

A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer

JULIE GODBOUT,* JUAN P. JARAMILLO-CORREA,* JEAN BEAULIEU*† and JEAN BOUSQUET*

*Chaire de recherche du Canada en génomique forestière et environnementale and Centre de recherche en biologie forestière, Pavillon C.-E. Marchand, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4, †Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, 1055 rue du P.E.P.S., PO Box 3800, Sainte-Foy, Québec, Canada G1V 4C7

Abstract

Jack pine (*Pinus banksiana* Lamb.) is a broadly distributed North American conifer and its current range was covered by the Laurentian ice sheet during the last glacial maximum. To infer about the history and postglacial colonization of this boreal species, range-wide genetic variation was assessed using a new and highly variable minisatellite-like marker of the mitochondrial genome. Among the 543 trees analysed, 14 distinct haplotypes were detected, which corresponded to different repeat numbers of the 32-nucleotide minisatellite-like motif. Several haplotypes were rare with limited distribution, suggesting recent mutation events during the Holocene. At the population level, an average of 2.6 haplotypes and a mean haplotype diversity (H) of 0.328 were estimated. Population subdivision of genetic diversity was quite high with G_{ST} and R_{ST} values of 0.569 and 0.472, respectively. Spatial analyses identified three relatively homogeneous groups of populations presumably representative of genetically distinct glacial populations, one west and one east of the Appalachian Mountains in the United States and a third one presumably on the unglaciated northeastern coastal area in Canada. These results indicate the significant role of the northern part of the US Appalachian Mountains as a factor of vicariance during the ice age. A fourth distinct group of populations was observed in central Québec where the continental glacier retreated last. It included populations harbouring haplotypes present into the three previous groups, and it had higher level of haplotype diversity per population ($H = 0.548$) and lower population differentiation ($G_{ST} = 0.265$), which indicates a zone of suture or secondary contact between the migration fronts of the three glacial populations. Introgression from *Pinus contorta* Dougl. var. *latifolia* Engelm. was apparent in one western population from Alberta. Altogether, these results indicate that the mitochondrial DNA variation of jack pine is geographically highly structured and it correlates well with large-scale patterns emerging from recent phylogeographical studies of other tree boreal species in North America.

Keywords: contact zones, glacial populations and refugia, jack pine, mitochondrial DNA, phylogeography, postglacial colonization

Received 4 February 2005; revision received 21 May 2005; accepted 15 June 2005

Introduction

During the Pleistocene, strong climate oscillations affected dramatically the living world, resulting in the repetitive

advance and retreat of the polar ice sheet on circumpolar region. The ensuing cycles of contraction and expansion of natural ranges may have further resulted in reduced genetic diversity and/or altered geographical distribution of genetic diversity in postglacial populations of the surviving species (Hewitt 1996; Perron *et al.* 2000). Phylogeography attempts to unravel spatio-temporal dynamics

Correspondence: Jean Bousquet, Fax: (+1) (418) 656-7493; E-mail: bousquet@rsvs.ulaval.ca

of populations and their consequences for evolution (Avice *et al.* 1987; Cruzan & Templeton 2000; Petit *et al.* 2004). Two complementary approaches have been followed for reconstructing species history: inferences from (i) the fossil record (Davis 1983; Jackson *et al.* 1997) and from (ii) the distribution and geographical patterns of genetic variation of living organisms (Hewitt 1996; Petit & Vendramin 2005). Molecular markers, especially cytoplasmic markers, have been previously used to deduce the location of glacial populations and postglacial colonization paths for different forest tree species in Europe (Dumolin-Lapègue *et al.* 1997; Sinclair *et al.* 1999; Soranzo *et al.* 2000; Gugerli *et al.* 2001; Burban & Petit 2003; Palmé *et al.* 2003; Heuertz *et al.* 2004). For most of the tree species surveyed, the analysis of their population structure indicated fragmentation from disjunct glacial populations or refugia. Recently, a mitochondrial DNA (mtDNA) study of the biogeography of black spruce [*Picea mariana* (B.S.P.) Mill.], a North American conifer species with transcontinental distribution, indicated four genetically distinct glacial populations (Jaramillo-Correa *et al.* 2004). In other North American plant species studied at the range-wide or regional levels, glacial populations correlating in location with those observed in transcontinental black spruce have also been recently reported (Tremblay & Schoen 1999; Mitton *et al.* 2000a; Griffin & Barrett 2004). North American tree species with large distribution have only received limited attention and their phylogeographical study should help to shed light on past population fragmentation patterns and causes of vicariance at the continental level.

Jack pine (*Pinus banksiana* Lamb.) is a hard pine typical of the boreal forest in North America, with nearly transcontinental distribution. Pines are abundantly represented in the fossil record by their pollen, which they produce usually in large amounts (Critchfield 1985). According to palynological studies, hard pine species dominated the forests of southeastern United States at the Wisconsin glacial maximum, 18 000 years before present (BP) (Jackson *et al.* 2000; Williams *et al.* 2004). However, in most palynological studies (Davis 1983; Ritchie 1987; Jackson *et al.* 1997), it was not possible to distinguish well the pollen of jack pine from that of the partially sympatric red pine (*Pinus resinosa* Ait.) (Whitehead 1964), causing the lumping of these two species (e.g. Watts 1979; Jackson *et al.* 1997). In addition, the number of genetically distinct glacial populations could not be inferred from the sole study of the pollen or macrofossil records and in jack pine, no evidence for geographically distinct glacial populations has been observed from the palynological record (Ritchie 1987; Jackson *et al.* 2000; Williams *et al.* 2004).

In Pinaceae, mtDNA is maternally inherited and dispersed through seeds (Dong & Wagner 1993). This mode of dispersion should lead to reduced gene flow and higher population structure, compared to biparentally inherited

nuclear DNA or chloroplast DNA (cpDNA) markers which, in Pinaceae, are paternally inherited and dispersed through pollen and seed (Neale & Sederoff 1988). Recently, such a contrast has been repeatedly documented for a variety of conifers species (Wu *et al.* 1998; Gugerli *et al.* 2001; Sperisen *et al.* 2001; Burban & Petit 2003; Gamache *et al.* 2003; Petit *et al.* 2005). In a phylogeographical context, maternal inheritance implies that the imprint of ancient genetic structures established at the time of glacial periods should remain for longer periods of time with mtDNA markers than with cpDNA or nuclear DNA markers in Pinaceae. However, in conifers, only low levels of variation have been observed in mtDNA (Jaramillo-Correa *et al.* 2003), which has hampered the development of informative markers. In addition, the conifer mitochondrial genome appears to be very large (in excess of 750 kb in *Larix*, Kumar *et al.* 1995) and in angiosperms, it has a high rate of structural rearrangement (Palmer & Herbon 1988). Generally, the mitochondrial genome has also the lowest sequence mutation rate among the three genomes present in plants (Laroche *et al.* 1997). These features represent major obstacles to the design of markers of more variable intergenic regions that could be easily transferable between species. In addition, in jack pine, a previous study could not reveal any clear large-scale geographical structure using mtDNA restriction fragment length polymorphisms (Dong & Wagner 1993).

