Molecular Ecology (2001) **10**, 965–985

Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation

G. LU,* D. J. BASLEY† and L. BERNATCHEZ*
*GIROQ, Université Laval, Département de Biologie, Ste-Foy, Québec, Canada G1K 7P4, †Maine Department of Inland Fisheries and Wildlife, Ashland, Maine, USA

Abstract

We performed a combined analysis of mitochondrial DNA (mtDNA) and microsatellite loci among lake whitefish (*Coregonus clupeaformis*) populations in order to assess the levels of congruence between both types of markers in defining patterns of genetic structuring, introgressive hybridization and inferring population origins in the hybrid zone of the St. John River basin. A second objective was to test the hypothesis that secondary contact between glacial lineages always resulted in the occurrence of sympatric dwarf and normal whitefish ecotypes. Fish were sampled from 35 populations and polymorphism was screened at mtDNA and six microsatellite loci for a total of 688 and 763 whitefish, respectively. Four lakes harbouring a single whitefish population of normal ecotype admixed with mtDNA haplotypes of different lineages were found. This confirmed that secondary contact between whitefish evolutionary lineages did not always result in the persistence of reproductively isolated ecotypes. Microsatellites further supported the definition of distinct glacial lineages by identifying lineage-specific allelic size groups. They also further supported the hypothesis that ecotypes originated from either a single founding lineage (sympatric divergence) or following secondary contacts between lineages (allopatric divergence), depending on the lake. In general, however, the pattern of population differentiation and introgressive hybridization observed at microsatellites was in sharp contrast with that depicted by mtDNA variation. Both factorial correspondence analysis and analysis of admixture proportion revealed a much more pronounced pattern of introgressive hybridization than depicted by mtDNA analyses. Variable levels of introgression indicated that environmental differences may be as important as the historical contingency of secondary contact in explaining the persistence of sympatric ecotypes and the differential pattern of introgressive hybridization among lakes. Whitefish populations from the St. John River basin hybrid zone represent a rare illustration of a continuum of both morphological and genetic differentiation within a given taxon, spanning from complete introgression to possibly complete reproductive isolation, depending on lakes. Thus, each lake may be viewed as a different temporal snapshot taken throughout the gradual process of speciation.

Keywords: *Coregonus*, hybrid zone, introgression, microsatellites, mtDNA, speciation

Received 2 July 2000; revision received 9 November 2000; accepted 9 November 2000

Introduction

Phylogeographic studies have contributed importantly to our understanding of evolutionary history, a fundamental goal in evolutionary biology (Avise 2000). A clear success of phylogeography has been the improved description of geographical distribution, phylogenetic relationships and genetic distances among evolutionary lineages (Bermingham & Moritz 1998). The great majority (> 80%) of phylogeographic studies has relied on the analysis of

Correspondence: L. Bernatchez. Fax: +1 418 656 2043; E-mail: Louis.Bernatchez@bio.ulaval.ca

mitochondrial DNA (mtDNA) variation, mainly because of its rapid evolution and maternal transmission without intermolecular recombination (Avise 2000). The reliability of such studies, however, has often been criticized because the mitochondrial genome represents only a minuscule fraction of the total historical record within a sexual organismal pedigree (e.g. Degnan 1993; Palumbi & Baker 1994). This may potentially lead to erroneous interpretations of phylogeographic patterns, namely in situations in which introgressive hybridization may be involved (Rieseberg 1998).

Assessing the congruence among independent genetic markers has therefore become an important issue in phylogeography (Avise 2000). Indeed, an increasing number of studies is inferring relationships among populations based on the combined analysis of different DNA regions (e.g. Bernatchez & Osinov 1995; Glémet *et al*. 1998; Buonaccorsi *et al*. 1999; Oyler McCance *et al*. 1999; Ross *et al*. 1999; Allendorf & Seeb 2000). While concordant results will provide a robust perspective on species evolutionary history (Cummings *et al*. 1995), findings of incongruent patterns at different markers would also be valuable because they may provide a better understanding of important evolutionary processes such as the role of selection and introgressive hybridization in speciation (Dowling & Secor 1997; Rieseberg 1998).

In this study, we performed a combined analysis of mtDNA and microsatellite loci to infer the evolutionary history and the extent of genetic differentiation of lake whitefish (*Coregonus clupeaformis*) populations in a zone of extensive secondary contact between distinct glacial races. The whitefish *Coregonus* sp. (Salmonidae) has a continuous circumpolar distribution in the northern hemisphere, exhibiting extreme phenotypic variation in morphological characteristics (Nelson 1984). Palearctic populations are generally referred to as European whitefish *C. lavaretus*, whereas Nearctic populations are referred to as lake whitefish (*C. clupeaformis*) (Bernatchez & Dodson 1994). Previous mtDNA phylogeographic surveys have revealed that this species complex is composed of five major evolutionary lineages that evolved allopatrically in distinct glacial refugia during the Pleistocene (Bernatchez & Dodson 1994). These studies also indicated that secondary intergradation among these lineages has been limited to several geographically disjointed areas (Bernatchez & Dodson 1990, 1991, 1994; Bodaly *et al*. 1992; Foote *et al*. 1992; Bernatchez *et al*. 1996).

The St. John River basin in northeastern North America is one such zone that has been studied intensively over recent years (reviewed in Bernatchez *et al*. 1999). It is characterized by the occurrence of several lakes harbouring so-called dwarf and normal sympatric whitefish ecotypes (Fenderson 1964; Lindsey *et al*. 1970; Kirkpatrick & Selander 1979; Bernatchez & Dodson 1990; Chouinard *et al*. 1996; Pigeon *et al*. 1997; Bernatchez *et al*. 1999). Dwarf whitefish mature by the age of 1 or 2 years and seldom live beyond their fourth year, whereas the normal form does not generally mature until 4 years of age and may reach 12 years. In fact, a strong bimodal distribution in sizes of sexually maturing fish is the first indication for the existence of sympatric whitefish forms (Chouinard *et al*. 1996). Dwarf fish seldom exceed 20 cm in length and 100 g in weight, whereas the normal fish commonly exceed 40 cm and 1000 g. Generally speaking, dwarf and normal ecotypes primarily occupy limnetic and benthic trophic niches, respectively.

Previous mtDNA studies indicated that the occurrence of sympatric ecotypes in the St. John River basin could be the result of either secondary contact between distinct glacial races or incipient divergence from a single founding population, depending on lakes (Bernatchez & Dodson 1990, 1991; Pigeon *et al*. 1997). The main weakness of these previous studies, however, is that they have not been corroborated by analysis of nuclear genes. In this context, a first objective of this study was to assess the levels of congruence between mtDNA and microsatellite analyses in defining patterns of population structuring and inferring the historical origin of whitefish populations in the St. John River zone of secondary contact.

More recent studies revealed that the extent of gene flow between sympatric dwarf and normal whitefish ecotypes varied from lake to lake (Pigeon *et al*. 1997; Lu & Bernatchez 1999). This pattern was also correlated with the potential for trophic niche partitioning between ecotypes and the extent of their morphological specialization observed in different lakes (Bernatchez *et al*. 1999; Lu & Bernatchez 1999). This raised the hypothesis that environmental differences among lakes may be as important as historical contingency (secondary contact among distinct evolutionary lineages) in explaining the persistence of sympatric ecotypes, as well as their level of divergence. The finding of lakes that were colonized by two glacial lineages, but harbouring a single ecotype representing a hybrid swarm, would provide additional support to this hypothesis. A second objective of this study was therefore to test the hypothesis that secondary contact between distinct glacial races always results in the occurrence of sympatric pairs of whitefish ecotypes. A related objective was to assess the levels of congruence between mtDNA and microsatellite analyses in quantifying the extent of introgression between distinct glacial lineages.

By assessing those specific objectives, we also aimed at more generally illustrating how the analysis of microsatellite and mtDNA variation may reveal very contrasting patterns of population structuring in situations of pronounced introgressive hybridization, and lead to very different inference of population origins.

Materials and methods

Sampling

Fish were sampled from 35 populations in lakes of northeastern North America (Table 1). Twenty-eight of these populations were collected in the contact zone between the Atlantic and Acadian lineages, that is the St. John River drainage of northern Maine and southestern Quebec (Bernatchez & Dodson 1991). A total of 688 and 763 whitefish were used for mtDNA and microsatellite analyses, respectively. This included specimens that had been analysed previously (Bernatchez & Dodson 1990, 1991; Pigeon *et al*. 1997; Bernatchez *et al*. 1999; Lu & Bernatchez 1999), as well as 354 and 299 new specimens for mtDNA and microsatellite analyses. Tissues of liver or muscle collected from these specimens were preserved in 95% ethanol and subsequently used to extract total DNA as described in Bernatchez *et al*. (1992).

mtDNA analysis

Mitochondrial DNA variation was analysed using restriction fragment length polymorphism (RFLP) performed on polymerase chain reaction (PCR) products, which ensure compatibility with previous data sets. Mitochondrial segments comprised ND5/6 subunits of the NADH dehydrogenease gene, the cytochrome *b* gene and the control region (D-loop). Restriction enzymes used were *Alu*I, *Cfo*I, *Ava*II, *Dde*I, *Mbo*I, *Hin*cII, *Hae*III, *Rsa*I, *Ban*I and *Bsp*1286I. Procedures for PCR amplification, enzyme digestion and electrophoresis were as described in Bernatchez & Osinov (1995) and Pigeon *et al*. (1997). Identification of endonuclease patterns and mtDNA haplotypes were as detailed in Pigeon *et al*. (1997). In addition, a 559-bp mtDNA D-loop segment from one individual of each haplotype defined by RFLP was also sequenced in order to compare the congruence in defining major evolutionary lineages using either RFLP or sequence analyses. The D-loop region was PCR-amplified using the primers Tpro-2 (5′-ACCCTTAACTCCCAAAGC-3′) (Gilbert *et al*. 1988) and HN20 (5′-GTGTTATGCTTT-AGTTAAGC-3′) (Bernatchez & Danzmann 1993). PCR conditions were as described in Bernatchez *et al*. (1992). These primers are, respectively, located in the proline and the phenylalanine transcriptional RNA (tRNA) genes. The PCR products were purified using the QIAGEN gel extraction kit, cycle sequencing reactions were carried out with the Tpro-2 primer using the DNA sequencing kit (Perkin–Elmer Applied Biosystems), and nucleotide sequences were determined using an ABI 377 DNA sequencer.

Microsatellite analysis

Nuclear DNA variation was characterized at six microsatellite loci (Table 3) with dinucleotide repeats specifically developed for *Coregonus*. Primer sequences were 5′-GATCAGAGAA-ATACACACAACGCATCAA-3′ and 3′-CACGAGTCA-TTACCTTGGAGAC-5′ for locus *Bwf1* (Patton *et al*. 1997), 5′-GGGATACATCGGCAACCTCTG-3′ and 3′-AAAAGAGT-AACCCCTGACAGA-5′ for *Bwf2* (Patton *et al*. 1997), 5′-GAGAGGGGGTATGTCTGT-3′ and 3′-CACCAATGATTT-GAGGCTA-5′ for *Cocl22* (Bernatchez 1996), 5′-GCTGTATG-AGGATAGCATTC-3′ and 3′-TGTGTTTTGCTGGATTACG-5′ for *Cocl23* (Bernatchez 1996), 5′-CTTAGATGATGGCTT-GGCTCC-3′ and 3′-CCACAACATTCTCACTAACGTGG-5′ for *C2-157* (Turgeon *et al*. 1999) and 5′-GCTTGTTGAGG-GCGTTTACC-3′ and 3′-CGAACCACTTTGTAGACACAC-5′ for *C4-157* (Turgeon 2000). Procedures for PCR amplification, electrophoresis, autoradiography and determination of allelic size were as described in Lu & Bernatchez (1999).

