic of either clade (Tremblay and Dodson, unpublished data). The distribution of the two clades suggests that the Labrador Sea may be a zone of sympatry, but it is unclear whether the two clades are actually in contact and whether they interbreed. Genetic relationships between Atlantic and Pacific capelin wait to be established. Capelin thus represents an appropriate species model for studying phylogeographical patterns associated with multiple openings of the Bering Strait and Pleistocene climatic oscillations because of its circumpolar distribution, its probable Pacific origin, and its distributional plasticity related to temperature variations.

The overall objective of this study was to evaluate to what extent trans-Arctic dispersals have contributed to either promoting or impeding evolutionary diversification within capelin. The first objective was to document the degree of intraspecific cladogenesis by reconstructing the phylogeny of capelin from samples collected over the species' range based on mtDNA polymorphisms. Second, we resolved the demographic history of the major clades within the species to illuminate the effect of Pleistocene climatic oscillations on both demographic and range expansions. Third, we aimed to quantify the mixing of clades within the Labrador Sea, previously hypothesized to be a zone of sympatry of the two North Atlantic clades (Dodson *et al*. 1991). Finally, we used the distribution of contemporary clades to infer biogeographical history and the number, timings and directions of trans-Arctic dispersals.

Materials and methods

Sampling

All sampling sites are listed in Table 1 and their locations identified on Fig. 1. A total of 270 capelin from 19 sampling sites were sequenced, including 11 samples previously obtained by Dodson *et al*. (1991). New samples include the northeast Pacific (two sites: Kodiak Island. and Glacier Bay, Alaska), the Okhotsk Sea (one site), the Bering Sea [two sites: the US Exclusive Economic Zone (EEZ) (61°59′N, 169°69′W) and the Russian EEZ (64°01′N, 179°58′W)], and the Beaufort Sea (one site). Additional samples from Canadian waters (two sites) were obtained from Coats Island, northern Hudson Bay, and Pangnirtung, Cumberland Strait, Baffin Island. In the presumed area of sympatry, an additional 238 capelin were obtained from the shrimp fishery along the west coast of Greenland from 59°56′N to 70°19′N, along the east coast of Labrador from 46°45′N to 58°34′N and in Davis Strait (67°49′N 58°17′W). These latter samples were used for RFLP–based clade identification (see below).

PCR amplification and sequencing

Total DNA was extracted from tail muscle using a protocol adapted from the standard proteinase-K/phenol-chloroform procedure (Hillis *et al*. 1996). A total 609 base-pair (bp) fragment of the mitochondrial 5′-cytochrome *b* gene (including 9 bp before the start codon) was amplified. We amplified two contiguous fragments (overlapping by 11 bp), one of 411 bp and another of 209 bp. Primer sequences for the 411 bp fragment were Cyb2 (5′-CCCTCAGAATGA TATTTCTCCTA-3′; derivative of universal primers of Kocher *et al*. 1989) and P1 (5′-AAAACCACCGTTGTTATTCAAC TACA-3′). For the 209 bp fragments, primers were P2 (5′-GGGTTGTTTGATCCTGTTTCGTG-3′) and P3 (5′-TATGA TGATGCCTTTGTAGGCT-3′). PCR was performed under the following profile: 5 min at 95 °C followed by 40 cycles of 1 min at 95 °C, 1 min at 45 °C and 1.5 min at 72 °C, completed with 5 min of dwell period at 72 °C. PCR products were purified using the QIAquick Gel Extraction Kit following the manufacturer' protocol (Quiagen). Fragments were sequenced in both directions on ABI PRISM 377 DNA Sequencer and analysed using ABI PRISM Sequencing Analysis 3.0 program.

Data analysis

Sequences were aligned be eye and no ambiguities were encountered. The data set was reduced using DAMBE (Xia & Xie 2001) so as to retain only unique haplotypes for phylogenetic reconstruction. Neighbour-joining (NJ) and maximum parsimony (MP) algorithms were performed on PAUP* 4.0b10 (Swofford 2002). The NJ tree was linearized

Table 1 Description of the haplotypes. *m* refers to the location identification letter in Fig. 1, Location indicates the names of sampling location (abbreviated clade ID), *n* is the number of sequenced individuals, *s* indicates the number of unique haplotypes, and ω is the mean number of pairwise differences. The right part of the table presents all 154 unique haplotypes, their clade-specific numerical identification and occurrence for individual sites and clades

					Clade specific haplotype identification and occurence																																														
m	Location (clade)	*n*	*s*	ω	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
NE-Pacific (PAC)																																																			
a.	Glacier Bay, Al.	16	12	3.33 ± 1.81	1	1	1	1	1	1	1	1	1	3	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.																				
b.	Kodiak Island, Al.	16	16	4.48 ± 2.33	.	.	.	.	.	.	.	.	.	1	.	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1																				
Total		32	27	4.38 ± 0.98	1	1	1	1	1	1	1	1	1	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1																				
Arctic (ARC)																																																			
c.	Us EEZ	20	11	1.51 ± 0.95	1	1	1	1	1	1	1	1	1	5	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.													
d.	Russian EEZ	11	9	2.07 ± 1.25	1	.	.	.	.	.	.	.	.	2	.	1	1	1	1	1	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.													
e.	Beaufort Sea	18	7	0.88 ± 0.64	.	1	.	.	.	.	.	.	.	11	.	.	.	.	.	.	.	.	1	1	1	2	1	.	.	.	.	.	.	.	.	.	.	.													
f.	Baffin Island	10	7	1.53 ± 1.00	.	.	.	.	.	.	.	.	.	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	.	.	.	.	.	.													
g.	Okhotsk Sea	13	9	1.60 ± 0.963	.	2	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1													
Total		72	34	2.69 ± 0.60	2	4	1	1	1	1	1	1	1	26	7	1	1	1	1	1	1	3	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1													
NW Atlantic (NWA)																																																			
h.	St. Flavie, Qué	12	8	2.12 ± 1.27	4	2	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
i.	St. Irénée, Qué	12	9	2.36 ± 1.38	4	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	1	1	1	.	.	
j.	Placentia Bay, Nfld	15	12	1.98 ± 1.18	3	.	.	.	.	.	.	1	1	1	1	1	1	1	1	1	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
k.	Conception Bay, Nfld	11	8	1.71 ± 1.08	4	1	.	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
l.	Notre-Dame Bay, Nfld	13	8	1.87 ± 1.14	4	1	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
m.	Southest Shoal, Nfld	13	10	2.38 ± 1.39	4	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	
n.	Nain, Labrador	13	9	2.15 ± 1.28	5	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	.	.	.	.	.	.	
o.	Hudson Bay	11	6	1.38 ± 0.92	6	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1	1	
Total		100	46	3.25 ± 0.69	34	7	1	1	1	1	1	8	1	2	2	1	2	1	1	1	2	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	
NE-C Atlantic (NECA)																																																			
p.	Disko Island	17	13	3.09 ± 1.69	1	1	1	1	1	1	1	1	1	1	1	4	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
q.	South Greenland	14	11	3.31 ± 1.81	.	.	.	.	.	.	.	.	.	.	.	1	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	1	1	1	1	1	1	1
r.	Iceland	16	14	2.94 ± 1.63	.	.	.	.	.	.	.	.	.	.	.	1	3	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
s.	Barents Sea	19	18	3.37 ± 1.81	.	.	.	.	.	.	.	.	.	.	.	2	1	.	.	1	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.
Total		66	47	4.03 ± 0.75	1	1	1	1	1	1	1	1	1	1	1	8	10	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1
All *M. villosus*		270	154	17.44 ± 7.77																																															