In the present study, we report on a new and highly variable marker from the mitochondrial genome of jack pine. We have further used this marker to assess the modern population structure of the species across its natural range and to infer the number and location of glacial populations and postglacial colonization routes and patterns. The wide distribution of jack pine should allow useful phylogeographic inferences at the continental level, in search of correlated patterns with other widely distributed plant and tree species in North America.

Materials and methods

DNA isolation and detection of mitochondrial polymorphism

A total of 543 mature jack pine trees from 35 seed sources distributed throughout the species range were sampled in three provenance tests located in Québec and Ontario. Seed sources were composed from 8 to up to 100 trees each. For each sampled tree, needles were collected, put on ice and then stored at -25°C until DNA extraction. Total DNA was extracted from 40 mg of ground needle tissue using the DNeasy Plant Mini Kit (QIAGEN) and following manufacturer's instructions. A preliminary test for detecting mtDNA polymorphism was conducted using an exploratory panel of 22 out of the 543 sampled trees. Each tree

Table 1 Target regions, annealing temperatures and expected size of PCR products for primer pairs used to amplify mtDNA regions

Genomic region	Annealing temperature (°C)	PCR product size (bp)	Primer source
<i>coxI</i> intron 1	58	750 (m.f.)†	Lu <i>et al.</i> 1998
<i>matR</i> intron 1	58	500	Jaramillo-Correa <i>et al.</i> 2003
<i>mh05</i>	54	NA	Jeandroz <i>et al.</i> 2002
<i>mh09</i>	52	300	Jeandroz <i>et al.</i> 2002
<i>mh33</i>	52	150 (m.f.)†	Jeandroz <i>et al.</i> 2002
<i>mh44</i>	56	not amplified	Jeandroz <i>et al.</i> 2002
<i>nad1</i> intron b/c	58	~3000	Demesure <i>et al.</i> 1995
<i>nad3-rps12</i> (i.r.)*	58	350	Soranzo <i>et al.</i> 1999
<i>nad3</i> intron 1	58	250	Soranzo <i>et al.</i> 1999
<i>nad5</i> intron 1	62	900	Jaramillo-Correa <i>et al.</i> 2003
<i>nad5</i> intron 4	61	900	Wu <i>et al.</i> 1998
<i>nad7</i> intron 1	57	900–2000	Jaramillo-Correa <i>et al.</i> 2004
<i>rpl5</i>	50	580	Duminil <i>et al.</i> 2002
SSU rRNA region V1	64	380	Duff & Nickrent 1999

*Intergenic region.

†Multiple fragments.

was selected from a different population and the sample covered most of the natural range. Fourteen primer pairs specific to plant mtDNA were tested to detect polymorphisms: *coxI* intron 1, *matR* intron 1, *mh05*, *mh09*, *mh33*, *mh44*, *nad1* intron b/c, *nad3-rps12* intergenic-region, *nad3* intron 1, *nad5* intron 1, *nad5* intron 4, *nad7* intron 1, *rpl5* and SSU rRNA region V1 mitochondrial regions (see Table 1). DNA was amplified in a PTC-225 thermal cycler (MJ Research) with 25–50 ng of DNA template, 0.2 µM of each primer, 0.2 mM of each dNTP, 1× of reaction buffer, 1.5 mM MgCl₂, and 0.65 unit of Platinum *Taq* DNA polymerase (Invitrogen). Polymerase chain reaction (PCR) conditions were as described in Jaramillo-Correa *et al.* (2004). To detect the presence of indels, PCR products were electrophoresed through 2% agarose gels with TAE. Monomorphic PCR products were further digested with different restriction enzymes (*CfoI*, *HaeIII*, *RsaI*, *Sau3AI*, and *TaqI*) and separated through 8% nondenaturing polyacrylamide gels (in TBE) in order to detect putative cleaved amplified polymorphic sites (CAPS). The undigested monomorphic PCR products were also denatured for 5 min at 94 °C and separated through 6% polyacrylamide gels at 10 mA and 10 °C for 18 h to detect possible single-strand conformation polymorphisms (PCR-SSCP).

Of the 14 primer pairs tested, polymorphism was observed for only one pair. Using the conditions and primers of Jaramillo-Correa *et al.* (2004), a minisatellite-like motif was detected for the intron 1 of *nad7* after migration on agarose gel. Given the large size of some of the variants obtained (up to 1500 bp) and the inconsistent presence of a second but weak amplification product of smaller fixed size, which was not observed in black spruce (Jaramillo-

Correa *et al.* 2004), and given the objective to sequence the various fragments, we designed two new internal primers for this locus: *nad7*(intron1)-internal-forward (5'-GAGGGACAACCCTGGAATACT-3') and *nad7*(intron1)-internal-reverse (5'-AAGGCCTCTCCATTTCCAAT-3'). PCR amplification was conducted with 25–50 ng of DNA template using 0.2 µM of each primer, 0.2 mM of each dNTP, 1× of reaction buffer, 2.4 mM of MgCl₂, and 0.65 unit of Platinum *Taq* polymerase (Invitrogen). The amplifications were performed in a PTC-225 thermal cycler (MJ Research) with the following program: a first step of 94 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 69 °C for 15 s, and 72 °C for 1 min, followed by a final extension step at 72 °C for 5 min. The use of these new primers and higher annealing temperature permitted to eliminate the inconsistent presence of the smaller and weak fragment of fixed size sometimes observed when using the original primers of Jaramillo-Correa *et al.* (2004).

The observed variants were verified at the DNA sequence level, by direct sequencing of the two DNA strands using an automated DNA sequencer (ABI 377, Applied Biosystems), the dideoxynucleotide chain-termination procedure, the appropriate primers and a Sequenase GC-rich kit (Applied Biosystems). The identity of the sequences obtained was also compared to the GenBank nonredundant nucleotide sequence database by using BLASTN (Altschul *et al.* 1990). Individuals from different populations bearing the same putative allele were also sequenced to detect possible homoplasy in fragment length, resulting in a total of 90 distinct DNA fragments sequenced on both strands.