Data analyses

Phylogeographic relationships among mtDNA haplotypes were assessed using the restriction site presence–absence matrix to generate a phylogenetic tree according to Wagner parsimony criteria using the MIX program of PHYLIP 3.5c (Felsenstein 1993). A majority-rule consensus tree and confidence statements on branches were obtained using the CONSENSUS program performed on 10 000 different trees generated by MIX from bootstrap replicates produced by the SEQBOOT program.

Mitochondrial gene diversity was estimated by the number of haplotypes, nucleon (haplotype) diversity (h) and nucleotide diversity (π) indices of Nei & Tajima (1981), using the DA program of REAP package (McElroy *et al*. 1992). Relationships among all populations were assessed by using the resulting distance matrix of net nucleotide divergence to construct a neighbour-joining dendrogram using the NEIGHBOUR program of PHYLIP (Felsenstein 1993). Bootstrapping values were computed with the MTDST program (Danzmann *et al*. 1991), using 1000 randomized data sets produced by the SEQBOOT of PHYLIP (Felsenstein 1993).

Hierarchical analyses of molecular variance (AMOVA) (Excoffier *et al*. 1992) was performed using ARLEQUIN 1.1 (Schneider *et al*. 1997). Groups were defined by their glacial origin, i.e. Mississippian, Atlantic and Acadian (see Results). Populations were assigned to a given group based on the fixation or a predominant proportion of mtDNA haplotypes characteristic of a given glacial race (Bernatchez & Dodson 1991). Separate analyses were conducted for allopatric and sympatric populations. All analyses were performed excluding samples with $n < 10$. Analyses were conducted both with and without nucleotide divergence between pairs of haplotypes taken into account.

Microsatellite gene diversity was quantified as the number of alleles per locus (A), observed heterozygosity (H_O) and unbiased gene diversity (H_E) using GENETIX 4.0 (Belkhir 1999). Departures from Hardy–Weinberg equilibrium

Table 1 Sampling information and mtDNA haplotype distribution of lake whitefish in eastern North America. Haplotypes are grouped according to phylogenetic assemblages representative of three whitefish lineages

				N		mtDNA haplotype(absolute frequency)		
Sampling location	Latitude	ongitude	Ecotype	nDNA	mtDNA	Mississippian	Atlantic	Acadian
Mira River	47°0′40″	65°51′34″	Normal*	14	12			25(12)
Aylmer Lake	45°50′00″	71°26′00″	Normal	24	26	1(20)	17(5), 117(1)	
Champlain Lake	45°03′00″	73°09′00″	Normal	20	23	1(23)		
East Lake	47°11′00″	69°33′00″	Dwarf*	41	48			25(18), 106(30)
			Normal*	41	41			25(5), 106(33), 107(3)
Matane Lake	48°42′40″	67°4′41″	Normal		12	1(11), 128(1)		
Pohénégamook Lake	47°29′00″	69°16′00″	Normal	21	21			25(21)
St. Lawrence River	47°46′40″	69°15′00″	Normal	20	24	1(13), 122(4)	17(5), 21(1), 123(1)	
Témiscouata Lake	47°36′00″	68°45′00″	Dwarf*	44	32			25(25), 26(3), 28(1), 111(2),112(1)
			Normal*	35	27	1(4)		25(21), 109(1), 110 (1)
Allagash Lake	46°19′24″	69°31′23″	Normal	5	5		21(1), 119 (1)	25(3)
Big Pleasant Lake	46°28′24″	69°10′48″	Normal	20	23		21(22), 118(1)	
Carr Pond	46°45′58″	68°43′29″	Normal	8	8		21(1)	25(7)
Clear Lake	46°31′14″	69°7′44″	Normal	16	17		21(7), 121(10)	
Cliff Lake	46°23′59″	69°15′11″	Dwarf*	40	30			25(28), 105(1), 116(1)
			Normal*	40	30		21(29), 117(1)	
Crescent Pond	46°25′36″	69°36′20″	Dwarf*	42	19		21(2), 117(1)	25(15), 110 (1)
			Normal*	20	8		21(3), 117(1)	25(4)
Echo Lake	46°25′53″	69°6′57″	Normal	20	24			25(24)
Harrow Lake	46°30′45″	69°12′47″	Normal	20	20		21(20)	
Haymock Lake	46°18′8″	69°11′10″	Normal	20	19		21(9)	25(10)
Indian Pond	46°15′27″	69°17′29″	Dwarf*	44	47			25(29), 107(9), 124(2), 125(4), 127(1)
			Normal*	37	20			25(19), 124(1)
Ross Lake	46°30′43″	69°37′26″	Normal	20	24		21(23), 129(1)	
Rowe Lake	46°37′29″	68°48′42″	Normal	8	8		21(8)	
Sebago Lake	43°52′6″	70°35′48″	Normal	13	13			25(13)
South Pond	44°23′4″	70°40′53″	Normal	6	6	1(6)		
Spider Lake	46°26′57″	69°12′37″	Dwarf*	8	8		21(6), 130(1)	25(1)
			Normal	5	5		21(2)	25(2), 133(1)
St. Francis Lake	46°45′14″	69°45′14″	Normal	6	6			25(6)
Umsaskis Lake	46°35′41″	69°23′04″	Dwarf	8	8		120(8)	
			Normal	3	3		21(1)	25(2)
Wesbter Lake	46°09′21″	69°04′45″	Dwarf*	40	19		21(3)	25(15), 124(1)
			Normal*	40	40		21(30), 119(2)	25(5), 111(1), 110(1), 131(1)
West Grand Lake	45°14′23″	67°50′32″	Normal	14	14			25(13), 132(1)

*Population previously studied by either mitochondrial (mt)DNA (Bernatchez and Dodson 1990, 1991, 1994; Chouinard *et al*. 1996; Bernatchez *et al*. 1999) or microsatellite analyses (Bernatchez *et al*. 1999; Lu & Bernatchez 1999). nDNA, nuclear DNA.

(HWE) were tested with the specific alternative hypothesis of either heterozygote deficiency or excess using the score (*U*) test of Rousset & Raymond (1995) available in GENEPOP 3.1d. Significance values were computed for each locus by unbiased estimates of Fisher's exact test using the Markov chain method through 1000 iterations (Guo & Thompson 1992). Global tests across loci were also made. The probability values were adjusted for multiple simultaneous table-wide tests using the sequential Bonferroni adjustments to minimize type-I errors (Rice 1989). Genetic differentiation between pairs of populations was first quantified by estimates of pairwise fixation indices based on variance in allelic frequencies (θ of Weir & Cockerham 1984) using FSTAT, version 1.2 (Goudet *et al*. 1996). Genetic differentiation was also estimated using mutational differences among alleles and assuming the stepwise mutation model by computing pairwise standardized R_{ST} estimate with the program R_{ST} Calc (Goodman 1997). We conducted AMOVAS using the same hierarchical designs described for mtDNA analyses. The analyses were also conducted excluding small samples ($n < 10$).

The $(\delta\mu)^2$ distance (Goldstein *et al.* 1995) and the Chord distance, D_{CE} (Cavali-Sforza & Edwards 1967), were used to construct phylogenetic trees using the neighbour-joining (NJ) algorithm (Saitou & Nei 1987) and implemented in the NJBPOP program (J.M. Cornuet, INRA, France: http://www.ensam.inra.fr/URLB). The robustness of tree topologies was assessed by replicating 2000 bootstrap over loci.

In order to compare the congruence in genetic relationships among populations inferred from mtDNA and microsatellite markers, we tested the significance of the correlation between matrices of mtDNA nucleotide divergence, the D_{CE} genetic distance and the $(\delta\mu)^2$ genetic distance using the Mantel test (Mantel 1967) available in GENETIX 4.0.

The extent of population differentiation based on microsatellite allelic frequency was further assessed by performing a factorial correspondence analysis (FCA) using GENETIX 4.0. The principle of FCA has been described in Benzécri (1973), and She *et al*. (1987) adopted the method to analyse genetic data. Briefly, genetic data are transformed into a contingency table (samples × alleles), in which each sample is described by the allelic frequencies. The χ^2 distance centred on the marginal distribution of the contingency table is used to measure the relatedness between any two samples in the *k*-dimensional space (*k* = number of alleles). The resulting factorial axes can be ordered according to their eigenvalues and the first axis corresponding to the largest eigenvalue explains the more general pattern or structure contained in the data set. The ordination of populations along these axes can a priori be used to express which populations are the most different or the most similar for a given axis (Guinand 1996). All factorial analyses were conducted excluding small samples ($n < 10$).

Admixture proportions of different evolutionary lineages within each population in the zone of secondary contact were estimated from the microsatellite allelic composition of putative source populations by means of the weighted least squares (WLS) using the program ADMIX (Long 1991). Ideally, source populations of distinct lineages should have been obtained from outside the contact zone. This was not feasible here, however, especially for the Atlantic race for which pure populations outside the zone of secondary contact may no longer exist (Bernatchez & Dodson 1991). We thus considered that populations with highest FCA values on the first axis of FCA 1 (Ross, Clear and Cliff normal ecotype, see Results) were the most representative of the Atlantic lineage source population as also determined by mtDNA analyses, whereas those with lowest FCA values (Crescent, East and Cliff dwarf ecotype) were considered to represent the Acadian lineage source population. The combined allelic composition of the putative source populations was subsequently used to quantify their admixture proportion within the other populations found in the contact zone. An obvious potential bias in such a procedure comes from the possible presence of nonancestral alleles in putative source populations, which would cause an underestimation of the admixture proportion in other populations. Thus, our estimations are likely to be conservative.

In order to further investigate whether the observed levels of admixture were indeed the result of variable levels of introgressive hybridization among distinct lineages, alleles found in populations of the contact zone were combined for each locus, and their size distributions were observed to identify allelic size groups potentially representative of each glacial lineages. With reference to allelic size distribution within the putative Atlantic (Ross, Clear and Cliff normal ecotype) and Acadian (Crescent, East and Cliff dwarf ecotype) source populations, we then defined allelic size modes that were characteristic of both lineages at three of six loci (see Results). Alleles of a given modal size and belonging to the same glacial lineage were combined and considered as a single 'lineage-specific allelic size group'. The proportion of lineage-specific allelic size groups in each population was averaged over the three loci. We then performed a linear correlation between the proportion of putative Acadian alleles and the admixture proportion of the Acadian lineage allelic size group among populations from the contact zone using STATISTICA 5.1 (StatSoft Inc.). Again, small samples ($n < 10$) were excluded from this analysis.