using MEGA 3.1 (Kumar *et al*. 2004) to represent the approximate divergence times of clades (Takezaki *et al*. 1995) using a 2% nucleotide divergence/Myr (see below). Trees were rooted with *Osmerus mordax* (Northwest Atlantic, clade A; Pigeon *et al*. 1998) and *Osmerus dentex* (GenBank AB114911), both belonging to the genus *Osmerus* closely related to *Mallotus villosus*, and *Plecoglossus altivelis* (GenBank AB047553), a sister genus of Osmeridae and Salangidae (icefishes) (Johnson & Patterson 1996; J. J. Dodson, J. Laroche & F. Lecomte, unpublished). Settings for phylogenetic analyses were obtained using MODELTEST version 3.6 (Posada & Crandall 1998) and corresponded to a TrN + I + G model. Finally, we constructed mtDNA haplotype networks for each clade using TCS 1.2 (Clement *et al*. 2000) and connected them according to the NJ reconstruction done previously.

We estimated mean intra- and interclade genetic distances with Tamura–Nei distance (Tamura & Nei 1993). We also estimated the time elapsed since the last shared ancestor of the major mtDNA clades with a molecular clock calibration of 2% nucleotide divergence/Myr for marine fishes (Bowen *et al*. 2001). Departures from this rate would increase or decrease the time elapsed since the last shared ancestor, but would not alter the relative timing of such events. As there is undoubtedly a wide error associated with such molecular clock calibrations, estimated divergence times must be considered only as rough approximations of the timing of historical events. To verify if the clades were evolving in a clock-like fashion, we performed a likelihood ratio test of rate constancy (Huelsenbeck & Rannala 1997; Govindarajan *et al*. 2005). Likelihood scores were calculated in PAUP* version 4.0b10 (Swofford 2002) using the best-fit model with and without the molecular clock constraint and were used to generate a χ^2-test statistic with $n-2$ degrees of freedom (n = number of haplotypes).

An analysis of molecular variance (AMOVA) was used to detect genetic structure within clades by partitioning the observed genetic variation into variance components among and within sampling sites. Given the large number of singletons, Φ_{ST} was employed to estimate F_{ST}. Statistical significance was estimated by permutation analysis (1000 permutations). Haplotype and nucleotide diversity were also estimated using ARLEQUIN 3.1 (Excoffier *et al*. 2005). To calculate the time since divergence of the major clades, we considered the net sequence divergence instead of the mean sequence divergence between clades to reduce the overestimation caused by ancestral polymorphism predating the split of the clades (Avise & Walker 1998; Klicka & Zink 1999). The net sequence divergence between clades A and B is estimated using $P_{AB(net)} = P_{AB} - (P_A + P_B)/2$, where P_{AB} is the mean sequence divergence between clades A and B, P_A and P_B are the mean sequence divergence between the haplotypes composing each clade. Calculations were performed using MEGA 3.1 (Kumar *et al*. 2004).

When population fission leads to divergent populations of unequal sizes, standard genetic distances between populations may severely overestimate divergence times. Thus, we also estimated divergence times among clades using the method of Gaggiotti & Excoffier (2000). This method removes the possible effect of bottlenecks and unequal population sizes. Calculations were performed using ARLEQUIN 3.1 (Excoffier *et al*. 2005).

Mismatch analysis

As all of the identified mtDNA clades are composed of clusters of diagnostic haplotypes derived from distinct evolutionary lineages genetically isolated for extensive periods of time, mismatch analyses were performed to search for evidence of potential impacts of glaciation events on the molecular architecture of the contemporary clades. However, because of significant Φ_{ST} estimates within two clades (see below), some capelin samples were also analysed separately. Mismatch analysis provides information about demographic processes such as the magnitude and timing of population expansion based on the observed nucleotide pairwise distribution (demographic mismatch analysis: Rogers & Harpending 1992). It also characterizes the effect of range expansion independently of demographic expansion (spatial mismatch analysis: Ray *et al*. 2003; Excoffier 2004). A unimodal distribution is considered here as the signature of a population expansion, although it is not possible to say whether it is a consequence of demographic or spatial processes (see Excoffier 2004). In the specific case of a range expansion, preferential reproduction with neighbouring individuals (local demes) should lead to some level of population structure while the local deme should nevertheless exhibit a unimodal mismatch distribution (Excoffier 2004; Excoffier *et al*. 2005). Range expansion may thus produce the same genetic signature as a demographic expansion. As capelin have most definitely experienced both demographic and range expansions and contractions over successive cycles of glaciation, we conducted both demographic and spatial mismatch analyses to gain insight into both processes. Considering a pure demographic expansion, the timing of expansion events was evaluated using τ ($\tau = 2ut$: unit of mutation time before present) where $u = m_T\mu$ (μ is the mutation rate per generation and m_T is the number of nucleotides used) (Rogers & Harpending 1992). Female effective population sizes before (N_{f0}) and after expansion (N_{fi}), were estimated using the mean pairwise difference in the mismatch distribution before (θ_0) and after the expansion (θ_1), respectively ($N_{fi} = \theta_i/2\,\mu$). We used a mutation rate of 1% per Myr within lineages to convert mismatch estimators to time since expansion and N_f. We used the age at which 95% of the individuals are mature to estimate generation time. Using the percentage age distribution of spawning

capelin in the period 1954–1983 (Gjøsæter 1998), we calculated the mean generation time to be 3.8 years. Alternatively, considering a spatial expansion, we estimated three parameters, τ, $\theta = \theta_0 = \theta_1$ (assuming that $N = N_0$), and $M = 2Nm$ (where m is the rate at which the sampled deme would exchange migrants with a unique population of infinite size) using the same method as that used in the estimation of the parameters of a demographic expansion (Excoffier *et al*. 2005). The analysis of spatial mismatch, based on the assumption of no demographic expansion within the sampled deme, provides an estimate ($N_f = \theta/2\mu$) of the effective female population size necessary to explain the mismatch distribution uniquely in terms of range expansion. For both methods, the parameters calculated reflect only the average processes involved, either spatial or demographic and represent the minimal value (e.g. smallest population size or smallest range contraction) that has imprinted the molecular architecture we see today. There is no doubt that capelin populations have expanded and contracted on several occasions, but not all such events are imprinted on the genetic architecture of the species. The interpretation of the mismatch must thus be taken with caution as the expansion parameter cannot be interpreted as a single expansion spanning a timeframe that includes several glacial cycles. Calculations for both methods (spatial and demographic) were performed in ARLEQUIN 3.1 (Excoffier *et al*. 2005).