In order to verify the presence of the minisatellite-like polymorphism in other pines, a screening was conducted for three hard pines, *Pinus contorta* Dougl. Ex Loud. var.

latifolia Engelm., *Pinus resinosa* Ait., *Pinus sylvestris* L., and for one soft pine, *Pinus strobus* L. About 25 individuals of various populations for each species were amplified using the first set of primers and PCR conditions above for *P. resinosa*, *P. sylvestris*, and *P. strobus*, and using the second set of primers and PCR conditions above for *P. contorta*. The PCR products were visualized on 2% agarose gels with TAE.

mtDNA diversity

Mitochondrial DNA diversity was estimated using two different parameters, the number of haplotypes per population (nh) and the total mtDNA diversity per population, H (equivalent to expected heterozygosity H_E or H_S for diploid data; Weir 1996). The index of among-population differentiation (G_{ST}) was also determined among the 35 seed sources sampled and following a hierarchical design where the total population was divided in a number of regions delineated by spatial analysis of genetic diversity (see below). R_{ST} , which is a measure of population differentiation that takes into account the similarity in haplotype size (Slatkin 1995), was also estimated. An R_{ST} significantly higher than G_{ST} would indicate that mutations following a stepwise-mutation model (SMM) contributed to population differentiation. The test was conducted with 1000 permutations using the program CPSSR, a modification of the program HAPSTEP (Pons & Petit 1996) (<http://www.pierroton.inra.fr/genetics/labo/Software/CpSSR/index.html>).

Analysis of geographical structure

Different numerical analyses were used to decipher geographical population structure and to identify possible zones of suture between glacial populations. The first two methods used, Bray–Curtis ordination (Bray & Curtis 1957) and cluster analysis using the unweighted pair-group method with arithmetic averages (UPGMA) (Sneath & Sokal 1973), are based on the analysis of a distance matrix without consideration of the geographical location of populations. Such distance methods would allow for the clustering of geographically distant populations with similar genetic structure. The other two methods used, spatial analysis of molecular variance (SAMOVA) (Dupanloup *et al.* 2002) and Monmonier analysis (Monmonier 1973), are spatially constrained methods based on connection networks between populations. Their aim is to identify most differentiated groups of populations that are as spatially clustered as possible.

For Bray–Curtis ordination and UPGMA, a matrix of pairwise relative Sorensen distances (McCune *et al.* 2002) between populations was estimated. This distance is a metric based on the sum, over the haplotypes, of the pairwise differences in haplotype frequencies between populations such as:

$$D_{(x_1, x_2)} = \frac{1}{2} \sum_{j=1}^p \left| \frac{y_{1j}}{\sum_{j=1}^p y_{1j}} - \frac{y_{2j}}{\sum_{j=1}^p y_{2j}} \right|$$

where $D_{(x_1, x_2)}$ is the relative Sorensen distance between populations x_1 and x_2 , p is the number of different haplotypes in the populations, j varying from 1 to p , y_{1j} is the number of individuals with haplotype j in population x_1 , and y_{2j} is the number of individuals with haplotype j in population x_2 . It is constrained between 0 when the haplotype frequencies are identical between two populations, and 1 when the populations have no haplotypes in common. Contrary to some other distances, it makes possible to estimate divergence even when populations are fixed for different haplotypes. Bray–Curtis ordination and UPGMA analyses were performed using the PC-ORD software version 4.25 (MjM Software Design 1999).

SAMOVA was conducted using the software SAMOVA version 1.0 (<http://web.unife.it/progetti/genetica/Isabelle/samova.html>). This method is based on a simulated annealing procedure which aims to maximize the proportion of total genetic variance due to differences between groups of populations (Dupanloup *et al.* 2002). A Delaunay triangulation (Delaunay 1934) was constructed from the geographical location of the 35 populations to ensure that the inferred groups were composed of adjacent populations. An arbitrary partition of the sampling in K groups was initially made (initial K -values were set between 2 and 12). An F_{CT} index (Wright 1978) associated to genetic differentiation among the initial K groups was then computed and an iterative simulated annealing process made it possible to obtain the optimal configuration of the groups as well as the final F_{CT} values. The simulated annealing process was repeated 10 000 times and it was conducted with a speed constant set to 0.9158. The new K groups were accepted with a probability of 1% if the new F_{CT} value was lower than that obtained in the previous iteration by 0.001. In order to make sure that the final configuration of the K groups was not affected by the initial configuration, the entire procedure was repeated 100 times with a randomly picked different initial configuration of groups, as suggested by Dupanloup *et al.* (2002). The configuration with the largest F_{CT} value among the 100 tested was retained as the best grouping of populations.

The Monmonier algorithm (Monmonier 1973) was also used to detect spatial structure in the mtDNA dataset. The method allows defining zones of maximum genetic change along a network of connecting populations, i.e. the 35 jack pine populations sampled. First, a Delaunay triangulation network (Delaunay 1934) was created to connect the adjacent populations. A set of Voronoi polygons was further superimposed on the corresponding Delaunay network. Pairwise relative Sorensen distances were then calculated between

populations that were connected by direct edges in the Delaunay network. The last step was to apply the maximum-difference algorithm of Monmonier to identify zones of discontinuity. A barrier was initiated by tracing a line on the edge of a Voronoi polygon that was associated with the highest distance between two neighbouring populations. The barrier was then extended progressively to the adjacent edge that was associated to the highest distance until it reached the edge of the network, or until it closed a circle around one or more populations. This analysis was performed with the software BARRIER version 2.2 (Manni *et al.* 2004).

Results

mtDNA variation

Of the 14 regions of the mitochondrial genome tested, only the intron 1 of *nad7* showed length polymorphism in jack

pine. Five fragment length variants were detected with the exploratory panel of 22 trees. Nine more length variants were found when the DNA from the 521 remaining trees was surveyed, for a total of 14 haplotypes. Their length was found to vary between 323 and 803 bp. Sequencing of each length variant observed in each population resulted in the sequencing of 90 distinct DNA fragments, for an average of 6.4 sequences per length variant. The fragment length polymorphism was found to be caused by variation in the number of copies of one perfect tandem repeat of 32 nucleotides (5'-TCTCCTTATCAGTCGAGCCTGCTTGCGCATCC-3'). No polymorphism was found in the sequence motif of the repeat among the 90 haplotypes sequenced. The number of copies of the tandem repeat varied from 3 to 18. None of the 543 sampled individuals harboured the 4x and the 15x repeat variants (Fig. 1A). The sequences for the first intron of *nad7* are available on GenBank (Accession numbers DQ060039–DQ060052). For the other four pine species