Results

mtDNA phylogenetic differentiation

A total of 30 composite haplotypes was identified, which differed by one to nine restriction sites (Fig. 1). Haplotypes

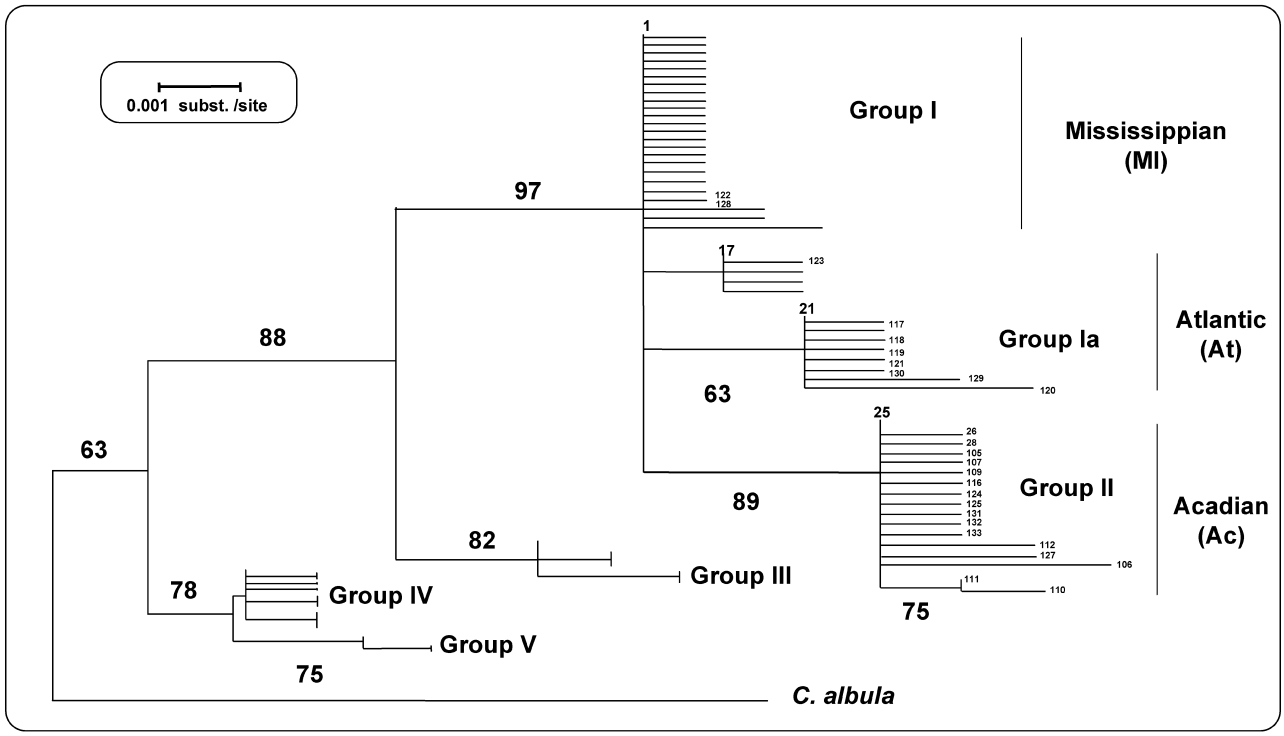

Fig. 1 Condensed majority-rule tree clustering of 30 mitochondrial mtDNA (mtDNA) haplotypes observed in this study (branch tips with numbers) with all other haplotypes defined previously in Bernatchez & Dodson (1994). Bootstrap estimates (in %) are given along the branches. The scale of 0.001 substitution/site corresponds to one restriction site difference. Three mtDNA phylogenetic groups distinguished three glacial lineages (Mississippian, Atlantic and Acadian) of whitefish based on phylogenetic and/or geographical distinctiveness reported in previous studies.

118–133 had not been described previously, whereas the other haplotypes were reported in Bernatchez & Dodson (1994) and Pigeon *et al*. (1997). All haplotypes belonged to either one of the mtDNA clades defined previously (reviewed in Bernatchez *et al*. 1999), which included all mtDNA haplotypes found thus far in North America outside Beringia (Alaska and the Yukon). These clades characterized three glacial lineages of whitefish, based on both their phylogenetic and/or geographical distinctiveness: the Acadian, the Atlantic and the Mississippian (Fig. 1) (Bernatchez & Dodson 1991, 1994; Pigeon *et al*. 1997; Bernatchez *et al*. 1999). The Mississippian lineage was typically distributed outside the St. John River basin (Fig. 2). The clade comprising the largely predominant haplotype 21, as well as haplotypes 117–121, 129–130, and characteristic of the Atlantic lineage exhibited a localized distribution in lakes found in the Allagash River of the St. John River basin. The Acadian lineage was also observed mainly in the St. John River basin but extended eastward to Nova Scotia. This lineage was largely dominated by the single haplotype 25. Both Atlantic and Acadian lineages co-occurred in the St. John River basin. Admixture of haplotypes characteristic of both lineages was observed in four lakes (Aylmer, Carr, Allagash and Haymock lakes) harbouring a single population (always of normal ecotype) (Fig. 2).

Sequence polymorphism in the D-loop region was completely congruent with RFLP analyses in defining distinct evolutionary lineages, with the exception of haplotype 17 (Table 2). A total of 559 bp was sequenced, which contained 14 bp in the proline tRNA gene and the adjacent 545 bp at the 5′-end of the D-loop region. D-loop sequence polymorphism was very low, with only seven variable sites detected. Haplotypes of the Mississippian lineage differed from those of the Atlantic by a single position (24), whereas two apomorphies (positions 29 and 88) distinguished the Acadian lineage from the other two. RFLP haplotype 17, which had previously been associated with the Atlantic lineage (Bernatchez & Dodson 1991), had a D-loop sequence identical to that of haplotype 1, characteristic of the Mississippian lineage. This haplotype was, however, absent in the contact zone, and consequently, this ambiguity had no consequence in interpreting patterns of population differentiation and introgressive hybridization in this area.

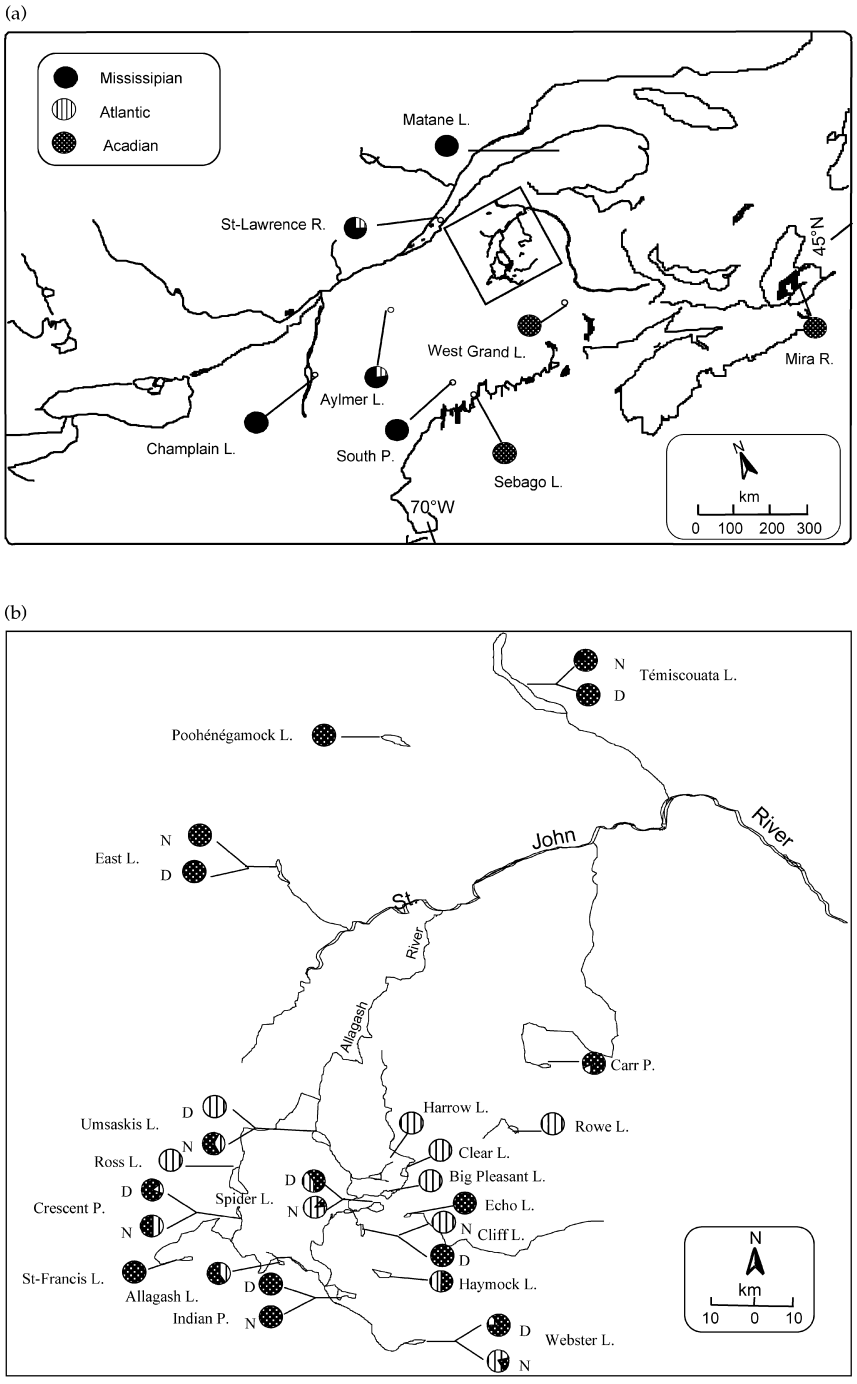

Fig. 2 Geographic distribution of major mitochondrial DNA phylogenetic groups representing three glacial lineages (Acadian, Atlantic and Mississippian) among lake whitefish populations in northeastern North America. Sympatric populations include dwarf (D) and normal (N) ecotypes. (a) The whole sampling region in northeastern North America; (b) the secondary contact zone of Atlantic and Acadian lineages in southern Quebec and northern Maine.

mtDNA diversity and genetic differentiation among populations

The number of haplotypes, mtDNA haplotype and nucleotide diversity within populations varied from minimal to moderate (Table 3). The number of haplotypes per sample was 2.5 on average (range: 1–6). The mean haplotype diversity was 0.286 (range: 0.000–0.800), whereas mean nucleotide diversity was 0.0008 (range: 0.000–0.0028). Based on their mtDNA haplotype composition, whitefish populations formed three distinct clusters that were well supported by bootstrapping values (Fig. 3). These clusters corresponded to the predominance of mtDNA haplotypes characteristic of distinct glacial lineages. The Mississippian

Table 2 Variable nucleotide positions of haplotypes resolved by sequence analysis of a 5′-end 545 bp segment of the mtDNA d-loop region from lake whitefish. Sequence for haplotype 1 has been deposited in GenBank (Accession no. AF239253). For other haplotypes, variable nucleotides are given when different from haplotype 1, while identity is indicated by dots. Haplotypes are grouped according to phylogenetic assemblages representative of three whitefish lineages, Mississippian, Atlantic and Acadian

Phylogenetic group	Haplotypes	Variable nucleotide site (related to position of the D-loop 5′-end)						
		24	29	58	88	240	297	489
Mississippian								
	1	C	G	A	A	A	G	T
	108*	.	.	.	.	.	.	.
	113*, 114*, 115*	.	.	G	.	.	.	.
Atlantic								
	17	.	.	.	.	.	.	.
	21, 118	G	.	.	.	.	A	.
	119, 120, 121, 129, 130	G	.	.	.	.	.	.
	123	G	.	.	.	.	.	C
Acadian								
	25, 106, 109, 111, 110, 112, 124, 127, 131	.	A	.	G	.	.	.
	26, 28, 133	.	A	.	G	.	.	C
	125	.	A	.	G	G	.	.