Clade identification in the putative zone of sympatry

To detect if the distributions of the two North Atlantic clades overlap in the Labrador Sea, we developed an RFLP identification method to assign individuals to one of the two clades. A sample of 228 capelin were so analysed to which we added 10 609-bp mtDNA sequences of individuals from west Greenland (70°19′N, 55°03′W), for a total of 238 capelin sampled throughout the Labrador Sea and northern Hudson Bay. Following the 411 bp mtDNA cytochrome *b* gene amplification, amplified fragments were digested with restriction enzyme *Stu*I (AGG′CCT) at 37 °C overnight in a 20 μL total volume containing 5–15 μL of the DNA fragments (depending on the concentration), 0.5 μL of restriction enzyme and 2 μL of digestion buffer. Digestion fragments were separated by electrophoresis and viewed under UV light. Digestion patterns were clade-specific and easily recognizable as *Stu*I produced two distinct length fragments for the NWA clade (153 + 258 bp), and for the NECA clade (99 + 312 bp).

Results

Mitochondrial phylogeny

A total of 154 haplotypes were identified (GenBank Accession nos DQ457421–DQ457574) from the 270 sequences obtained and 129 haplotypes were singletons (Table 1). Phylogenetic reconstructions using both neighbour-joining (NJ) and maximum-parsimony algorithms supported the same principal relationships with strong bootstrap support (Fig. 2). Four principal clades were identified and all were characterized by one or two prevalent haplotypes, occupying the central position in clade-specific haplotype networks from which other haplotypes were separated by one to three mutation steps (Fig. 3). The four mtDNA clades included the two previously described Atlantic clades (NWA and NECA; Dodson *et al*. 1991). A third clade included sequences from the Okhotsk Sea, the Bering Sea (US and Russian Exclusive Economic Zones), the Beaufort

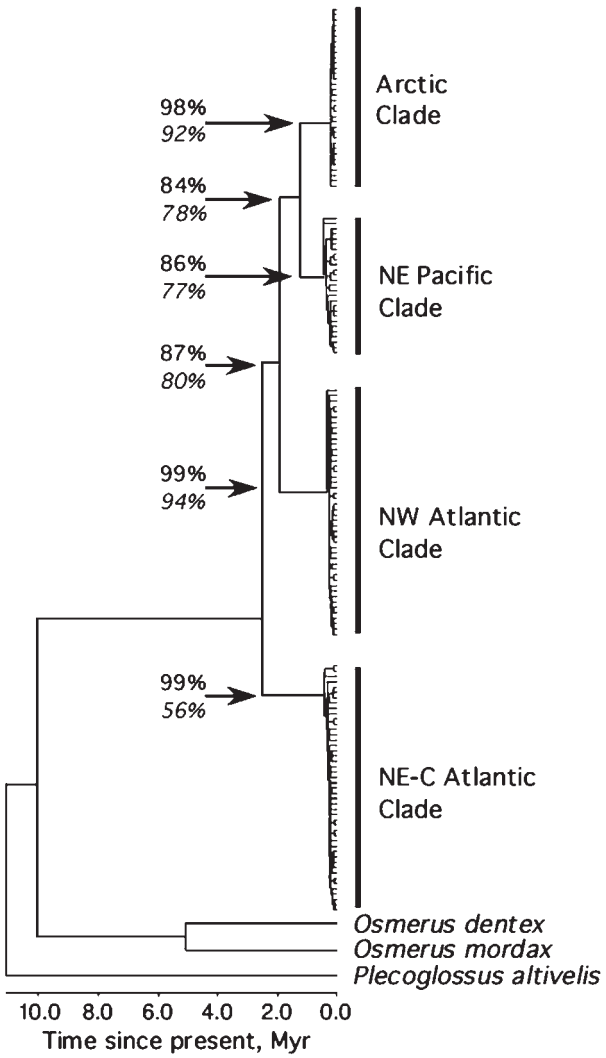

Fig. 2 *Mallotus villosus* neighbour-joining tree. Numbers on nodes are percentage bootstrap values for neighbour-joining and maximum parsimony, respectively (1000 bootstrap pseudo-replicates for each method). The tree was linearized to represent the time since present, calculated using a divergence rate of 2%/Myr (mutation rate of 1%/Myr).

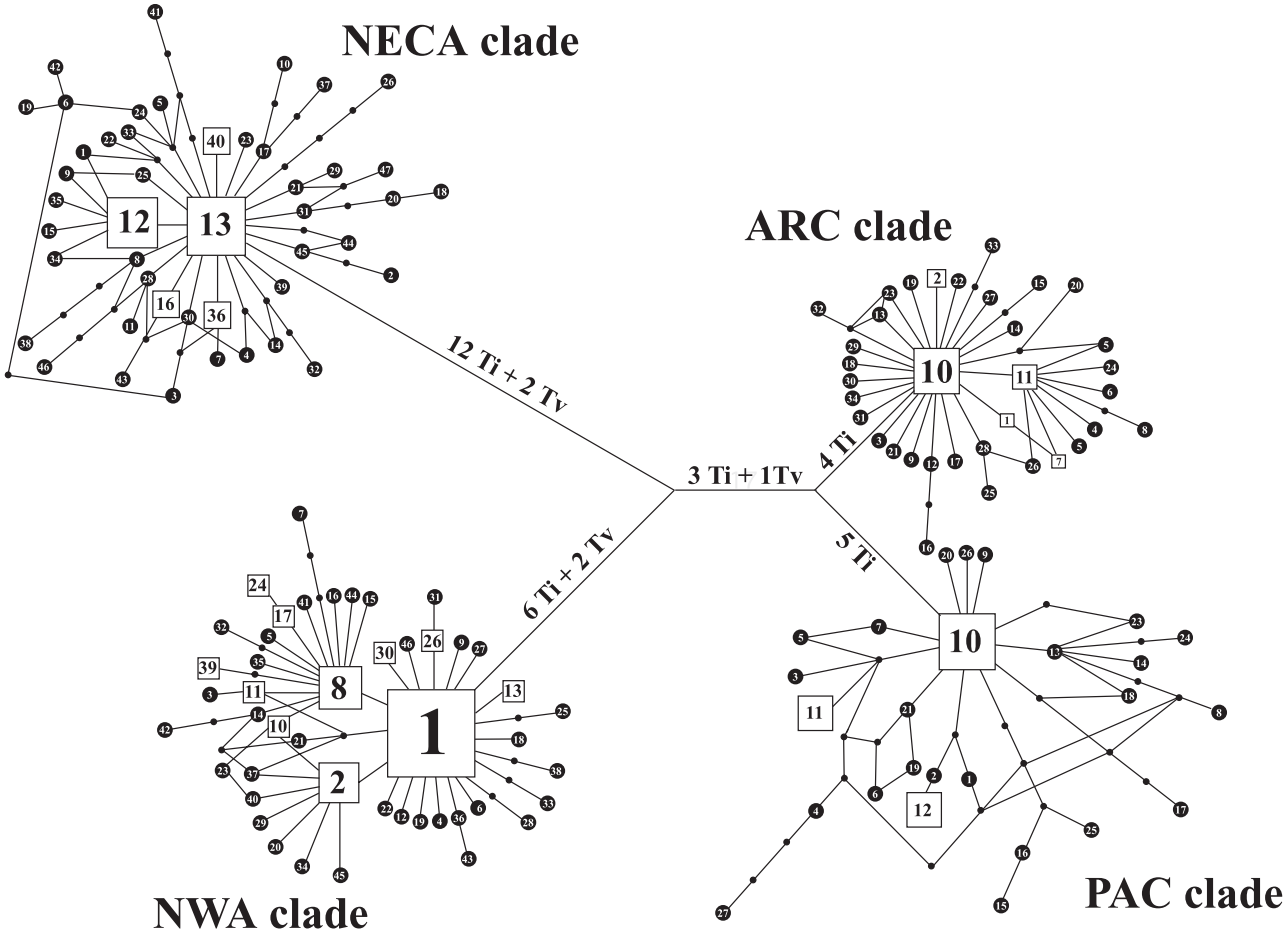

Fig. 3 Haplotype networks of the four mtDNA clades. Intraclade networks were obtained using TCS. Numbers of mutations separating the central haplotypes of the four clades are indicated (Ti, transitions; Tv, transversions). Central haplotypes were connected according to the neighbour-joining tree presented in Fig. 2. Haplotype codes refer to Table 1, and sizes of the symbols correspond to observed total numbers of haplotypes.