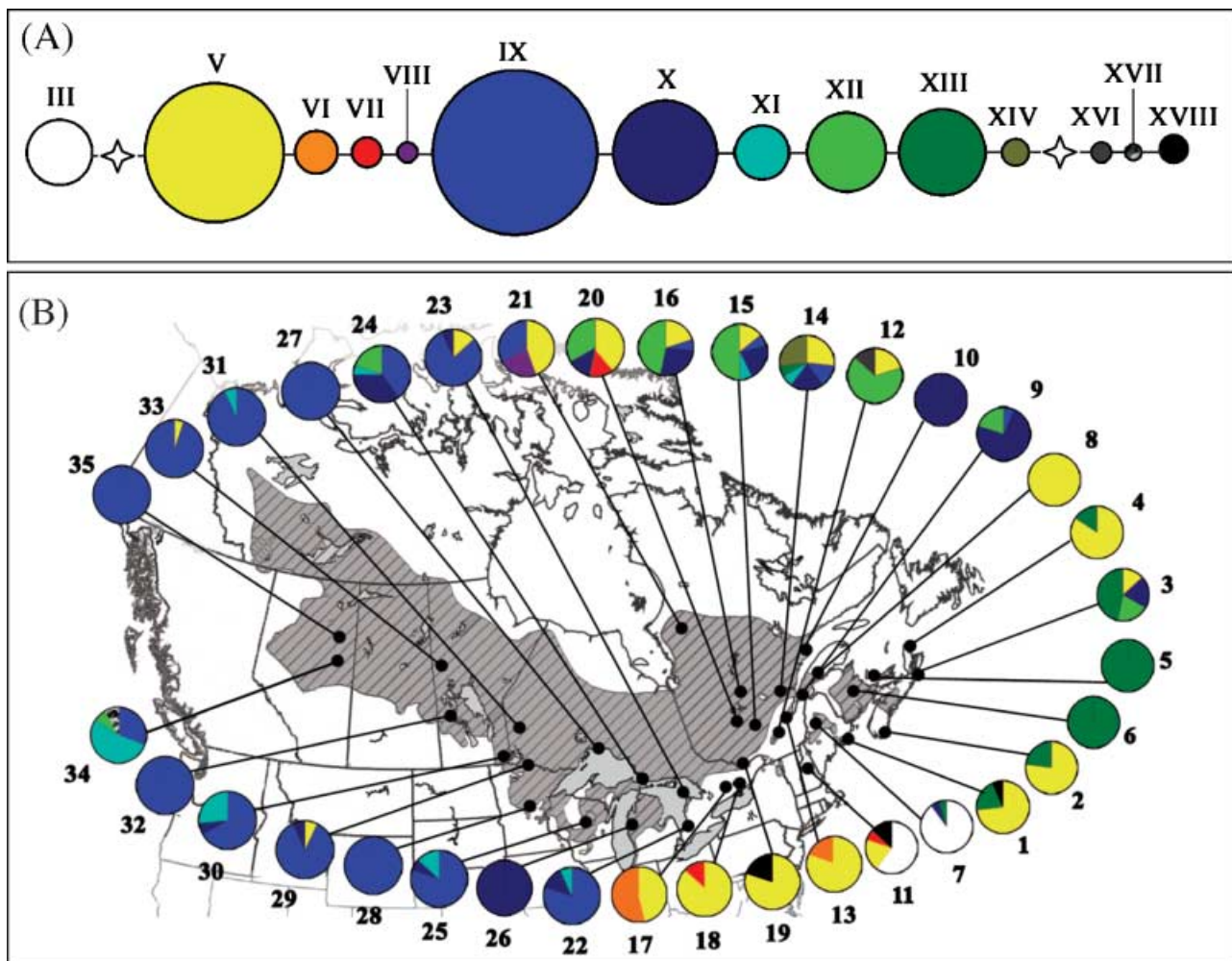


Fig. 1 (A) Haplotype network of the 14 mtDNA haplotypes found in jack pine. The size of circle is proportional to the relative frequency of haplotype in the study (for exact frequencies, see Table 2). Stars represent intermediate haplotypes not found in the sample. (B) Geographical distribution of mtDNA haplotypes in jack pine populations. The colours correspond to the haplotypes in Fig. 1A.

Table 2 Haplotype frequencies and genetic diversity estimates in 35 populations of jack pine

#	Population name	State or province*	Latitude (°N)	Longitude (°W)	Sample size	Haplotype counts														nht	H‡
						III	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XVI	XVII	XVIII		
1	Schoodic Head	ME	44.35	68.05	15	0	11	0	0	0	0	0	0	0	3	0	0	0	1	3	0.418
2	Birchtown Brook	NS	43.82	65.38	13	0	10	0	0	0	0	0	0	0	3	0	0	0	0	2	0.355
3	Durell Island	NS	45.33	61.00	15	0	2	0	0	0	0	3	0	3	7	0	0	0	0	4	0.684
4	Cape Breton H. P./Neils H.	NS	46.80	60.33	19	0	16	0	0	0	0	0	0	0	3	0	0	0	0	2	0.266
5	East Bideford	PEI	46.62	63.88	15	0	0	0	0	0	0	0	0	0	15	0	0	0	0	1	0
6	Cains River	NB	46.50	65.88	15	0	0	0	0	0	0	0	0	0	15	0	0	0	0	1	0
7	Lobster Lake	ME	45.85	69.55	20	18	0	0	0	0	0	1	0	0	1	0	0	0	0	3	0.185
8	Patapedia Depot	QC	48.12	67.50	16	0	16	0	0	0	0	0	0	0	0	0	0	0	0	1	0
9	Whiteworth/St-Alexandre	QC	47.71	69.49	15	0	0	0	0	0	1	11	0	3	0	0	0	0	0	3	0.418
10	Little Calumet River	QC	49.67	67.25	15	0	0	0	0	0	0	15	0	0	0	0	0	0	0	1	0
11	Welch Mt.	NH	43.92	71.58	15	9	3	0	1	0	0	0	0	0	0	0	0	0	2	4	0.578
12	St-Louis de France	QC	46.42	72.58	15	0	3	0	0	0	0	0	0	10	0	0	2	0	0	3	0.498
13	Château d'eau	QC	46.85	71.42	15	0	12	3	0	0	0	0	0	0	0	0	0	0	0	2	0.320
14	Port Alfred	QC	48.25	70.88	15	0	4	0	0	0	2	3	1	0	1	4	0	0	0	6	0.791
15	Lake Valade	QC	47.30	73.87	14	0	2	0	0	0	1	3	1	7	0	0	0	0	0	5	0.673
16	Ducharme River	QC	49.37	73.97	15	0	3	0	0	0	1	4	0	7	0	0	0	0	0	4	0.667
17	Twin Lakes	ON	44.65	77.88	15	0	7	8	0	0	0	0	0	0	0	0	0	0	0	2	0.498
18	Kaladar	ON	44.65	77.13	15	0	13	0	2	0	0	0	0	0	0	0	0	0	0	2	0.231
19	Constance Bay	ON	45.52	76.10	15	0	12	0	0	0	0	0	0	0	0	0	0	0	3	2	0.320
20	McKinnon Lake	QC	47.97	75.37	15	0	6	0	2	0	0	2	0	5	0	0	0	0	0	4	0.693
21	Kanaaupscow Lake	QC	54.03	76.52	9	0	4	0	0	2	3	0	0	0	0	0	0	0	0	3	0.642
22	Kettle Point	ON	43.22	81.95	15	0	0	0	0	0	12	2	1	0	0	0	0	0	0	3	0.338
23	Miller Lake	ON	45.13	81.45	15	0	2	0	0	0	12	1	0	0	0	0	0	0	0	3	0.338
24	Dunbar Forest	ON	46.43	84.27	20	0	0	0	0	0	8	7	1	4	0	0	0	0	0	4	0.675
25	Mosinee	WI	44.83	89.67	15	0	0	0	0	0	12	1	2	0	0	0	0	0	0	3	0.338
26	Freesoil	MI	44.13	86.12	15	0	0	0	0	0	0	15	0	0	0	0	0	0	0	1	0
27	Terrace Bay	ON	48.78	87.37	20	0	0	0	0	0	20	0	0	0	0	0	0	0	0	1	0
28	Brainerd	MN	46.33	94.17	15	0	0	0	0	0	15	0	0	0	0	0	0	0	0	1	0
29	Fort Frances	ON	48.77	93.50	14	0	1	0	0	0	12	1	0	0	0	0	0	0	0	3	0.255
30	Hadashville	MB	49.50	95.75	15	0	0	0	0	0	10	1	4	0	0	0	0	0	0	3	0.480
31	Red Lake	ON	51.02	94.12	15	0	0	0	0	0	14	0	1	0	0	0	0	0	0	2	0.124
32	Cowan	MB	52.05	100.38	15	0	0	0	0	0	15	0	0	0	0	0	0	0	0	1	0
33	Kississing Lake	MB	55.12	101.15	20	0	1	0	0	0	19	0	0	0	0	0	0	0	0	2	0.095
34	Lac la Biche	AB	55.23	111.92	13	0	0	0	0	0	4	0	7	1	0	0	0	1	0	4	0.604
35	Fort McMurray	AB	56.63	111.88	20	0	0	0	0	0	20	0	0	0	0	0	0	0	0	1	0
Total			—	—	543	27	128	11	5	2	181	70	18	40	48	4	2	1	6	2.6§	0.328§