*Haplotypes described in Pigeon *et al*. (1997). DNA samples for sequence analysis were not available for the following haplotypes: 105, 107, 116, 117, 122, 128, and 132.

cluster encompassed only allopatric populations, whereas the Acadian and Atlantic clusters comprised both allopatric and sympatric populations. Sympatric populations from a given lake belonged either to separate (e.g. Cliff and Webster) or the same (e.g. East) population cluster.

The overall level of genetic differentiation among populations was pronounced, as revealed by both F_{ST} [0.496 ± 0.338 (SD)] and Φ_{ST} [0.554 ± 0.357 (SD)] estimates. Hierarchical AMOVA revealed that genetic differentiation among groups defined as glacial lineages was important relative to that observed among populations within groups (Table 4). This translated into 65.0 and 81.7% of the total genetic variance among sympatric and allopatric populations attributed to differences among glacial lineages, respectively.

Microsatellite gene diversity, population differentiation, and patterns of introgression

Variable levels of genetic diversity were observed within samples. The mean number of alleles per locus averaged five (range: 1–15). This translated into gene diversity estimates averaging 0.546 and varying from 0.000 to 0.855 (Table 3). The null hypothesis of HWE was rejected at the 5% level in 14 of 204 cases following the Bonferroni sequential adjustment for multiple tests, a number slightly higher than expected by chance. These departures were due mainly to a clustered heterozygote deficit at *C4–157* (seven of 11 deficits detected; *P*-value of global test across populations < 0.00001), strongly suggesting the presence of a null allele at this locus. This is consistent with our previous findings (Lu & Bernatchez 1999). Prior to performing additional statistical analyses, we thus approximated the frequencies of a null allele in populations detected with heterozygote deficits as described in Lu & Bernatchez (1999).

No significant association was observed by Mantel tests either between the $(\delta\mu)^2$ and the mtDNA net nucleotide divergence (Pearson $r = 0.160$; $P = 0.239$) or between the D_{CE} and the mtDNA net nucleotide divergence (Pearson $r = 0.125$; $P = 0.150$). This indicated the overall lack of congruence in population differentiation inferred from mtDNA and microsatellite markers. Similarly, neighbour-joining dendrograms constructed using either D_{CE} or $(\delta\mu)^2$ genetic distances showed striking differences in tree topologies compared with the one built with mtDNA net nucleotide divergence (Fig. 4). The $(\delta\mu)^2$ dendrogram revealed no significant grouping among populations (all bootstrap values were < 50%). Populations alternatively fixed for mtDNA haplotypes that were characteristic of distinct glacial lineages clustered closely, such as Sebago (AC) with Cliff normal ecotype (AT) and Clear (AT), or Echo (AC) with Ross (AT) and Rowe (AT). There were, however, several elements of congruence between the $(\delta\mu)^2$ and mtDNA topologies. For example, populations from East Lake always clustered together, as did populations of Aylmer (MI), Champlain (MI) and St. Lawrence (MI).

Table 3 Summary of variation at mtDNA and microsatellites from lake whitefish populations in eastern North America: number of haplotypes (*n*), haplotype diversity (h), nucleotide diversity (π), number of alleles at each locus (A), most common allele (AC), frequency of the most common allele (FC), range of allelic size (AR), observed heterozygosity (H_O) and gene diversity (H_E)

	Allagash	Aylmer	Big Pleasant[D]	Carr	Champlain	Clear	Cliff (D)	Cliff (N)	Crescent (D)	Crescent (N)	East[D] (D)	East (N)
mtDNA												
n	3	3	2	2	1	2	3	2	4	3	2	3
h	0.700	0.385	0.087	0.250	0	0.515	0.131	0.067	0.380	0.679	0.479	0.340
π	0.003	0	0	0.001	0	0.001	0	0	0.001	0.002	0.001	0.001
BWF1												
A	5	9	5	3	11	6	9	4	7	6	8	6
AC	220	216	220	220	216	214	220	214	220	220	220	220
FC	0.38	0.46	0.48	0.88	0.48	0.41	0.44	0.70	0.68	0.76	0.61	0.61
AR	212–224	206–226	214–226	216–220	204–228	212–226	212–228	210–220	208–228	214–228	212–216	212–224
H_O	0.50	0.51	0.65	0.25	0.75	0.81	0.93	0.54	0.53	0.37	0.62	0.51
H_E	0.75	0.63	0.59	0.23	0.73	0.75	0.75	0.45	0.52	0.42	0.60	0.57
BWF2												
A	2	6	2	2	3	1	3	3	6	3	6	4
AC	157	163	157	157	155	147	157	157	157	157	157	157
FC	0.63	0.42	0.78	0.88	0.53	1.00	0.91	0.69	0.82	0.84	0.55	0.91
AR	147–157	147–167	147–157	147–157	147–163	147	153–161	147–157	141–161	147–161	147–159	147–159
H_O	0.25	0.87	0.33	0.25	0.55	0	0.18	0.48	0.29	0.21	0.67	0.15
H_E	0.47	0.68	0.35	0.22	0.57	0	0.16	0.44	0.32	0.28	0.63	0.16
C2–157												
A	2	15	9	8	11	5	9	7	8	4	9	10
AC	147	147	147	147	145	167	147	147	147	147	147	145
FC	0.88	0.35	0.23	0.31	0.33	0.50	0.51	0.68	0.62	0.46	0.39	0.29
AR	147–159	145–179	143–167	143–163	145–173	147–167	133–167	147–169	143–171	147–171	121–167	121–167
H_O	0.25	0.87	0.70	0.88	0.80	0.75	0.68	0.60	0.58	0.43	0.63	0.63[d]
H_E	0.22	0.82	0.84	0.81	0.83	0.64	0.70	0.49	0.55	0.62	0.76	0.82
C4–157												
A	2	9	7	5	10	5	6	3	5	4	7	5
AC	293	277	293	293	293	293	293	297	293	293	293	293
FC	0.50	0.47	0.39	0.50	0.38	0.31	0.50	0.72	0.59	0.76	0.68	0.36
AR	293–297	273–299	281–297	289–297	275–301	289–297	281–301	293–301	289–301	289–301	285–301	285–301
H_O	0.56	0.95[e]	0.39[d]	0.50	0.60	0.50	0.49	0.39	0.27	0.42	0.30[d]	0.66
H_E	0.50	0.69	0.77	0.66	0.79	0.70	0.66	0.44	0.40	0.67	0.52	0.76
COCL22												
A	2	5	3	2	9	3	4	3	6	4	6	6
AC	121	121	121	121	123	121/123	123	123	123	123	125	123
FC	0.50	0.50	0.48	0.63	0.45	0.47	0.60	0.41	0.49	0.55	0.55	0.46
AR	121–123	119–127	121–127	121–123	107–129	119–123	117–127	121–127	117–127	119–127	105–127	105–127
H_O	0.57	0.53	0.95	0.75	0.70	0.94	0.75	0.70	0.65	0.81	0.89[e]	0.76
H_E	0.50	0.56	0.62	0.47	0.65	0.56	0.57	0.57	0.58	0.55	0.60	0.64

Table 3 *Continued*

	Allagash	Aylmer	Big Pleasant[D]	Carr	Champlain	Clear	Cliff (D)	Cliff (N)	Crescent (D)	Crescent (N)	East[D] (D)	East (N)
COCL23												
A	3	6	2	4	7	3	4	4	4	4	5	6
AC	266	264	266	262	270	262	266	262	266	266	266	266
FC	0.5	0.28	0.58	0.63	0.38	0.54	0.59	0.58	0.82	0.72	0.63	0.55
AR	260–266	258–270	262–266	258–266	252–272	262–268	260–266	260–266	262–268	264–270	260–270	260–270
H_O	0.75	0.62	0.42	0.63	0.80	0.77	0.75	0.65	0.36	0.56	0.59	0.58
H_E	0.59	0.78	0.49	0.56	0.74	0.56	0.55	0.6	0.32	0.46	0.56	0.58

	Echo	Harrow	Haymock	Indian[D] (D)	Indian[D] (N)	Mira	Pohénégamook	Ross	Rowe	Sebago (D)	South	Spider
mtDNA												
n	1	1	2	5	2	1	1	2	1	1	1	3
h	0	0	0.600	0.585	0.100	0	0	0.083	0	0	0	0.464
π	0	0	0.002	0.001	0	0	0	0	0	0	0	0.001
BWF1												
A	5	10	7	7	5	5	9	6	4	6	3	5
AC	220	214	220	220	220	216	220	220	220	214	232	220
FC	0.40	0.33	0.37	0.5	0.38	0.33	0.44	0.5	0.56	0.27	0.67	0.31
AR	212–222	212–238	208–224	212–224	212–224	216–232	208–228	212–224	214–226	210–224	212–232	214–224
H_O	0.90	0.95	0.79	0.57	0.53	0.89	0.72	0.78	0.75	0.85	0.50	0.84
H_E	0.70	0.83	0.73	0.70	0.73	0.75	0.73	0.68	0.60	0.80	0.49	0.76
BWF2												
A	3	2	3	5	4	4	5	3	1	3	3	2
AC	151	157	157	157	157	157	157	147	157	147	147	157
FC	0.55	0.68	0.53	0.84	0.78	0.45	0.57	0.72	1.00	0.69	0.67	0.50
AR	151–159	153–157	147–159	147–161	147–161	147–159	147–163	145–151	157	145–151	147–159	147–157
H_O	0.65	0.25	0.30	0.16[d]	0.35	0.79	0.62	0.44	0	0.46	0.67	0.50
H_E	0.60	0.22	0.56	0.29	0.36	0.68	0.60	0.43	0	0.46	0.49	0.50
C2–157												
A	6	5	9	12	5	5	9	5	7	6	5	6
AC	147	147	147	147	147	147	147	147	147	147	145/147	147
FC	0.63	0.48	0.55	0.57	0.84	0.40	0.33	0.53	0.25	0.50	0.33	0.56
AR	143–163	143–159	147–169	147–171	145–159	145–163	143–163	143–163	147–163	145–159	143–159	147–163
H_O	0.70	0.70	0.75	0.73	0.06[d]	0.70	0.90	0.63	0.88	0.85	0.50	0.75
H_E	0.56	0.67	0.66	0.65	0.29	0.69	0.80	0.65	0.83	0.68	0.74	0.64
C4–157												
A	3	10	6	10	6	2	3	5	4	7	4	4
AC	297	291/293	297	293	297	293	293	293	297	293	291	293
FC	0.50	0.24	0.45	0.50	0.66	0.55	0.57	0.44	0.50	0.69	0.58	0.63
AR	293–301	277–307	289–303	278–303	273–305	293–297	289–297	231–301	231–297	283–301	291–303	283–297
H_O	0.56	0.68	0.74	0.58[d]	0.41	0.59	0.38	0.61	0.64	0.53	0.67	0.58
H_E	0.62	0.83	0.67	0.71	0.55	0.50	0.54	0.66	0.60	0.50	0.60	0.54