Table 2 Capelin mtDNA clades (ARC, Arctic; PAC, northeast Pacific; NWA, northeaster Atlantic; NECA, northeast-central Atlantic) intra-clade pairwise differences (%) and standard deviation (SD) on diagonal; between-clades net sequence differences (%) and standard error (SE) below diagonal; and the time of clade divergence (Ma) and SE (molecular clock: 2% divergence/Myr) above diagonal. All calculations are based on Tamura–Nei distances

	PAC	ARC	NWA	NECA
PAC	**0.66 (0.34)**	0.96 (0.26)	1.58 (0.36)	1.90 (0.40)
ARC	1.91 (0.52)	**0.26 (0.16)**	1.56 (0.37)	2.27 (0.45)
NWA	3.15 (0.72)	3.11 (0.73)	**0.33 (0.19)**	2.06 (0.42)
NECA	3.80 (0.80)	4.53 (0.90)	4.12 (0.83)	**0.52 (0.27)**

Sea and Baffin Island. We named this the Arctic clade (ARC) considering its wide trans-Arctic distribution. The fourth clade grouped Glacier Bay (GB) and Kodiak Island (KI) samples and was named the northeast Pacific clade (PAC; Fig. 1). The likelihood ratio test did not refute the molecular clock assumption ($\chi^2 = 95.49$, d.f. = 99, rejection value at $\alpha = 0.05$: 123.23).

The four monophyletic clades were highly divergent, differing by 1.9–4.5% net sequence divergence, while intraclade divergences varied between 0.3% and 0.7% (Table 2). Based on net pairwise distance estimates (Table 2), the northeast-central Atlantic clade first diverged approximately 2 Ma and the northwest Atlantic clade diverged from the Pacific-Arctic clade approximately 1.6 Ma (Table 2). The Arctic and northeast Pacific clades represent the most recent divergence, occurring approximately 1 Ma. The net interclade divergences represent reductions of 0.22 Myr (SD = 0.05) relative to the mean divergence times illustrated in Fig. 2. Applying the Gaggiotti & Excoffier (2000) correction reduced the net interclade sequence divergence estimates by an additional 0.29 Myr (SD = 0.07). As such, the standard genetic distances between clades somewhat overestimated divergence times, but the corrected estimates were well

within the SE of the net pairwise interclade divergences (Table 2).

The 32 sequenced individuals of the northeast Pacific clade (PAC) revealed 27 haplotypes characterized by two diagnostic substitutions. One haplotype (PAC10), shared between the two sites, was the most frequent at GB (25%), but appeared only once in KI, which revealed 15 other haplotypes all of which were singletons (Table 1, Fig. 3). AMOVA revealed that 4.3% of the genetic variation occurred between sampling sites ($\Phi_{ST} = 0.043$ $P = 0.037$). The significant Φ_{ST} value indicates that this clade is genetically structured. Mismatch analyses (see below) were thus conducted separately for the two sampling areas which were considered as separate populations.

The 72 sequenced individuals of the Arctic clade (ARC) revealed 34 haplotypes. One (ARC10) was common to all samples (Table 1, Fig. 3). A second haplotype (ARC18) was restricted to the Russian EEZ, accounting for 22% of the sample. All other haplotypes from this site were unique. AMOVA showed that 10.8% of the genetic variation found in the ARC clade was among the sampling sites ($\Phi_{ST} = 0.108$, $P < 0.001$), suggesting that these sampling sites are genetically structured. Pairwise Φ_{ST} analyses (results not shown) revealed three groups; a first group including the Okhotsk Sea (OS) and the Beaufort Sea (BS) samples (pairwise $\Phi_{ST} = 0.005$; $P = 0.298$), a second group including the US EEZ and Baffin Island samples (pairwise $\Phi_{ST} = 0.061$; $P = 0.070$) and finally the Russian EEZ sample. Mismatch analyses (see below) were thus conducted separately for these three groups. Although these groupings do not conform to a simple geographical model, the mismatch parameters model the history of coalescence of genes within distinct gene pools, thus dictating against pooling sampling sites based on geographical proximity.

The northwest Atlantic clade (NWA) comprised eight sampling sites, all from Canadian waters (Table 1; Fig. 1). From 100 capelin sequenced, 46 haplotypes were found (Table 1, Fig. 3). One (NWA1) was the most frequent at all sites. The AMOVA revealed no significant genetic structure within this clade ($\Phi_{ST} = -0.005$; $P = 0.626$). Finally, the northeast and central Atlantic clade (NECA) encompassed four sites (Table 1, Fig. 1). A total of 66 sequences were obtained involving 47 haplotypes of which two (NECA12 and NECA13) were present in all four sites (Table 1, Fig. 3). As for the NWA clade, AMOVA revealed no significant genetic structure within the NECA clade ($\Phi_{ST} = -0.007$; $P = 0.761$).

Mismatch analyses

All mtDNA clades and their component populations analysed conformed to the sudden population expansion hypothesis ($P > 0.05$). We considered the unimodal mismatch distributions both in terms of demographic and spatial processes; estimated times of expansion were similar under the assumptions of both demographic and range expansion (Table 3). There is no doubt that capelin populations have expanded and contracted on several occasions in concert with glaciations, but only the most recent event to have reduced effective population size to a minimal value may be detected by mismatch analysis. The results of both mismatch analyses suggested that all clades experienced relatively recent expansions. The Pacific clade and its component populations were inferred to expand the earliest, approximately 300 000 years ago, whereas the Arctic clade exhibited the most recent expansion, ranging in age from 180 000 to 110 000 years ago for the three component populations (Table 3). The northwest Atlantic clade was inferred to have expanded approximately 100 000 years after the expansion of the northeast-central Atlantic clade that occurred approximately 270 000 years ago, at about the same time as that of the Pacific clade. Under the assumption of a purely demographic expansion, clades and component populations were inferred to have expanded from very small effective female population sizes to huge effective female population sizes. Under the assumption of spatial expansion, all three populations composing the Arctic clade required relatively small female effective population sizes to explain the inferred spatial expansion and exhibited an exceedingly large rate of exchange with neighbouring populations (Table 3). Neither the NWA nor the NECA clades were genetically structured and, under the spatial expansion assumption, both clades expanded from relatively small effective female population sizes and exhibited large rates of exchange with neighbouring populations (Table 3). The analysis of spatial mismatch also revealed that the northeast Pacific clade required the smallest migration rates relative to all other clades to explain the observed diversity based uniquely on range expansion. In general, the older the expansion event, the smaller the rate of exchange needed to explain the mismatch distribution.