*States/Provinces abbreviations: AB, Alberta (Canada); MB, Manitoba (Canada); ME, Maine (USA); MN, Minnesota (USA); MI, Michigan (USA); NB, New Brunswick (Canada); NH, New Hampshire (USA); NS, Nova Scotia (Canada); ON, Ontario (Canada); PEI, Prince Edward Island (Canada); QC, Quebec (Canada); WI, Wisconsin (USA).

†Number of distinct haplotypes per population.

‡Haplotype diversity.

§Average.

tested (*P. contorta* var. *latifolia*, *P. resinosa*, *P. sylvestris*, and *P. strobus*), positive amplification was obtained but fragment length polymorphism was only detected for *P. contorta* var. *latifolia*, a species closely related to *P. banksiana* (Dancik & Yeh 1983; Wagner *et al.* 1987) (data not shown).

Haplotype diversity and distribution

The number of distinct haplotypes per population (nh) ranged from 1 to 6 with an average value of 2.6 (Table 2). The average diversity index (H) per population ranged from 0 to 0.791, with an overall average value of 0.328 (Table 2). The geographical distribution of haplotypes was clearly not random in jack pine (Fig. 1B). Population differentiation was quite high among the 35 populations, with G_{ST} and R_{ST} values of 0.569 and 0.472, respectively. When excluding presumed zones of secondary contact between ancestral populations (populations in central Québec and from the easternmost part of Maine and Nova Scotia, see below), the G_{ST} and R_{ST} values increased to 0.669 and 0.618, respectively, among the 23 remaining populations.

Haplotypes V, IX, X, XII, and XIII (with roman numbers corresponding to the number of minisatellite repeat motifs) were the most abundant (Fig. 1A). Altogether, 86% of the sampled individuals had one of these haplotypes. Variation among populations and regions was evident for these haplotypes (Fig. 1B). The 'medium size' haplotypes (IX, X, and XI repeat motifs) were mainly present in the western part of jack pine natural range (i.e. Alberta to Ontario and the adjacent United States) with some occurrence in central Québec. Comparatively, in the eastern part of the species' range (i.e. eastern Ontario, Québec, the Maritime provinces in Canada, and Maine and New Hampshire in the United States), the distribution of haplotypes was more heterogeneous. Haplotype XIII was fixed in populations of New Brunswick and Prince Edward Island and it was also present in each of the easternmost populations (#1 to #4, Fig. 1B). Haplotype XII, a likely close relative of haplotype XIII, was mainly located in central Québec. The short haplotype V was widespread in the eastern region, with a major concentration in southern Québec and eastern Ontario where it was also found with other short and likely closely related haplotypes (III, VI, and VII). Moreover, haplotype diversity was higher in the eastern than in the western part of the continent: all haplotypes but one (haplotype XVIII) were observed in eastern populations (#1 to #21, average number of haplotypes per population = 2.8) in contrast to only six distinct haplotypes observed in western populations (#22 to #35, average number of haplotypes per population = 2.3).

The rare haplotypes VIII, XIV, XVI, and XVII were each found in only one population, i.e. populations #21, #14, #34, and #12, respectively. Other rare haplotypes (III, VI, VII, and

XVIII) were clustered in two or three more or less adjacent populations.

Distance analyses

The Bray–Curtis ordination obtained was consistent with the geographical distribution of the haplotypes described above (Fig. 2A). All western populations, except population #26, were clustered together in the lower-left corner of the graph (group W, Fig. 2A) because of the similarity of these populations regarding their haplotype composition (dominance of haplotypes IX, X, and XI). Another cluster of populations, at the other end of the first ordination axis (X-axis), was located on the right side of the scatter plot in the reduced space. Most of these populations were mainly composed of individuals carrying haplotype V and associated rare haplotypes III, VI, and VIII (group S, Fig. 2A). The third distinct group of populations on the graph was made of populations #3, #5, and #6, with the last two populations being fixed for haplotype XIII (group E, Fig. 2A). They were distinct from other populations relative to the third ordination axis (Z-axis). Populations with the largest number of haplotypes were scattered in the middle of the graph between the groups of populations previously described (group C, Fig. 2A). Populations #9, #10, and #26 stood apart on the second ordination axis (Y-axis), due to the high prevalence of haplotype X (group A, Fig. 2A).