Table 3 *Continued*

	Echo	Harrow	Haymock	Indian[D] (D)	Indian[D] (N)	Mira	Pohénégamook	Ross	Rowe	Sebago (D)	South	Spider
COCL22												
A	3	5	6	7	5	5	3	7	3	6	3	4
AC	123	123	123	123	123	123	123	123	123	121	121	123
FC	0.53	0.43	0.55	0.61	0.62	0.60	0.64	0.50	0.75	0.42	0.58	0.50
AR	121–127	117–127	117–127	109–129	119–129	107–131	121–127	103–129	121–129	117–131	115–123	121–127
H_O	0.85	0.90	0.74	0.67	0.73	0.40	0.67	0.79	0.50	0.85	0.83	0.88
H_E	0.59	0.71	0.63	0.59	0.55	0.59	0.52	0.68	0.41	0.68	0.54	0.65
COCL23												
A	3	3	4	6	5	3	3	1	2	2	2	2
AC	262	266	262	266	266	262	266	262	266	262	266/278	266
FC	0.9	0.87	0.76	0.57	0.68	0.67	0.68	1.00	0.75	0.54	0.50	0.69
AR	262–268	262–266	258–266	256–270	260–270	258–272	262–270	262	262–266	262–266	266–278	262–266
H_O	0.20	0.16	0.26	0.70	0.42	0.33	0.63	0	0.25	0.15[d]	0.33	0.63
H_E	0.18	0.23	0.39	0.58	0.49	0.47	0.48	0	0.38	0.50	0.50	0.43

	Spider (N)	St. Francis	St. Lawrence[D]	Témiscouata (D)	Témiscouata (N)	Umsaskis (D)	Umsaskis (N)	Webster[D] (D)	Webster[D] (N)	West Grand	Matane
mtDNA											
n	3	1	5	5	4	1	2	3	6	2	2
h	0.800	0	0.659	0.387	0.385	0	0.667	0.368	0.428	0.143	0.167
π	0.003	0	0.001	0	0.001	0	0.003	0.001	0.002	0	0
BWF1											
A	6	1	4	8	10	6	2	8	6	5	
AC	220	220	214	220	214	212	212	220	214	216	
FC	0.40	1.00	0.35	0.58	0.31	0.25	0.50	0.49	0.38	0.39	
AR	212–224	220	208–216	212–226	208–226	212–226	212–220	210–224	212–226	210–224	
H_O	0.84	0	0.50	0.68	0.69	0.82	0.60	0.72	0.60	0.93	
H_E	0.76	0	0.72	0.63	0.77	0.79	0.50	0.70	0.73	0.73	
BWF2											
A	2	2	7	4	5	3	2	5	2	2	
AC	157	157	155	157	157	157	157	157	147	157	
FC	0.63	0.83	0.70	0.64	0.57	0.88	0.67	0.82	0.74	0.89	
AR	147–157	147–157	147–165	147–159	147–163	147–159	147–157	147–161	147–157	147–157	
H_O	0.25	0.33	0.45	0.48	0.54	0.25	0.34	0.29	0.33	0.21	
H_E	0.47	0.28	0.48	0.56	0.59	0.23	0.44	0.32	0.39	0.19	
C2–157											
A	5	5	13	8	9	3	2	12	6	6	
AC	147	147	149	147	147	147	147	147	147	163	
FC	0.38	0.42	0.25	0.83	0.63	0.81	0.83	0.54	0.84	0.38	
AR	147–165	145–167	127–175	143–163	145–163	147–153	147–163	143–171	145–171	145–167	

Table 3 *Continued*

	Spider (N)	St. Francis	St. Lawrence[D]	Témiscouata (D)	Témiscouata (N)	Umsaskis (D)	Umsaskis (N)	Webster[D] (D)	Webster[D] (N)	West Grand	Matane
H_O	0.84	0.67	0.91	0.31	0.54	0.38	0.33	0.70	0.33	0.75	
H_E	0.75	0.69	0.86	0.30	0.58	0.32	0.28	0.69	0.29	0.74	
C4–157											
A	4	1	10	4	4	3	1	11	5	6	
AC	287	293	299	297	297	297	297	293	297	293	
FC	0.60	1.00	0.28	0.64	0.43	0.75	1.00	0.40	0.38	0.35	
AR	287–299	293	209–311	293–301	289–301	283–301	297	283–303	277–303	287–297	
H_O	0.40	0	0.65[d]	0.14[d]	0.36	0.17	0	0.35[d]	0.33[d]	0.85	
H_E	0.58	0	0.83	0.51	0.66	0.40	0	0.75	0.72	0.77	
COCL22											
A	3	4	6	3	4	3	2	8	5	3	
AC	123	123	123	123	123	123	123	123	123	123	
FC	0.70	0.58	0.53	0.62	0.60	0.44	0.75	0.54	0.46	0.85	
AR	121–127	105–127	107–127	121–127	119–127	121–127	115–123	117–131	119–131	119–127	
H_O	0.60	0.67	0.84	0.65	0.71	0.75	0.50	0.78	0.88[e]	0.15	
H_E	0.46	0.58	0.68	0.54	0.54	0.63	0.38	0.65	0.68	0.27	
COCL23											
A	2	2	6	5	5	2	2	7	5	3	
AC	266	262	262	266	266	266	262	266	262	262	
FC	0.80	0.75	0.58	0.51	0.46	0.94	0.83	0.71	0.38	0.89	
AR	262–266	258–268	260–272	258–268	158–168	262–266	262–266	252–270	262–272	260–266	
H_O	0.40	0.50	0.67	0.63	0.69	0.13	0.33	0.41	0.53	0.21	
H_E	0.32	0.38	0.60	0.63	0.65	0.12	0.28	0.49	0.6	0.20	

Superscripts e and d indicate significant heterozygote excess and deficit, respectively, following the sequential Bonferroni correction ($\alpha = 0.05$; $k = 204$), whereas superscripts D indicates significant heterozygote deficit (global test, Fisher's method) prior to correcting for the presence of a null allele at *C4–157*. (D): dwarf ecotype (N): noraml ecotype.

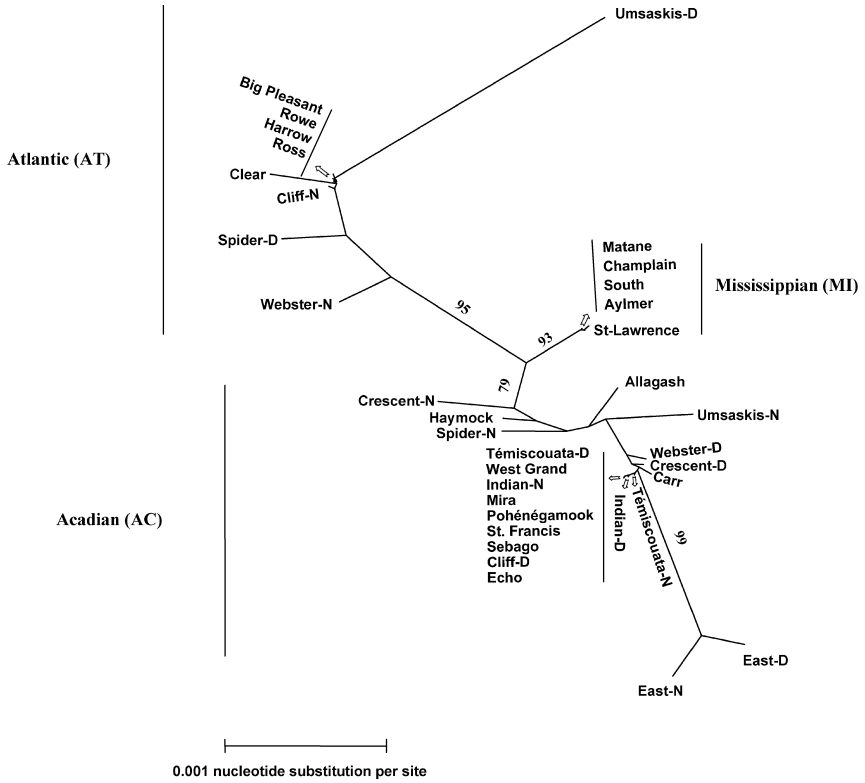

Fig. 3 Neighbour-joining dendrogram inferred from mitochondrial DNA random fragment length polymorphism analyses, showing population relationships of lake whitefish from northeastern North America, according to the distance matrix resulting from net nucleotide divergence between population. Bootstrapping estimates > 50% are given along branches. The three population clusters correspond to distinct glacial population groups: Mississippian (MI), Atlantic (AT) and Acadian (AC). Letters D and N following population names indicate sympatric dwarf and normal ecotypes, respectively.

Source of variation	mtDNA	Microsatellites
Groups defined by glacial origin		
Allopatric populations		
Among groups	Φ_{CT} 0.817 (< 0.001)	ρ_{CT} 0.071 (0.065)
Among populations within groups	Φ_{SC} 0.372 (< 0.001)	ρ_{SC} 0.253 (< 0.001)
Sympatric populations		
Among groups	Φ_{CT} 0.650 (0.016)	ρ_{CT} 0.153 (< 0.001)
Among populations within groups	Φ_{SC} 0.432 (< 0.001)	ρ_{SC} 0.038 (0.022)

Table 4 Hierarchical analysis of molecular variance (AMOVA) based on mtDNA and microsatellite data. Glacial origin were determined based on mtDNA phylogenetic relationships among populations. Populations with sample size < 10 were excluded from the analyses. *P*-values are given in parentheses

Population relationships were somewhat better supported by bootstrap values in the D_{CE} dendrogram than the $(\delta\mu)^2$ one (Fig. 4). Yet, little congruence with the mtDNA population tree was depicted. As for the $(\delta\mu)^2$ dendrogram, no distinct Acadian and Atlantic population clusters were defined. Populations characterized by the same mtDNA lineage did not always cluster more closely relative to those possessing different mtDNA haplotypes, e.g. Sebago (AC) and Echo (AC). Only two small population groups were supported by > 50% bootstrapping values; the three Mississippian populations (Aylmer, Champlain and St. Lawrence), and three Acadian populations (Crescent dwarf and normal ecotypes, and Cliff dwarf ecotype). There were, however, several elements of congruence between the D_{CE} distance tree and the mtDNA tree in sympatric populations. Namely, dwarf and normal populations of Cliff Lake were still genetically distant, whereas those of East Lake still clustered together.