Diversity indices

Haplotype diversity was high for the four clades and relatively higher for both PAC and NECA clades compared to ARC and NWA clades. Nucleotide diversity was reduced and relatively low for ARC and NWA clades (Table 3).

The Labrador Sea

The results of the RFLP analysis (Fig. 4) confirmed the distribution of each clade around Davis Strait. We detected no important areas of sympatry and confirmed previous reports (Dodson *et al*. 1991). However, a single individual from the western side of Davis Strait (58°34′-N 60°06′-W), approximately 600 km to the north of a sample that belonged to the NWA clade, was attributed to the NECA

Table 3 Haplotype (h) and nucleotidic (π) diversity (SD, standard deviation) and mismatch analysis results and mismatch distribution parameter conversions. Number of pairwise comparisons (n)

				Demographic mismatch parameter			Spatial mismatch parameter		
Clade Population	h (SD)	π (SD)	n	τ (age in years)	θ_0 ($N_{ft=0}$)	θ_1 ($N_{ft=1}$)	τ (age in years)	θ N_f	M m
PAC	0.990	0.007	496	4.055	0.027	> 99 999	4.040	0.17	> 99 999
	(0.011)	(0.004)		332 900	579	$> 2 \times 10^9$	331 700	3600	13.72
Glacier Bay	0.958	0.005	120	3.492	0.025	> 99 999	3.521	0.001	276.3
	(0.036)	(0.003)		286 700	579	$> 2 \times 10^9$	286 700	21	6.44
Kodiak Island	1.000	0.007	120	3.234	1.596	> 99 999	3.234	1.651	> 99 999
	(0.022)	(0.004)		265 500	579	$> 2 \times 10^9$	265 500	34 341	1.46
ARC	0.962	0.003	2556	1.715	0	> 99 999	1.703	0.001	> 99 999
	(0.012)	(0.002)		140 800	(0)	$> 2 \times 10^9$	139 800	21	2332.45
US EEZ + Baffin Island	0.929	0.003	435	1.693	0	> 99 999	1.689	0.004	> 99 999
	(0.030)	(0.002)		139 000	(0)	$> 2 \times 10^9$	138 700	86	583.11
Okhotsk Sea + Beaufort Sea	0.865	0.002	465	1.328	0.002	> 99 999	1.330	0.001	> 99 999
	(0.054)	(0.001)		109 000	(43)	$> 2 \times 10^9$	109 200	21	2332.45
Russian EEZ	0.964	0.003	55	2.180	0	> 99 999	2.179	0.001	> 99 999
	(0.051)	(0.002)		180 000	(0)	$> 2 \times 10^9$	178 900	21	2332.45
NWA	0.987	0.003	4950	2.105	0	> 99 999	2.099	0.006	> 99 999
	(0.004)	(0.002)		172 800	(0)	$> 2 \times 10^9$	172 300	129	388.74
NECA	0.992	0.005	1627	3.266	0	> 99 999	3.206	0.062	> 99 999
	(0.005)	(0.003)		268 100	(0)	$> 2 \times 10^9$	263 200	1329	37.46

ARC, Arctic; PAC, northeast Pacific; NWA, northeastern Atlantic; NECA, northeast-central Atlantic.

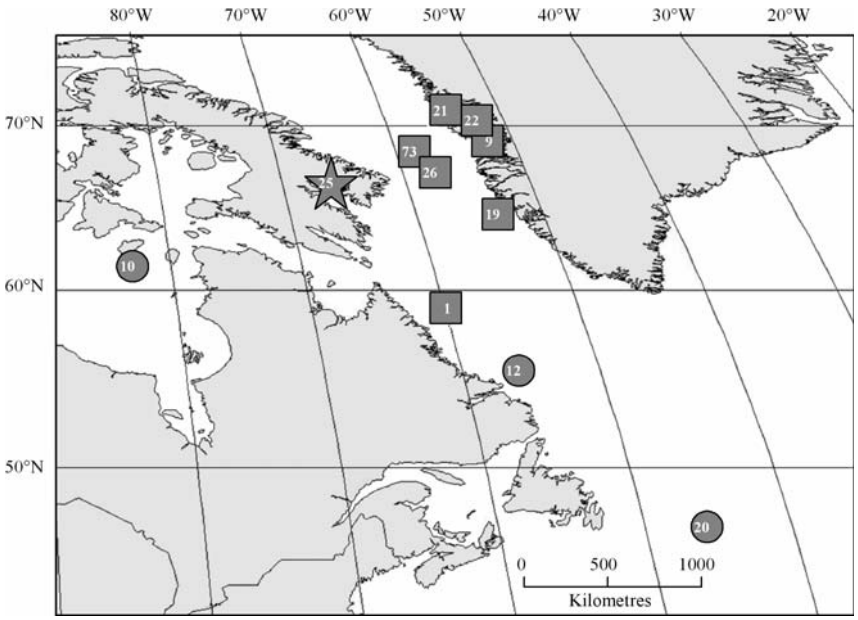

Fig. 4 Labrador Sea secondary contact zone. Circles represent NWA clade, squares NECA clade, and stars the ARC clade sample from Pangnirtung, Baffin Island. Numbers within symbols are sample sizes.

clade. In addition, the discovery of capelin assigned to the ARC clade at Pangnirtung, southeast Baffin Island, demonstrated that the Labrador Sea is in fact a potential geographical contact zone for three of the four capelin clades.

Discussion

Capelin is composed of four highly divergent and geographically distinct mtDNA clades. These deep divisions evolved through several glacial cycles and climate oscillations during which time the range of the species and populations sizes would have dramatically changed. Capelin populations comprising all clades experienced at least one demographic reduction–expansion episode during their recent history, sufficiently strong enough to create a unimodal mismatch distribution. Similarly, all clades provided evidence consistent with extensive range expansion, with the exception of the Glacier Bay population. Although there is no doubt that capelin populations expanded and contracted in concert with Pleistocene glaciations, not all of these episodes have left an imprint on the molecular architecture of the populations. Furthermore, there is no evidence that past range expansions have led to the mixing of the four historical clades; all four clades remain geographically isolated. Each of the four clades is characterized by several prevalent haplotypes, distributed over the entire range of the clade. These common haplotypes occupy the central position in clade-specific haplotype networks from which other haplotypes are separated by one to three mutation steps (Fig. 3). This is associated with high haplotype diversity and low nucleotide diversity (Table 3) and attributable to demographic expansion after a period of low effective population size (Avise *et al*. 1984; Rogers & Harpending 1992; Grant & Bowen 1998).

Two capelin clades are widespread in the North Atlantic, one associated uniquely with Canadian Atlantic waters, including Hudson Bay, and the second distributed from West Greenland to the Barents Sea. Two additional clades occur in the Arctic and northeast Pacific Oceans, representing the most recent divergence within the capelin phylogenetic tree. The Arctic clade is widespread and considering that two identical haplotypes were shared between the Labrador Sea (Baffin Island) and the Bering Sea (US and Russian EEZ), and one with the Beaufort Sea, it is likely that capelin from the Canadian Arctic also share these haplotypes. Although the Labrador Sea contains three capelin clades, little evidence of sympatry was detected. The solitary capelin found in the western Labrador Sea with a NECA haplotype (Fig. 4) suggests that there may be some geographical overlap of the two Atlantic clades in the region. However, detecting relatively restricted areas of sympatry or introgression will require a far greater sampling effort and the application of nuclear markers.