UPGMA cluster analysis made it possible to distinguish six main clusters if we arbitrarily retain groups with average population divergence smaller than 50% (Fig. 2B). These clusters were consistent with those delineated by the Bray–Curtis ordination. UPGMA separated eastern populations #5 and #6 from the rest (cluster E, Fig. 2B). A second group was composed of all western populations (#22 to #35) (cluster W, Fig. 2B) except populations #26 and #34. Populations #9, #10, and #26 formed the third cluster (cluster A, Fig. 2B). Another cluster contained populations bearing the haplotype V with a frequency of 75% or more (cluster S, Fig. 2B). Another cluster included populations from central Québec carrying haplotype V at low frequencies, together with the co-occurrence of other haplotypes (generally haplotypes IX and XII) (cluster C, Fig. 2B). A final group was made of populations #7 and #11, which presented a high prevalence of haplotype III (cluster B, Fig. 2B).

Spatial analyses

For SAMOVA, we repeated the analysis by increasing the value of K (2 to 12) until the F_{CT} value reached a plateau at $K = 8$. At this level, some groups were made of only one population, which implied that the geographical structure disappeared (Heuertz *et al.* 2004). Indeed, for all values of $K \geq 6$, some groups were made of only one population.

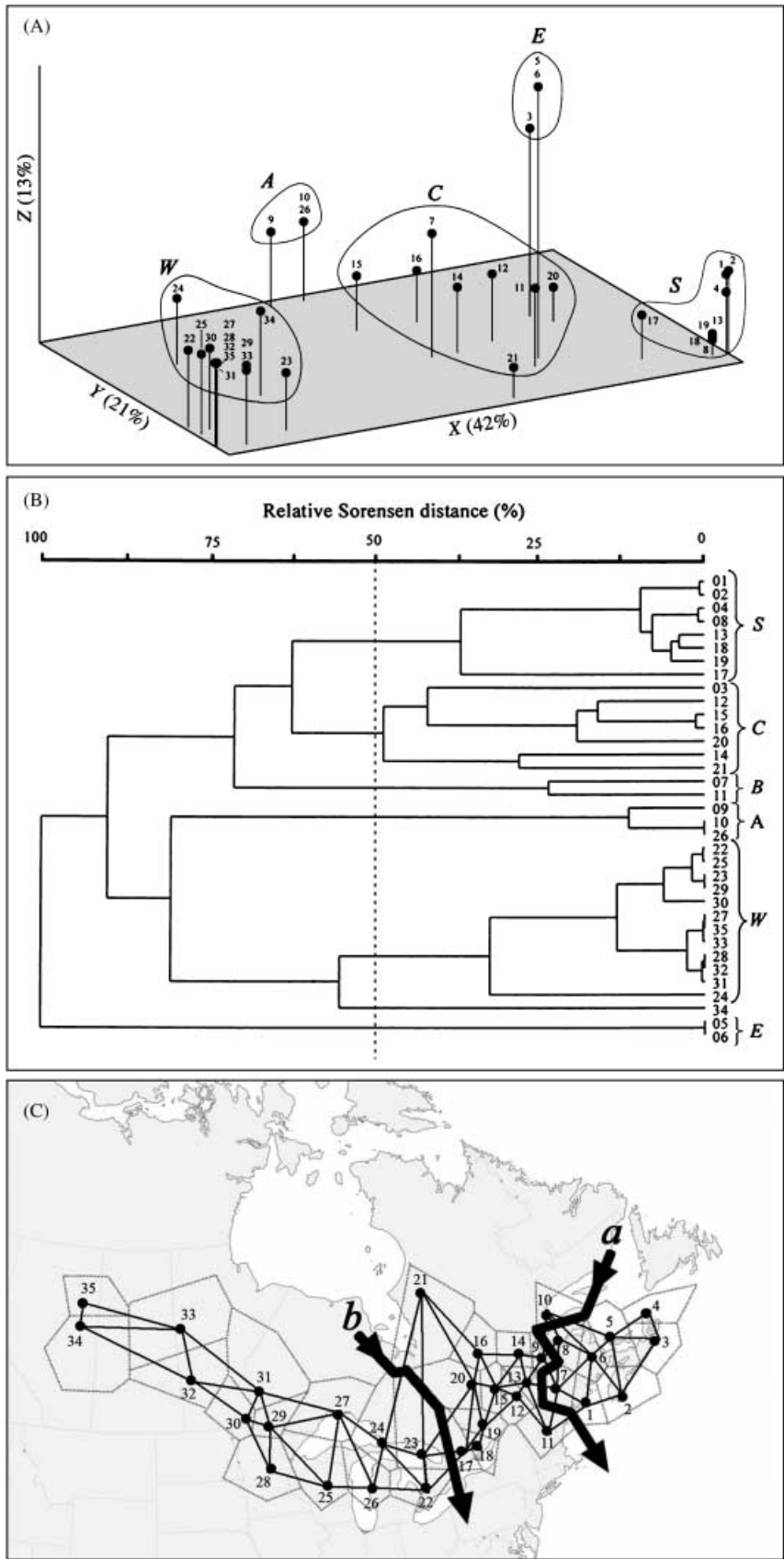


Fig. 2 (A) Bray–Curtis ordination of jack pine populations based on pairwise relative Sorensen distances estimated from mtDNA haplotype frequencies. The X, Y, and Z axes explained 42%, 21%, and 13% of the total variation, respectively. (B) UPGMA dendrogram of jack pine populations based on pairwise relative Sorensen distances. Groups were arbitrarily delineated with average population divergence smaller than 50%. (C) Genetic boundaries (bold lines *a* and *b*) obtained with Monmonier's maximum difference algorithm. The solid lines represent the Delaunay triangulation and the dashed lines represent the Voronoi tessellation.

Table 3 Group structure and genetic diversity indices estimated by SAMOVA

Group	Populations	H^*	F_{CT}^\dagger
1	22, 23, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35	0.214	0.500
2	1, 2, 4, 8, 17, 18, 19, 21	0.341	
3	9, 10, 12, 13, 14, 15, 16, 20, 24, 26	0.474	
4	7, 11	0.381	
5	3, 5, 6	0.228	

*Average haplotype diversity per population in each group.

†Differentiation among the five groups.

Thus, the grouping pattern corresponding to $K = 5$ was considered as valid and retained for interpretation (Table 3). As for UPGMA and Bray–Curtis ordination, four of these five clusters corresponded to the western, south-eastern, central Québec, and easternmost groups of populations, plus a fifth group containing populations #7 and #11 characterized by the presence of haplotype III. A maximum F_{CT} value of 0.500 was estimated between these five groups delineated by SAMOVA.