The overall θ and R_{ST} estimates were moderate and almost identical, being 0.161 ± 0.109 and 0.172 ± 0.137, respectively. The AMOVA showed both incongruent and congruent results derived from microsatellite and mtDNA markers. A reverse pattern of differentiation was observed in allopatric populations (Table 4). Thus, the genetic differentiation based on microsatellites among glacial lineages defined by mtDNA was not significant and represented

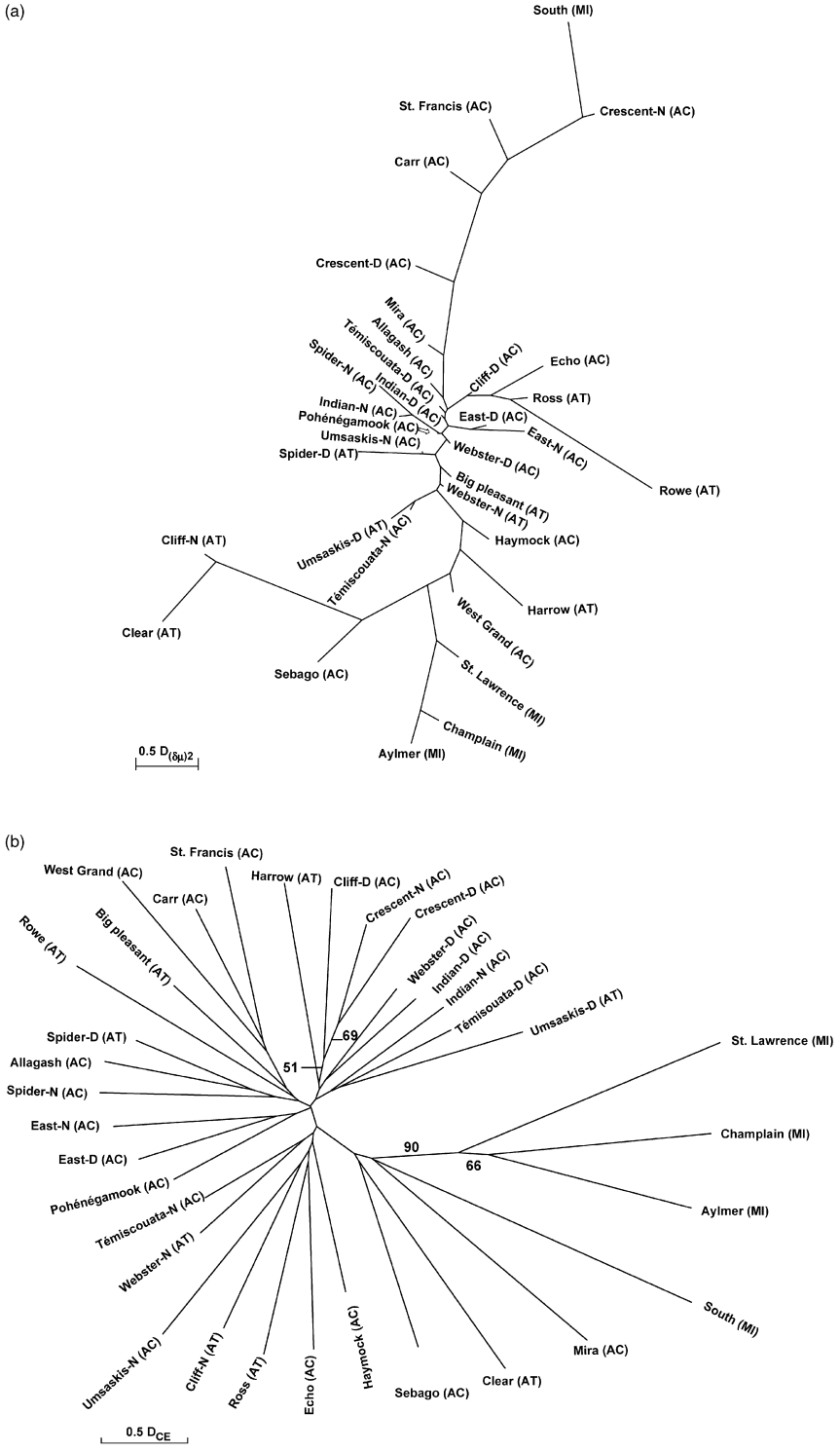

Fig. 4 Neighbour-joining dendrogram inferred from microsatellite data, illustrating population relationships of lake whitefish form northeastern North America, according to the $(\delta\mu)^2$ genetic distance (a) and Chord genetic distance (D_{CE}) (b). Only bootstrap values > 50% are provided. The abbreviations glacial groups, as defined by mitochondrial DNA random fragment length polymorphism analyses: MI, Mississippian; AT, Atlantic, AC, Acadian. Letters D and N following population names indicate sympatric dwarf and normal ecotypes, respectively.

less than a third of the component of genetic variance observed among populations within each group. For sympatric populations, genetic differentiation among glacial groups was much more important than that observed among populations within group.

A more detailed pattern of population differentiation was revealed by the FCA of microsatellite allelic composition. The first two axes explained 30.8% of the total genetic variance. Mississippian populations and the geographically remote Mira River (AC) populations were very

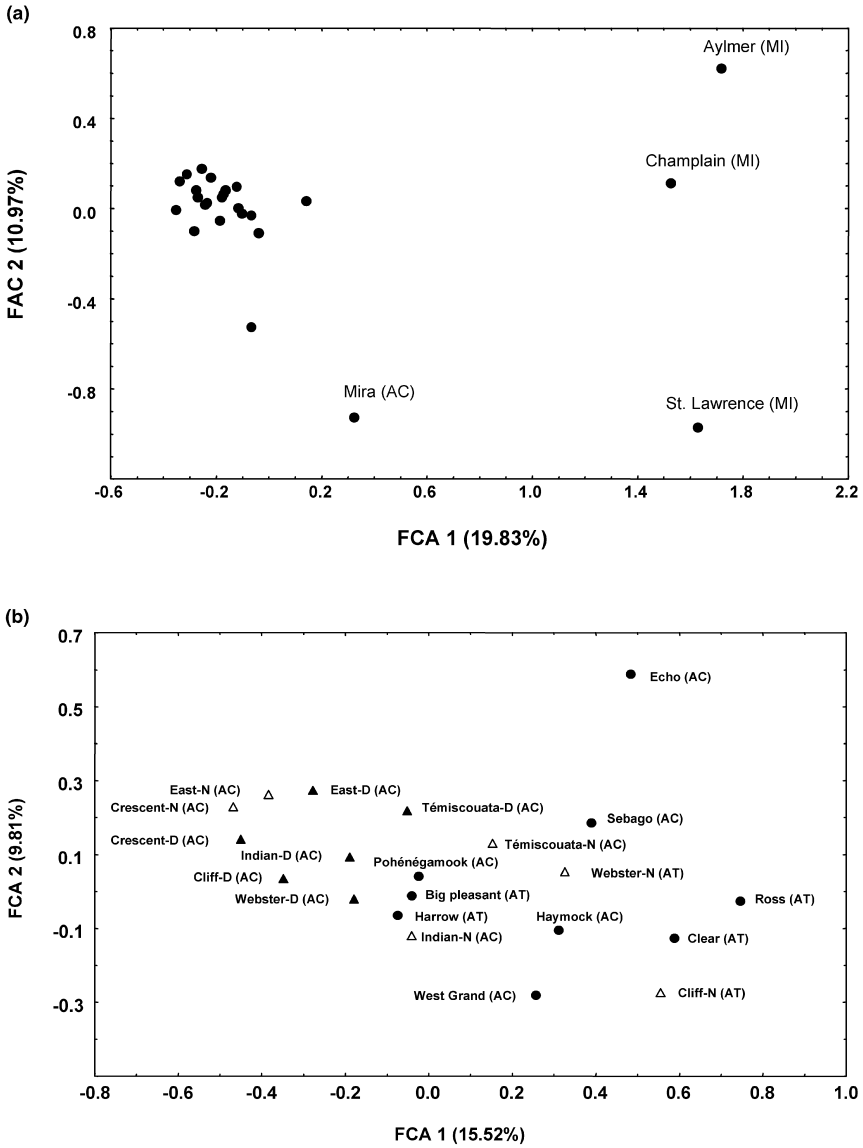

Fig. 5 Genetic differentiation among whitefish populations in northeastern North America based on factorial correspondence analysis (FCA) of allelic frequency at six microsatellite loci. (a) All samples, (b) populations of Mississippian origin and from Mira River are omitted. Populations with sample sizes < 10 were not included in this analysis. In (b) closed circles, closed triangles and open triangles represent allopatric populations, sympatric dwarf populations, sympatric normal populations, respectively. The meanings of abbreviations in parentheses and letters D and N following population names are the same as in Fig. 4.

distinct from all populations of the St. John River drainage (Fig. 5a). We performed a second FCA excluding these populations. The first two axes still explained 25.3% of the total variance (Fig. 5b). Populations with the lowest FCA 1 values were also either fixed (East Lake dwarf and normal ecotypes, Cliff Lake dwarf ecotype) or largely dominated (Crescent Lake dwarf and normal ecotypes) by mtDNA haplotypes characteristic of the Acadian lineage. Similarly, those with highest FCA 1 values (Ross, Clear and Cliff normal ecotype) were all fixed for haplotypes characteristic of the Atlantic lineage. Populations with intermediate FCA 1 values were either fixed for Acadian or Atlantic mtDNA haplotypes or admixed for different mtDNA lineages, and found in either sympatry or allopatry. This pattern of continuum in genetic differentiation was in sharp contrast to the pronounced population clustering observed for mtDNA, and was possibly indicative of variable levels of introgressive hybridization between glacial lineages, depending on lakes.

This was further supported by the variable levels of introgression estimated from putative source populations of different lineages (Fig. 6). Admixture proportion of Acadian origin was lowest (0.407 ± 0.251) in allopatric populations, highest (0.874 ± 0.151) in sympatric dwarf populations and intermediate (0.556 ± 0.397) in sympatric normal populations, respectively. The difference between allopatric and sympatric dwarf populations was significant (*t*-test, $P = 0.0013$), whereas that between allopatric and sympatric normal populations (one-sample *t*-test, $P = 0.3869$) or between sympatric dwarf and normal populations (one-sample *t*-test, $P = 0.096$) was not significant, although a statistical trend was suggested in the latter case.

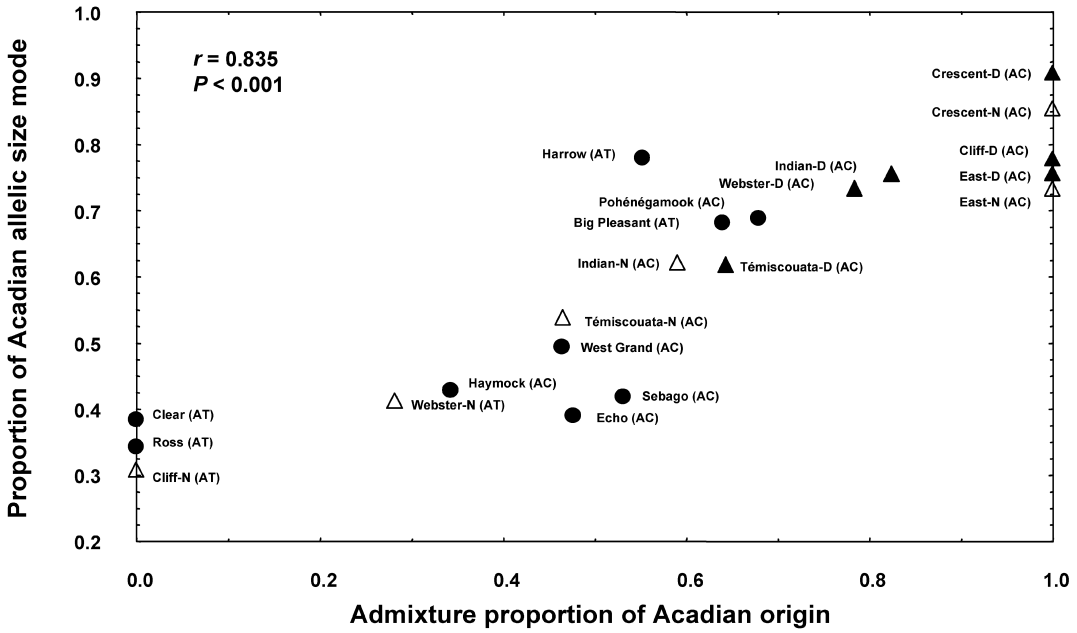

Fig. 6 Correlation between admixture proportion of Acadian origin and proportion of Acadian allelic size modes. Putative Acadian source populations: Crescent Pond, East Lake and Cliff Lake dwarf ecotype; Atlantic putative source populations: Ross Lake, Clear Lake and Cliff Lake normal ecotype. Allopatric, sympatric dwarf, and sympatric normal are depicted by closed circles, closed triangles and open triangles, respectively.