As expected, Pleistocene glaciations have left important imprints on the molecular architecture of the capelin populations. The inferred expansion of the Arctic and the northwest Atlantic capelin populations coincides approximately with the warm Sangamon (Eemian) interglacial that occurred 135 000–115 000 years ago at the beginning of the last glacial cycle (Hewitt 1996). The northeast-central Atlantic and northeast Pacific capelin populations also experienced significant reduction–expansion events, but these started earlier, approximately 270 000 years ago coincident with the earlier and even warmer Yarmouth (Hoxnian) interglacial that occurred 200 000–300 000 years ago (Pagé 1999).

The intervening glacial advance thus appears to have had a far more important impact in the northwest Atlantic and Arctic Oceans than in the northeast Pacific or northeast-central Atlantic. Range and demographic contractions caused by the most recent glaciations do not appear to have influenced the molecular architecture of any of the capelin populations.

The present geographical isolation of the four clades, apparently maintained over many glacial cycles, is particularly surprising given the capacity of the species to rapidly modify its distribution in response to temperature variation. Episodes of global warming and deglaciation would surely have favoured population expansion and dispersal, increasing the probability of mixing among populations derived from different clades. There is indirect evidence that this is occurring during the present interglacial. In addition to the striking occurrence of three clades in the Labrador Sea, observations of the diets of thick-billed murres (*Uria lomvia*) nesting on Prince Leopold Island in Lancaster Sound indicate ongoing dispersals in the eastern Arctic. From 1975 until 2000, Arctic cod (*Boreogadus saida*) was the dominant fish fed to murre chicks. During 2001 and 2002, capelin appeared in their diet for the first time in recent history (A. Gaston, personal communication). Continued dispersal in contemporary times may form the basis of another Atlantic invasion leading to the sympatric occurrence of the clades in secondary contact zones. It thus seems possible that during past expansion events, the extended ranges of capelin may have resulted in hybridization on several occasions, but evidence of this may have been lost by glacial advances extinguishing such peripheral populations (Hewitt 2004).

The contemporary genetic and geographical isolation of the four capelin clades is most likely a consequence of a recent range expansion. During the most recent glacial maximum, only those populations located in glacial refuges would be expected to survive and expand during the present interglacial period. The Arctic clade was probably located to the north of the Bering land bridge where it would have been isolated from the Pacific clade possibly confined to the south of the Cordilleran ice sheet, as proposed for Pacific herring (Grant 1986). The two Atlantic clades may also have been segregated along the North American and European coastlines to the south of their present-day distributions. During the last glacial maximum, these refuge populations would have been physically isolated but nevertheless huge, with enormous effective female population sizes. During the subsequent interglacial period, these source populations would have contributed to re-establishing the highly discrete geographical distributions of the major clades observed today. The north–south displacement of capelin clades along coastlines during the most recent glacial cycle may thus have served to reduce population bottlenecks and retain genetic diversity. This stands in marked contrast to South African populations of Southern Hemisphere 'Old World' anchovies (genus *Engraulis*) which have experienced recurring extinctions during ocean-climate changes and episodic recolonizations from Northern Hemisphere populations, largely because of the absence of north–south coastlines that allow range shifts during climatic extremes (Grant & Bowen 2006).

We may only speculate about the degree of reproductive isolation between clades. The average interval required for reproductive isolation and speciation in fishes is 1.7 Myr (Avise 2000), suggesting that some of the capelin clades may have achieved species status. The clade identification presented here is based on the maternally inherited mtDNA and does not permit the identification of hybrids. However, if the analysis of nuclear markers were to demonstrate extensive introgression between clades, this would necessarily be mediated uniquely by males. Given the huge population sizes and vagility of capelin, such asymmetric gene flow seems highly unlikely, but not impossible. Finally, a number of phenotypic differences between clades is also suggestive of speciation. Capelin populations in the north Pacific and northwest Atlantic differ from northeast-central Atlantic populations in the number of vertebrae, the range of latitude for spawning, spawning mode (beach vs. deep water) and a number of other life-history traits (Stergiou 1989). Differences observed in some morphometric and meristic characters between the populations of capelin in the North Pacific and Atlantic Oceans have been used to argue for the recognition of two subspecies in the two oceans (Klyukanov & McAllister 1979).

Biogeographic inference

We propose that capelin originated in the Pacific Ocean and dispersed on two separate occasions to the North Atlantic. A first dispersal founded the northeast-central Atlantic clade and a later dispersal founded the northwest Atlantic clade. In both the clam (*Macoma balthica* complex) and mussel (*Mytilus edulis* complex), genetic and systematic diversity is best explained as a result of repeated trans-Arctic dispersals to the North Atlantic, and subsequent contact and mixing of successive offshoots of the original Pacific lineages in the North Atlantic (Väinölä 2003; Nikula *et al*. 2007). Many of these events occurred in the Late Pliocene and Early Pleistocene, as inferred here for the two Atlantic capelin clades. In the case of the fish genus *Osmerus*, the north Pacific/Arctic species (*O. dentex*) and the northwestern Atlantic species (*O. mordax*) shared a common ancestor more recently than either has with the northeastern Atlantic species (*O. eperlanus*), indicating that the species complex arose from two independent Pacific–Atlantic divergences. Divergence between the Atlantic forms of *Osmerus*, however, is about twice that of the two Atlantic capelin clades (Taylor & Dodson 1994). The net nucleotide

distance separating the two Atlantic capelin clades indicates isolation of the two clades in the Late Pliocene–Early Pleistocene approximately 2 Ma, at about the same time that large-scale glaciations were initiated in the Northern Hemisphere (Kleiven *et al*. 2002; Harris 2005). Following access to the Arctic, *Mallotus* would have been isolated from their Pacific congeners when growth of continental ice sheets caused a fall in sea level of approximately 100–200 m in the North Pacific. As a result, the Bering Strait was closed, isolating the North Pacific and Arctic Basins (Nechaev *et al*. 1994). Capelin, most probably already present in the North Atlantic, would have been genetically isolated from Pacific congeners.

The second dispersal that founded the northwest Atlantic clade is inferred to have taken place about 1.6 Ma. As in the case of *Osmerus*, northwestern Atlantic capelin shared a common ancestor more recently with the Pacific–Arctic clade than either has with the northeastern Atlantic clade. The contemporary geographical isolation of the two North Atlantic clades may be associated with the presence of the Labrador current that developed ~3.0 Ma, before the inferred time of arrival of the northwest Atlantic capelin clade (Berggren & Hollister 1974; Cronin 1988). Wares (2001) proposed that this event contributes to the west–east genetic cleavage of *Asterias* sp. The presence of an important input of cold water in the Labrador Sea may also contribute to isolating the northwest and northeast-central Atlantic capelin clades.