Monmonier's maximum difference algorithm identified two main genetic boundaries (Fig. 2C) that were previously delineated by other analyses. By definition, boundaries correspond to zones of most abrupt genetic change in space and are identified in a descending order. The first boundary (*a* on Fig. 2C) separated the eight easternmost populations of the Maritime provinces of Canada and Maine from the 27 others located west of this first group. All populations (except population #7) located eastward of this boundary harboured individuals with haplotype XIII. The second boundary (*b* on Fig. 2C) separated the 27 remaining populations in a central group (populations #9 to #21) and a western group (populations #22 to #35). The average diversity indices (H) estimated for each of the three main groups delineated by the Monmonier spatial analysis were 0.239, 0.487, and 0.232, respectively, from east to west. Thus, the large central group was the most heterogeneous.

Discussion

Mitochondrial DNA variation and mutational model

In this study, we detected a minisatellite-like motif in the intron 1 of *nad7* after screening for length polymorphism in 14 introns and intergenic regions of the mitochondrial genome of jack pine. No other length polymorphisms were observed and the sequencing of 90 individuals did not show any point mutation in the minisatellite repeat. Other minisatellite-like variation of mtDNA have been reported in the intron b/c of the gene *nad1* in *Picea abies* (Gugerli *et al.* 2001; Sperisen *et al.* 2001) and in three species of *Pinus*

(Mitton *et al.* 2000b), and for the locus *mh44* in *Picea abies* (Bastien *et al.* 2003). Between 10 and 18 size variants were observed for these two loci in *Pi. abies*, and they allowed for the characterization of the geographical structure of Norway spruce. However, the mutational mechanisms responsible for minisatellite-like patterns have not been clearly established yet (Lunt *et al.* 1998). Previous plant studies using this type of markers assumed an SMM (Bastien *et al.* 2003; Cozzolino *et al.* 2003).

However, it has also been shown that the model of mutation for minisatellites may deviate in the direction of an infinite allele model (IAM), especially for repeats of size observed in the present study (Shriver *et al.* 1993). In our study, haplotypes with similar number of repeats tended to cluster in areas not affected by Holocene secondary contact between ancestral populations, for instance the clustering of haplotypes IX, X, and XI in western and central Canada, the clustering of haplotypes XII and XIII in eastern Canada, or the clustering of haplotypes III, V, VI, and VIII in the southeast part of the natural range. Such spatial clustering of haplotypes of similar size would lend support to the SMM model. However, R_{ST} , which takes into account size differences between haplotypes, was not higher than G_{ST} , even when Holocene contact zones between presumed ancestral populations were excluded from the analysis. Such a result goes against SMM. Outside these suture zones, there were obvious cases where minisatellite-like variation did not seem to follow SMM. For instance, within the southeast part of the range, the occurrence of the large size haplotype XVIII with small size haplotypes III and V in populations #11 and #19 suggests that the mechanism that is generally presumed to lead to minisatellite-like patterns, slipped-strand mispairing, may also produce multistep mutations (Shriver *et al.* 1993). Alternately, the presence of these widely different haplotypes in the same populations could also result from ancient mixing between different ancestral populations well before the Holocene. Altogether, these observations suggest that when evidence for putative mixing between ancestral populations is taken into account, SMM might hold reasonably well for our data. Indeed, when excluding altogether the zones of Holocene contact between ancestral populations and populations harbouring haplotype XVIII, a significant difference ($P < 0.05$) was observed between G_{ST} and R_{ST} with respective values of 0.685 and 0.861 among the 21 remaining populations analysed.

Ancestral vs. more recent haplotypes

The frequencies of the 14 *nad7* intron 1 size variants were highly structured at the geographical level, resulting in useful inferences about the glacial and postglacial history of jack pine. Because the linear haplotype network did not allow for the formal recognition of ancient haplotypes from

Table 4 Genetic diversity indices for groups of populations delineated from interpretation of the results of Bray–Curtis ordination, UPGMA clustering, Monmonier, and SAMOVA spatial analyses

Region	Populations	N^*	nht	$A‡$	$H§$	$G_{ST}¶$
West	22–35	227	6	2.3	0.231	0.419
Southeast	8, 11, 13, 17, 18, 19	91	5	2.2	0.324	0.345
Easternmost	1–7	112	6	2.3	0.273	0.597
Central Québec	9, 10, 12, 14, 15, 16, 20, 21	113	10	3.6	0.548	0.265
Total		543	14	2.6**	0.328**	0.569

*Sample size.

†Total number of haplotypes observed in each region.

‡Average number of haplotypes per population in each region.

§Average H over the populations contained in each region.

¶Differentiation among populations in each region.

**Average.

their more recent derivatives, only tentative interpretations of the relative age of haplotypes remained possible based on their abundance (Crandall & Templeton 1993). For instance, the frequent haplotype IX [frequency (f) = 0.33] might be ancestral to the less frequent haplotypes X and XI, which were often found with haplotype IX and which could represent more recent derivatives. The same rationale could apply to the rare haplotypes VI and VII, which were geographically strongly associated with the frequent and likely more ancient haplotype V (f = 0.24). Similarly, the rare haplotypes VIII, XIV and XVI had presumably mutated recently from the more abundant and widely dispersed haplotypes IX, XII, or XIII.

Geographical structure and gene flow

The overall differentiation among populations of jack pine estimated from the mtDNA minisatellite was rather high (G_{ST} = 0.569), especially when excluding putative zones of suture (see below) between ancestral glacial populations (G_{ST} = 0.669). Such level of population differentiation are comparable, albeit in the lower average of that reported for conifer and angiosperm maternally transmitted markers (Petit *et al.* 2005). For example, a range-wide investigation on the mtDNA population structure of black spruce, a North American transcontinental conifer species sympatric to *Pinus banksiana* on most of its range, resulted in a G_{ST} of 0.671, with strong geographical structure suggestive of several disjunct glacial populations (Jaramillo-Correa *et al.* 2004). In contrast, limited population differentiation was previously observed at the regional scale in jack pine using nuclear or cpDNA markers. For instance, G_{ST} or F_{ST} values derived from allozyme markers ranged between 0.021 and 0.070 among populations from Alberta (Dancik & Yeh 1983), from Manitoba (Ross & Hawkins 1986), from Ontario (Danzmann & Buchert 1983), from Québec (Gauthier *et al.* 1992), from the states of New Hampshire and New York

(Misenti & DeHayes 1989) or from Wisconsin (Sáenz-Romero *et al.* 2001). In our study, if we were to estimate G_{ST} among populations within regional groups delineated by all our analyses (Table 4), we would obtain values in the range of 0.265 to 0.597, which is much higher than preceding values derived from allozyme markers.