Populations with an approximately equal proportion (near 50%) of both Acadian and Atlantic composition were either allopatric or sympatric, indicating that introgressive hybridization has been as important in lakes now harbouring either a single population or the two genetically differentiated dwarf and normal ecotypes. A bimodal allelic size distribution was observed for three loci (Fig. 7). These defined a lineage-specific allelic size group for which the proportion in each population was highly correlated with the admixture proportion of Acadian lineage quantified from the allelic composition of putative source populations (Pearson $r = 0.835$; $P < 0.001$) (Fig. 6). These results further indicated that the pattern of population differentiation observed in the St. John River basin reflected extensive introgressive hybridization between glacial lineages that varied from lake to lake.

Discussion

Extending mtDNA analyses to additional samples supported previous interpretations of whitefish evolutionary history in northeastern North America. Three mtDNA groups characteristic of different whitefish evolutionary lineages were identified. The Atlantic and Acadian lineages largely overlapped in the St. John River basin, whereas the Mississippian lineage was absent in this region. Members of sympatric pairs of ecotypes were characterized by mtDNA haplotypes of either distinct or the same lineage. The AMOVA revealed that glacial origin was largely responsible

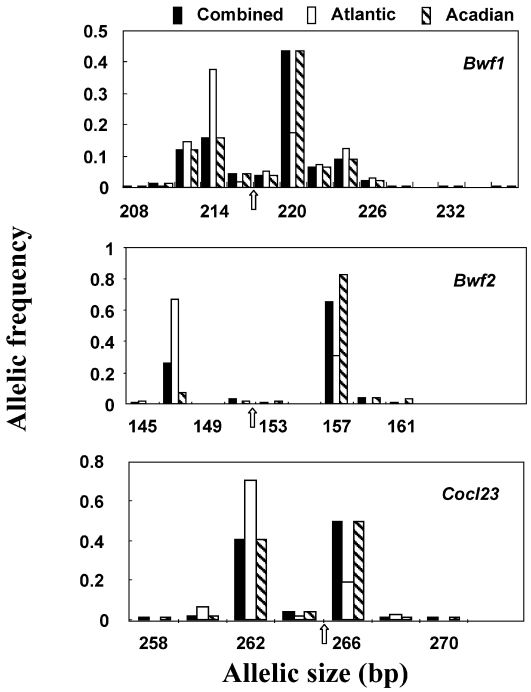

Fig. 7 Frequency histograms of allele sizes for loci *Bwf1*, *Bwf2* and *Cocl23* among lake whitefish populations in the secondary contact zone between the Atlantic and Acadian lineages. A bimodal allelic distribution was observed with each mode predominating within putative source populations for the Atlantic (Ross, Clear and Cliff normal ecotype) and Acadian (Crescent, East and Cliff dwarf ecotype) lineages. Arrows show the positions where distinct allelic size modes characteristic of distinct glacial race origin were divided.

for the extent of genetic divergence among either sympatric or allopatric populations. A new and important finding was the identification of lakes harbouring a single whitefish population of normal ecotype that was admixed with mtDNA haplotypes of different lineages. This confirmed that secondary contact between whitefish evolutionary lineages did not always result in the persistence of reproductively isolated ecotypes in lakes from the St. John River drainage.

Microsatellite results further supported the definition of the three glacial lineages by revealing the strong differentiation of Mississippian populations in all analyses, and by identifying lineage-specific allelic size group at three loci for the Acadian and Atlantic lineages. These results further supported the hypothesis that ecotypes originated from either a single founding lineage (sympatric divergence) or following secondary contacts between lineages (allopatric divergence) (Pigeon *et al*. 1997). Thus, an allelic composition typical of the Acadian lineage was observed for both ecotypes in East Lake Alternatively, ecotypes were in some cases clearly characterized by an allelic composition typical of different lineages, such as in Cliff Lake.

Generally speaking, however, the pattern of population differentiation and introgressive hybridization observed at microsatellites within the contact zone of the St. John River basin was in sharp contrast to that depicted by mtDNA variation. There was no significant population grouping associated with mtDNA lineages based on either $(\delta\mu)^2$ or D_{CE} distance. This was further illustrated by the AMOVA, which showed no significant component of genetic variance among allopatric populations imputable to glacial groups defined by mtDNA. The partitioning of genetic variance was also in sharp contrast to that observed for mtDNA in sympatric populations. Lack of genetic differentiation in association with distinct lineages defined by mtDNA was also illustrated by the FCA, in which all populations from the contact zone grouped closely relative to the Mississippian and the geographically remote Mira River (AC) populations. Overall, both the FCA and the analysis of admixture proportion based on microsatellite data revealed a much more pronounced pattern of introgressive hybridization between glacial lineages than was depicted by mtDNA analyses.

Several factors can potentially cause discrepancies in patterns of population differentiation at both markers. Differential mutation rates between markers could have resulted in the more important accumulation of homoplasic mutations in microsatellites than in mtDNA since postglacial recolonization. This would have contributed to the reduction in the extent of population differentiation at microsatellites (e.g. Viard *et al*. 1998). Departures from the stepwise mutation model implying even a small proportion of random mutations could also erase an important part of the mutation process, thereby strongly reducing the extent of population differentiation measured by R_{ST} (or ρ_{ST}) (Slatkin 1995). However, the predominance of the same microsatellite alleles in many populations found in physically isolated lakes suggests that the accumulation of new mutations has not been important since postglacial recolonization, ≈ 12 000 years ago. Limited accumulation of new mutations on a similar time scale was also reported in the brook charr (*Salvelinus fontinalis*) (Angers & Bernatchez 1998), and is congruent with the rule-of-thumb of 2000 generations since population founding for mutations to significantly effect patterns of population differentiation at microsatellites (Estoup & Angers 1998). Underestimation of population divergence based on microsatellites could also stem from its generally higher level of intrapopulation allelic diversity observed in this study relative to mtDNA. Thus, it has been well documented that the maximum extent of differentiation is negatively correlated with that of allelic diversity at a given locus (e.g. Charlesworth 1998; Hedrick 1999).

Discrepancies between both markers can also be effected by their differential modes of inheritance. The mitochondrial genome is maternally inherited, effectively haploid and its effective population size is theoretically one-quarter that of a nuclear-autosomal gene in equilibrium conditions. The probability of congruence between nuclear and mtDNA genealogies is dependent on the duration of common ancestry and effective population size (Moore 1995). Given that the Acadian and Atlantic glacial lineages have probably been isolated from each other since the Illinoian glacial period (150 000 years ago, Bernatchez & Dodson 1990), and the theoretically smaller effective population size of mtDNA, it is possible that reciprocal monophyly among lineages had time to develop for mtDNA, whereas this period has been too short for this to happen at most microsatellite loci. A stepwise mutation process for which high homoplasy is inherent may also contribute to reduce the probability of allelic reciprocal monophyly at microsatellites.

In addition, discrepancies between mtDNA and microsatellites must have been enhanced by extensive introgressive hybridization between whitefish lineages, and differential effects of genetic drift at both markers following their secondary contact. Although mtDNA clades defining different whitefish lineages comprised many haplotypes, both the Atlantic and Acadian lineages were mainly characterized by a single haplotype in the contact zone (haplotype 21 in the Atlantic and haplotype 25 in the Acadian lineages, respectively, Table 1). Following secondary contact, these haplotypes may have been admixed into the same population by introgressive hybridization. Subsequently, one or other of the haplotypes of different origin may have been lost by the stochastic process of genetic drift. In such a case, many populations may have been incorrectly considered to be of single origin based on their mtDNA makeup. In contrast, six microsatellite loci were screened, each generally characterized by many alleles. In this case, it is unlikely that genetic drift alone would result in a genetic composition

falsely reflecting a single origin following introgressive hybridization (Hartl & Clark 1997).

Thus, while mtDNA was more useful for characterizing evolutionary distinct lineages, microsatellites were more likely to correctly depict introgressive hybridization events following secondary contact. Consequently, FCA based on microsatellite allelic composition showed a continuum of admixture between the Acadian and Atlantic glacial lineages, indicating that most populations in the contact zone were introgressed, even in those cases in which admixture was not depicted by mtDNA analyses. Interestingly, however, microsatellite analyses revealed that the least admixed populations at both ends of this genetic continuum were also generally fixed for mtDNA haplotypes characteristic of distinct races. These populations also tended to be characterized by distinct allelic size modes reflecting the ancestral genetic composition of different glacial lineages at three loci. Overall, the contrasting pattern of population structure between mtDNA and microsatellites illustrates the importance of using multiple markers in interpreting population structure and origins, especially when introgressive hybridization may be involved (Scribner & Avise 1993).

Ecological basis for variable levels of introgression among lakes

Both mtDNA and microsatellites revealed differential levels of introgression between whitefish lineages, depending on the lake. This may possibly result from differential numbers of colonizing individuals from both lineages. Although plausible, this scenario is not supported by the geographical pattern of admixture, in which adjacent lakes (e.g. Big Pleasant Lake and Cliff Lake), which were possibly colonized at the same time and via the same dispersal route, harboured populations with very distinct levels of introgression.

Alternatively, variable levels of introgression among lakes could reflect the differential impact of divergent natural selection on gene flow stemming from the use of alternative environments and/or from resources competition (*sensu* Huxley 1942; Dobzhansky 1951). Environment-dependent hybrid fitness was documented in a comparative study of hybrid three-spined sticklebacks (*Gasterosteus* sp.), in which the distribution of resources available in the natural environment generated a hybrid disadvantage not detectable in the laboratory (Hatfield & Schluter 1999). Divergent natural selection may also have influenced the level of introgression between whitefish lineages by affecting the abundance of hybrid genotypes and their probability of reproducing in a given lake. Thus, recent studies have revealed that divergent natural selection plays an important role in the evolution of whitefish populations in this zone of secondary contact. The demonstration of parallel evolution of dwarf and normal whitefish ecotypes provided strong indirect evidence for the role of divergent natural selection in promoting their phenotypic divergence (Pigeon *et al*. 1997). In addition, a negative correlation was observed between the extent of gene flow occurring between sympatric ecotypes and that of their morphological trophic specialization, indicating that the extent of reproductive isolation reached between whitefish ecotypes was driven mainly by that of resource-based divergent selection in different environments (Lu & Bernatchez 1999). As proposed by Bernatchez *et al*. (1999), a more persistent ecological opportunity to occupy distinct niches throughout the ontogeny in a given lake would select for the development of increased specialization. Under the hypothesis of ecological speciation (Schluter 1998), such selective pressures would be associated with the higher fitness costs of producing hybrids of intermediate phenotype in lakes with high opportunities of occupying distinct niches (such as Cliff Lake), which would be less fit than the parent species in exploiting trophic resources. In contrast, such costs would be expected to be less important in lakes offering less opportunity for niche partitioning, given the more pronounced phenotypic similarity of ecotypes in such circumstances. The extent of reproductive isolation between ecotypes would, therefore, evolve as a consequence of the intensity of selective pressures imposed by the relative cost of producing hybrids in a given lake. The finding of several lakes harbouring a single population of hybrid origin represents an extreme outcome of this continuum of population divergence, whereby both colonizing populations may have introgressed completely as a consequence reduced ecological opportunity to occupy distinct niches throughout the ontogeny (Dobzhansky 1937, 1951; Barton & Hewitt 1981; Liou & Price 1994; Butlin 1995).