Finally, in contrast to *Osmerus*, the Pacific-Arctic complex of capelin has most recently diverged to form two clades occupying the northeastern Pacific and the Arctic Ocean. The divergence of PAC and ARC clades is estimated to have taken place about 1 Ma. Environmental forcing of glacial–interglacial episodes may have caused their divergence, but it is actually not possible to identify an event that produced these distinct clades. Contemporary ARC and PAC clades appear to be genetically and geographically isolated by the Aleutian archipelago. This region may impair dispersal, although no oceanographic or geomorphological barriers are known. Vermeij (1989) observed a vicariant North Pacific pattern where many ancestral amphi-Pacific molluscan taxa gave rise to separate eastern and western descendants.

Two additional, but somewhat less plausible, biogeographical scenarios may be inferred from these observations: (i) capelin evolved in the Pacific Ocean, invaded the Atlantic on only one occasion, diverged within the Atlantic Ocean to form the two Atlantic clades with the northwest Atlantic clade subsequently re-invading the Pacific to found the present-day Arctic and northeast Pacific clades. This assertion however, requires the additional speculation that ancestral *Mallotus* haplotypes in the Pacific are either extinct or were not sampled; (ii) capelin evolved in the Atlantic Ocean and diverged to form the two Atlantic clades with the northwest Atlantic clade subsequently invading the Pacific for the first time to found the present-day Arctic and Pacific clades. This assertion is consistent with biogeographical scenarios proposed for some marine fishes. The Arctic–Pacific divergence in capelin mirrors the two geographical races of Pacific herring (*Clupea pallasi*) that are divided by the Alaskan peninsula. Pacific herring is considered to be derived from an Atlantic invader (Grant 1986). In the case of gadine fishes (cods, pollocks, haddock, whiting and pouts), their centre of endemism is the eastern North Atlantic (Carr *et al*. 1999). The three endemic Pacific species represent distinct lineages that invaded the Pacific on two or three independent occasions (Carr *et al*. 1999; Coulson *et al*. 2006). Thus, an Atlantic origin for capelin followed by an invasion of the Pacific is a plausible scenario. Although the results presented here do not permit us to unequivocally support either a Pacific or an Atlantic origin of capelin, the former appears to be the more likely scenario as the Pacific is the centre of endemism of osmerid fishes. Of the 15 valid species comprising Osmeridae (Nelson 1994; J. J. Dodson, J. Laroche & Lecomte F., unpublished), all are restricted to the Pacific Ocean with the exception of capelin and the two *Osmerus* species (*O. mordax, O. eperlanus*) that occur in the Atlantic Ocean, and *O. dentex* that occurs in the Arctic Ocean as well as in the Pacific Ocean.

In conclusion, successive trans-Arctic dispersals of capelin during the Late Pliocene and Pleistocene have contributed to the subdivision of the species into four geographically isolated mtDNA clades. Isolation during long-lasting glacial maxima and expansion during brief interglacials has resulted in extreme heterogeneity in the spatial distribution of intraspecific genetic diversity. These divergent clades have remained separated over several glacial cycles and climate oscillations during which time the range of the species would have dramatically expanded and contracted. During expansion, their extended ranges may have resulted in hybridization on several occasions, but evidence of this may have been lost during glacial advances extinguishing such peripheral populations (Hewitt 2004). Some of the capelin clades may now be reproductively isolated. The most recent contraction–expansion episodes of sufficient magnitude to have impacted capelin's molecular architecture are inferred to have occurred relatively recently, coincident with Pleistocene glaciation cycles. No evidence of genetic structuring was seen among populations comprising the northeast-central Atlantic or the northwest Atlantic clades. In contrast, such structure was observed among populations comprising the Arctic and the northeast Pacific clades suggesting the existence of closely related populations within the distributional area of these two clades. It appears evident that trans-Arctic dispersals and climate oscillations have enhanced opportunities for evolutionary diversification and have contributed to the process of speciation within the genus *Mallotus*.

Acknowledgements

This study was supported by the Science subvention programme of the Canadian Department of Fisheries and Oceans (DFO) and by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada to J.J. Dodson. We are very thankful to all those who provided samples: Joe Firth (DFO, St John's, Newfoundland), Dr Anthony Gaston (Canadian Wildlife Service, University of Ottawa), James Murphy (NOAA/NMFS/Alaska Fisheries Science Center), Robert J. Foy and Katryn Brenner (University of Alaska, Fairbanks), Mayumi Arimitsu (USGS-BRD), Per Kanneworff (Greenland Institute of Natural Resources), Martine Giangioppi. (DFO, Central & Arctic Region) and Lubov A. Skurikhina (Institute of Marine Biology, Far Eastern Branch of Russian Academy of Science, Vladivostok, Russia). We thank Maxime Leblond for laboratory assistance. We also acknowledge the Centre de Bioinformatique, Université Laval and Dr Jerôme Laroche for assistance in computing phylogenetic reconstructions. We thank Drs Louis Bernatchez and Julie Turgeon for advice during this project and for providing constructive criticisms of earlier drafts of this paper. Finally, we thank Drs Stew Grant and Brian Bowen, and anonymous referees for providing significant commentary on earlier versions of this paper.

References

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 457–463.

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial-DNA lineage survivorship in animal populations. *Journal of Molecular Evolution*, **20**, 99–105.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 1707–1712.

Berggren WA, Hollister CD (1974) Paleogeography, paleobiogeography and the history of circulation in the Atlantic Ocean. *Society of Economic Paleontologists and Mineralogists Special Publication*, **20**, 126–186.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.

Birt TP, Friesen VL, Birt RD, Green JM, Davidson WS (1995) Mitochondrial DNA variation in Atlantic capelin, *Mallotus villosus*: a comparison of restriction and sequence analyses. *Molecular Ecology*, **4**, 771–776.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfish (*Aulostomus* spp.): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Briggs JC (1970) A faunal history of the North Atlantic Ocean. *Systematic Zoology*, **19**, 19–34.

Briggs JC (1995) *Global Biogeography*. Elsevier, Amsterdam.

Burton MPM, Flynn SR (1998) Differential post-spawning mortality among male and female capelin (*Mallotus villosus* Müller) in captivity. *Canadian Journal of Zoology*, **76**, 588–559.

Carr SM, Kivlichan DS, Pepin P, Crutcher DC (1999) Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. *Canadian Journal of Zoology*, **77**, 19–26.

Carscadden JE, Frank KT, Leggett WC (2001) Ecosystem changes and the effects on capelin (*Mallotus villosus*), a major forage species. *Canadian Journal of Fisheries and Aquatic Science*, **58**, 73–85.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Collins TM, Frazer K, Palmer AR, Vermeij GJ, Brown WM (1996) Evolutionary history of northern hemisphere Nucella (*Gastropoda*, Muricidae): molecular, morphological, ecological, and paleontological evidence. *Evolution*, **50**, 2287–2304.

Coope GR (1979) Late Cenozoic fossil Coleoptera: evolution, biogeography and ecology. *Annual Review of Ecology and Systematics*, **10**, 247–267.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome*, **49**, 1115–1130.

Cronin TM (1988) Evolution of marine climates of the United States Atlantic coast during the past 4 million years. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **318**, 661–678.

Dodson JJ, Carscadden JE, Bernatchez L, Colombani F (1991) Relationship between spawning mode and phylogeographic structure in mitochondrial DNA of North Atlantic capelin *Mallotus villosus. Marine Ecology Progress Series*, **76**, 103–113.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Frank KT, Simon J, Carscadden JE (1996) Recent excursions of capelin (*Mallotus villosus*) to the Scotian Shelf and Flemish Cap during anomalous hydrographic conditions. In: *Impact of Anomalous Oceanographic Conditions at the Beginning of the 1990s in the Northwest Atlantic on the Distribution and Behaviour of Marine Life Symposium*. NAFO Scientific Council Studies no. 24.

Gaggiotti OE, Excoffier L (2000) A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **267**, 81–87.

Gjøsæter H (1998) The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. *Sarsia*, **83**, 453–496.

Gladenkov AY, Gladenkov YB (2004) Onset of connections between the Pacific and Arctic Oceans through the Bering Strait in the Neogene. *Stratigraphy and Geological Correlation*, **12**, 175–187.

Gladenkov AY, Oleinik AE, Marincovich L, Barinov KB (2002) A refined age for the earliest opening of Bering Strait. *Palaeogeography Palaeoclimatology Palaeoecology*, **183**, 321–328.

Govindarajan AF, Halanych KK, Cunningham CW (2005) Mitochondrial evolution and phylogeography in the hydrozoan *Obelia geniculata* (Cnidaria). *Marine Biology*, **146**, 213–222.

Grant WS (1986) Biochemical genetic divergence between Atlantic, *Clupea harengus*, and Pacific, *C. pallasi*, herring. *Copeia*, **1986**, 714–719.

Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, **89**, 415–426.

Grant WS, Bowen BW (2006) Living in a tilted world: climate change and geography limit speciation in old world anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society*, **88**, 673–689.

Harris SA (2005) Thermal history of the Arctic Ocean environs adjacent to North America during the last 3.5 Ma and a possible mechanism for the cause of the cold events (major glaciations and permafrost events). *Progress in Physical Geography*, **29**, 218–237.

Harrison MK, Crespi BJ (1999) Phylogenetics of Cancer crabs (Crustacea: Decapoda: Brachyura). *Molecular Phylogenetics and Evolution*, **12**, 186–199.

Hewitt GM (1996) Some genetic consequences of ice ages and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 183–195.

Hillis DM, Moritz C, Marble BK (1996) *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Huelsenbeck JP, Rannala B (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, **276**, 227–232.

Jansson R, Dynesius M (2002) The fate of clades in a world of recurrent climate change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, **33**, 741–777.

Johnson GD, Patterson C (1996) Relationships of lower euteleostean fishes. In: *Interrelationships of fishes* (eds Stiassny ML, Parenti LR, Johnson GD), pp. 251–332. Academic Press, San Diego.

Kleiven HF, Jansen E, Fronval T, Smith TM (2002) Intensification of Northern Hemisphere glaciations in the circum Atlantic region (3.5–2.4 Ma)—ice-rafted detritus evidence. *Palaeogeography Palaeoclimatology Palaeoecology*, **184**, 213–223.

Klicka J, Zink RM (1999) Pleistocene effects on North American songbird evolution. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **266**, 695–700.

Klyukanov VA, McAllister DE (1979) Osmeridae. In: *Checklist of the Fishes of the Northeastern Atlantic and of the Mediterranean*, Vol. 1 (eds Tortonese E, Hureau JC), pp. 158–159. UNESCO, Paris.

Kocher TD, Thomas WK, Meyer A *et al*. (1989) Dynamics of mitochondrial-DNA evolution in animals — Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, USA*, **86**, 6196–6200.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Marincovich L, Gladenkov AY (1999) Evidence for an early opening of the Bering Strait. *Nature*, **397**, 149–151.

Marincovich L, Gladenkov AY (2001) New evidence for the age of Bering Strait. *Quaternary Science Reviews*, **20**, 329–335.

McAllister D (1963) *A Revision of the Smelt Family, Osmeridae*, National Museum of Canada, Ottawa.

Nechaev VP, Sorochinskaya AV, Tsoy IB, Gorbarenko SA (1994) Clastic components in Quaternary sediments of the Northwest Pacific and their paleo-oceanic significance. *Marine Geology*, **118**, 119–137.

Nelson J (1994) *Fishes of the World*, 3rd edn. John Wiley & Sons, New York.

Nikula R, Strelkov P, Väinölä R (2007) Diversity and trans-Arctic invasion history of mtDNA lineages in the North Atlantic *Macoma Balthica* complex (Bivalvia: Tellinidae). *Evolution*, **61**, 928–941.

Pagé P (1999) *Les Grandes Glaciations: L'histoire et la Stratigraphie Des Glaciations Continentales Dans L'hémisphère Nord*. Guérin, Montréal.

Palumbi SR, Kessing BD (1991) Population biology of the trans-Arctic exchange: mitochondrial DNA sequence similarity between Pacific and Atlantic sea urchins. *Evolution*, **45**, 1790–1805.

Pigeon D, Dodson JJ, Bernatchez L (1998) A mtDNA analysis of spatio-temporal distribution of two genetically-distinct sympatric larval populations of rainbow smelt (*Osmerus mordax*) in the middle estuary of the St. Lawrence R., Québec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1739–1747.

Posada D, Crandall KA (1998) MODELTEST (version 3.6): testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Sher A (1999) Traffic lights at the Beringian crossroads. *Nature*, **397**, 103–104.

Stergiou KI (1989) Capelin *Mallotus villosus* (Pisces: Osmeridae), glaciations and speciation: a nomothetic approach to fisheries ecology and reproduction biology. *Marine Ecology Progress Series*, **56**, 211–224.

Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0. Beta 10. Sinauer Associates, Sunderland, Massachusetts.

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution*, **12**, 823–833.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Taylor EB, Dodson JJ (1994) A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). *Molecular Ecology*, **3**, 235–248.

Väinölä R (2003) Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. *Marine Biology*, **143**, 935–946.

Vermeij GJ (1989) Invasion and extinction: the last three million years of North Sea pelecypod history. *Conservation Biology*, **3**, 274–281.

Vermeij GJ (1991) Anatomy of an invasion—the trans-Arctic interchange. *Paleobiology*, **17**, 281–307.

Vilhjálmsson H (1994) The Icelandic capelin stock. Rit Fiskideildar. *Journal of the Marine Research Institute, Reykjavik*, **XIII**, 282 pp.

Wares JP (2001) Biogeography of *Asterias*: North Atlantic climate change and speciation. *Biological Bulletin*, **201**, 95–103.

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

Xia X, Xie Z (2001) DAMBE: data analysis in molecular biology and evolution. *Journal of Heredity*, **92**, 371–373.

Julian Dodson's laboratory is dedicated to the study of the evolutionary ecology of a variety of freshwater and marine fishes and to the application of these fundamental principles to better manage aquatic resources. This work is part of the M.Sc. thesis project conducted by Suzy Tremblay under the direction of J.J. Dodson. Her research interests include molecular evolution and marine phylogeography. Francoise Colombani is a molecular biologist who is responsible for maintaining J.J. Dodson's laboratory on the straight and narrow for all things molecular. James Carscadden is a senior research scientist with Canada's Department of Fisheries and Oceans specialising in the biology and management of capelin in the North-west Atlantic Ocean. Frederic Lecomte's major interests are the molecular ecology and evolution of smalll pelagic fishes.