In a recent review of hybrid zones as related to speciation, Jiggins & Mallet (2000) showed that contact zones identified among different taxa exemplify a continuum from unimodal to bimodal zones. They argued that multiple intermediate stages may imply that speciation is a gradual, cumulative process, but that relatively little empirical support for this hypothesis exists. Whitefish populations from the St. John River basin hybrid zone represent a rare, if not unique, illustration of a continuum of both morphological and genetic differentiation within a given taxon, ranging from complete introgression to possibly complete reproductive isolation, depending on lakes. Thus, each lake may be viewed as a different temporal snapshot taken throughout the gradual process of speciation.

This continuum also indicates that environmental differences among lakes may be as important as the historical contingency of secondary contact in explaining the persistence of sympatric pairs and the differential pattern of introgressive hybridization among lakes. This will, however, require further empirical investigations of the correlation between production of trophic resources and the extent of population divergence reached in different lakes. It will be

equally important to identify genotype-specific differences in mating preference, fertility or viability, which are often more readily identified in such hybrid settings, and to evaluate their effects on population genetic structure. This hybrid zone also provides a unique system for studying the genetic architecture of reproductive barriers (e.g. identifying the number, location, effects and interactions of genetic factors that contribute to reproductive isolation) (Rieseberg 1998). By identifying genes potentially associated with phenotypic specialization and reproductive isolation, one could then begin to reconstruct the sequence of genetic changes involved in the process of speciation in northern fishes (Coyne 1992; Bradshaw *et al*. 1995). This research is currently in progress.

Acknowledgements

The authors dedicate this article to the late of A. Chouinard who deceased in the field during his research on whitefish. We are also grateful to R. Tardif (Société de la Faune et des Parcs du Québec) for his precious collaboration in sampling, and to J.J. Dodson, T. Kocher and P. Morissette for their constructive criticisms. This work was supported by a research grant from the Natural Science and Engineering Research Council (NSERC, Canada) to L.B. and the Fonds FCAR (Québec) postgraduate scholarships and the Bourse Angélo-Chouinard from Département de biologie, Université Laval to GL Contribution to the programme of GIROQ.

References

Allendorf FW, Seeb LW (2000) Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution*, **54**, 640–651.

Angers B, Bernatchez L (1998) Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, **15**, 143–159.

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

Barton NH, Hewitt GM (1981) *Hybrid Zones and Speciation. Evolution and Speciation and Says in Honour of M.J.D. White*. Cambridge University Press, Cambridge.

Belkhir K (1999) *GENETIX V. 4.0*. Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Montpellier France.

Benzécri JP (1973) *L'Analyse Des Données*. Dunod, Paris.

Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.

Bernatchez L (1996) *Réseau de suivi environnemental du complexe la grande. Caractérisation génétique des formes naines et normales de grand corégone du réservoir caniapiscau et du lac Sérigny à l'aide de marqueurs microsatellites*. Rapport présenté par l'université Laval à la vice-présidence Environement et Collectivités d'Hydro-Québec.

Bernatchez L, Chouinard A, Lu G (1999) Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, Coregonus sp., as a case study. *Biological Journal of the Linnean Society*, **68**, 173–194.

Bernatchez L, Danzmann RG (1993) Congruence in control-region sequence and restriction-site variation in mitochondrial DNA of Brook Charr (*Salvelinus fontinalis* Mitchill). *Molecular Biology and Evolution*, **10**, 1002–1014.

Bernatchez L, Dodson JJ (1990) Allopatric origin of sympatric populations of Lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. *Evolution*, **44**, 1263–1271.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the Lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, **45**, 1016–1035.

Bernatchez L, Dodson JJ (1994) Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 240–251.

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology*, **1**, 161–173.

Bernatchez L, Osinov A (1995) Genetic diversity of trout (genus *Salmo*) from its most eastern native range based on mitochondrial DNA and nuclear gene variation. *Molecular Ecology*, **4**, 285–297.

Bernatchez L, Vuorinen JA, Bodaly RA, Dodson JJ (1996) Genetic evidence for reproductive isolation and multiple origins of sympatric trophic ecotypes of whitefish (*Coregonus*). *Evolution*, **50**, 624–635.

Bodaly RA, Clayton JW, Lindsey CC, Vuorinen J (1992) Evolution of lake whitefish (*Coregonus clupeaformis*) in North America during the Pleistocene: genetic differentiation between sympatric populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 769–779.

Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Minulus*). *Nature*, **376**, 762–765.

Buonaccorsi VP, Reece KS, Morgan L, Graves JE (1999) Geographic distribution of molecular variance within the blue marlin (*Makaira nigricans*): a hierarchical analysis of allozyme, single copy nuclear DNA, and mitochondrial DNA markers. *Evolution*, **53**, 568–579.

Butlin RK (1995) Reinforcement: an idea evolving. *Trends in Ecology and Evolution*, **10**, 432–434.

Cavali-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**, 550–570.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Chouinard A, Pigeon D, Bernatchez L (1996) Lack of specialization in trophicmorphology between genetically differentiated dwarf and normal forms of lake whitefish (*Coregonus clupeaformis* Mitchill) in Lac de l'Est, Québec. *Canadian Journal of Zoology*, **74**, 1989–1998.

Coyne JA (1992) Genetics and speciation. *Nature*, **355**, 511–515.

Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, **12**, 814–822.

Danzmann RG, Ferguson MM, Skúlason S, Snorrason SS, Noakes DLG (1991) Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland. *Journal of Fish Biology*, **39**, 649–659.

Degnan SM (1993) The perils of single gene trees — mitochondrial versus single-copy nuclear DNA variation in white-eyes (Aves: Zosteropidae). *Molecular Ecology*, **2**, 219–226.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Dobzhansky T (1951) *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.

Dowling TE, Secor C (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, **28**, 593–619.

Estoup A, Angers B (1998) Microsatellite and minisatellites for molecular ecology: theoretical and experimental considerations. In: *Advances in Molecular Ecology* (ed. Carvalho G), pp. 55–86. IOS Press, Amsterdam.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Felsenstein J (1993) PHYLIP *(Phylogeny Inference Package), Version 3.5c*. Department of Genetics, SK-50, University of Washington, Seattle, WA.

Fenderson OC (1964) Evidence of subpopulations of lake whitefish, *Coregonus clupeaformis*, involving a dwarf form. *Transactions of the American Fisheries Society*, **93**, 77–94.

Foote CJ, Clayton JW, Lindsey CC, Bodaly RA (1992) Evolution of the lake whitefish (*Coregonus clupeaformis*) in North America during the Pleistocene: evidence for Nahanni glacial refuge race in the northern Cordillera region. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 760–768.

Gilbert TL, Brown JR, O'Hara PJ, Buroker NE, Beckenbach AT, Smith MJ (1988) Sequence of tRNA (Thr) and tRNA (Pro) from white sturgeon (*Acipenser transmontanus*) mitochondria. *Nucleic Acids Research*, **16**, 11825.

Glémet H, Blier P, Bernatchez L (1998) Geographical extent of Arctic char (*Salvelinus alpinus*) mtDNA introgression in brook char populations (*S. fontinalis*) from eastern Québec, Canada. *Molecular Ecology*, **7**, 1655–1662.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

Goodman SJ (1997) R_{ST} Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.

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

Guinand B (1996) Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of the Linnean Society*, **58**, 173–195.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Hartl DL, Clark AG (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates Inc, Sunderland, MA.

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Huxley J (1942) *Evolution, the Modern Synthesis*. Allen & Unwin, London.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250–255.

Kirkpatrick M, Selander RK (1979) Genetics of speciation in lake whitefishes in the Allegash basin. *Evolution*, **33**, 478–485.

Lindsey CC, Clayton JW, Franzin WG (1970) Zoogeographic problems and protein variation in the *Coregonus clupeaformis* whitefish species complex. In: *Biology of Coregonid Fishes* (eds Lindsey CC, Woods CS), pp. 127–146. University of Manitoba Press, Winnipeg, Canada.

Liou LW, Price TD (1994) Speciation by reinforcement of premating isolation. *Evolution*, **48**, 1451–1459.

Long JC (1991) The genetic structure of admixed populations. *Genetics*, **127**, 417–428.

Lu G, Bernatchez L (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution*, **53**, 1491–1505.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

McElroy DP, Moran P, Bermingham E, Kornfield I (1992) REAP: an integrated environment for the manipulation and phylogenetic analysis of restriction data. *Journal of Heredity*, **83**, 157–158.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145–163.

Nelson JS (1984) *Fishes of the World*. John Wiley, New York.

Oyler McCance SJ, Kahn NW, Burnham KP, Braun CE, Quinn TW (1999) A population genetic comparison of large- and small-bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. *Molecular Ecology*, **8**, 1457–1465.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Patton JC, Gallaway BJ, Fechhelm R-G, Cronin MA (1997) Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (*Coregonus nasus*) in the Colville and Sagavanirktok rivers in northern Alasaka. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1548–1556.

Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in the parallel evolution of sympatric dwarf and normal morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, **51**, 196–205.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Rieseberg LH (1998) Molecular ecology of hybridization. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 243–265. IOS Press, Amsterdam.

Ross KG, Shoemaker D, Krieger MJB, Deheer CJ, Keller L (1999) Assessing genetic structure with multiple classes of molecualr markers: a case study involving the introduced fire ant *Solenopsis invicsta*. *Molecular Biology and Evolution*, **16**, 525–543.

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics*, **140**, 1413–1419.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. *Molecular Biology and Evolution*, **4**, 406–425.

Schluter D (1998) Ecological causes of speciation. In: *Endless Forms: Species and Speciation* (ed. Howard D), pp. 114–129. Oxford University Press, Oxford, UK.

Schneider S, Kueffer J-M, Roessli D, Geneva LE (1997) ARLEQUIN. Dept. of Anthropology and Ecology, University of Geneva, CP511, 1211 Geneva 24, Switzeland.

Scribner KT, Avise JC (1993) Cytonuclear genetic architecture in mosquitofish populations and the possible roles of introgressive hybridization. *Molecular Ecology*, **3**, 139–149.

She J-X, Autem M, Kotoulas G, Pasteur N, Bonhomme F (1987) Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biological Journal of the Linnean Society*, **32**, 357–371.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Turgeon J (2000) *Écologie moléculaire et phylogéographie des ciscos d'Amérique du Nord (TELEOSTEI:* Coregonus *spp.)*. PhD Thesis. Université Laval, Québec, Canada.

Turgeon J, Estoup A, Bernatchez L (1999) Species flock in the North American Great Lakes: molecular ecology of Lake Nipigon Ciscoes (Teleostei: Coregonidae: *Coregonus*). *Evolution*, **53**, 1857–1871.

Viard F, Franck P, Dubois MP, Estoup A, Jarne P (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution*, **47**, 42–51.

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

This study is a contribution to a long-term research program on the evolution and systematics of Coregonine species. It is part of Guoqing Lu's PhD thesis which deals with speciation issues in lake whitefish ecotypes from northeastern North America. David Basley is a fisheries biologist working for the Maine Department of Inland Fisheries and Wildlife. The work was carried out under supervision of Louis Bernatchez, whose major research interests focus on the understanding of patterns and processes of molecular and organismal evolution, and its consequences for conservation.