The relatively high value of G_{ST} estimated in this study indicates that gene flow by seed is much restricted in jack pine. Gene flow was estimated at Nm = 0.379 migrant per generation, using the overall G_{ST} estimate as an indirect measurement of gene flow (Takahata & Palumbi 1985). This level of gene flow is below that necessary to counteract the effects of genetic drift and it indicates that gene flow is much restricted at the level of mtDNA in jack pine, as previously noted for other conifers (Latta *et al.* 1998; Gamache *et al.* 2003). Despite such low levels of gene flow, there was some evidence for long-distance seed dispersal in the present study. Haplotypes V and XII were most abundant in the southeastern parts of Ontario and Québec as well as in central Québec, respectively. Their rare presence in some populations from Ontario (populations #24 and #29) and northern Manitoba (population #33) (Fig. 1B), where no close-relative haplotypes were noted, could be the result of long-distance dispersal from eastern populations where they were abundant. It has been argued that such long-distance migration could have played a major role in the colonization of several tree species (Petit *et al.* 1997; Clark *et al.* 1998; Cain *et al.* 2000) but in our case, these haplotypes did not spread. The rare occurrence of haplotypes V and XII in central and western Canada could also reflect their ancient presence in more than one genetically distinct glacial population. Alternatively, it is also possible that the rare occurrence of haplotype XII in population #24 is indicative of parallel evolution resulting from a mutational event from haplotype XI, also present in the population.

Spatial analyses also indicated a strong geographical structure, which is most likely related to ancestral glacial

populations. Monmonier analysis regrouped populations in three main clusters which corresponded to the western, central-eastern, and eastern regions. These regions were each characterized by one of the few most abundant haplotypes (IX/X, III/V/VI, and XII/XIII, respectively). In addition, the Bray–Curtis ordination, UPGMA, and SAMOVA revealed a fourth group of heterogeneous populations in the large region of central Québec (Table 4). This group was characterized by higher genetic diversity at the intra-population level ($H = 0.548$) and lower population differentiation ($G_{ST} = 0.265$), as compared to the other three regions (Table 4). The preceding considerations indicate that, most likely, three distinct glacial populations existed in jack pine, each dominated by a major haplotype and its length derivatives, one western population with haplotypes IX/X, a second population in the central-east with haplotypes III/V/VI, and a third population in the east with haplotypes XII/XIII. Merging fronts between these three putative distinct glacial populations are apparent in central Québec. In Nova Scotia and in the easternmost part of Maine, there is also evidence that the group dominated by haplotype V and that dominated by haplotypes XII/XIII had merged to some extent.

The reconstitution of postglacial history from fossils and molecular evidence

Fossil data indicate that the northern hard pines *P. banksiana* and *P. resinosa* have been abundantly distributed from the Atlantic coastal and interior highland regions during the last glacial maximum (LGM) (Rudolph & Yeatman 1982). This area corresponds approximately to the states of Tennessee, South and North Carolina, Alabama, Georgia, and northern Mississippi (Jackson *et al.* 2000). The presence of jack pine north of 34° latitude at the LGM is corroborated by macrofossils (needles) found in eastern Tennessee (Delcourt 1979) and northwestern Georgia (Watts 1970). Some authors have put forward that northern hard pines could have survived the glaciation as south as Florida (Jackson *et al.* 2000). The lower concentration of *Pinus* pollen and the absence of macrofossil in the mid-continent regions suggests that these species were absent or scattered in these regions (Jackson *et al.* 2000). Biomes reconstitution from fossils of the LGM suggested the presence of a conifer forest along the Atlantic coast and the existence of a steppe in northern-central United States (Williams *et al.* 2000).

Simulation maps of the postglacial recolonization of the modern range of several eastern North American trees have been constructed from the fossil record (Davis 1983; Webb 1988; Jackson *et al.* 1997). They indicate that northern hard pines had spread rapidly northwards and westwards from 14 000 to 10 000 BP. By 12 000 BP, the abundance of *Pinus* pollen in palynological records had declined in the

midcontinent, but remained high along most of the Atlantic coast (Jackson *et al.* 1997). By 10 000 BP, northern hard pines were a significant component of the vegetation from the Atlantic coast of Canada to west of the Great Lakes region (Webb 1988).

Despite that the fossil record revealed no clear evidence of fragmented glacial populations (Ritchie 1987; Jackson *et al.* 2000; Williams *et al.* 2004), the various qualitative and quantitative analyses conducted herein showed that the current genetic structure of jack pine natural range would result from colonization by three genetically distinct and presumably disjunct glacial populations. All numerical analyses have put populations from the west in a same group, those harbouring haplotypes IX/X (e.g. populations #25 to #33). This pattern supports the hypothesis that they could have all emerged from a sole genetically distinct western glacial population. The fossil data of the western part of United States do not suggest the possibility of a distinct refuge for jack pine as west as on the western coast of North America where the area is rather occupied by the closely related *P. contorta* (Critchfield 1985). Moreover, palynological studies in western Canada (Saskatchewan, Alberta, and Northwest Territories) showed a clear progression of jack pine from central Canada to the west after the retreat of the ice sheet (McLeod & MacDonald 1997), which suggests that a genetically distinct glacial population might have existed in the Midwest or at least, west of the Appalachian Mountains. The great glacial lake Agassiz had probably constituted a physical barrier to the northward progression of animal and plant taxa until 10 000 BP (Dyke & Prest 1987). Thus, according to pollen data, jack pine would have colonized Manitoba 10 000 BP (Jacobson *et al.* 1987) from that genetically distinct midwestern glacial population and later moved to northern Alberta around 8000 BP (McLeod & MacDonald 1997).

It should also be noted that population #34 has a genetic composition dominated by haplotype XI and the presence of haplotypes XII and XVII, in sharp contrast to other nearly monomorphic jack pine populations from Alberta and western Manitoba. These haplotypes could result from mutational events or long-distance seed dispersal from the east. However, some of these haplotypes (XI) and other ones of similar size (XIII and XVI) were also observed at a relatively high frequency in the preliminary screening of five populations of lodgepole pine in western Alberta (data not shown, population samples obtained from R. Beaudoin, MRNF, Québec Government). Such a presence of these haplotypes in lodgepole pine suggests that they might be ancestral and that mtDNA introgression might have taken place between jack pine and lodgepole pine, which are known to hybridize naturally (Dancik & Yeh 1983; Wagner *et al.* 1987). These observations suggest that the presence of haplotypes XI, XII, and XVII in population #34 of jack pine might result from seed dissemination from

P. contorta, followed by backcrossing into the nuclear background of *P. banksiana*.

The geographical structure of haplotype diversity in the southeastern range of jack pine suggests the colonization by two genetically distinct and presumably disjunct glacial populations, followed by the establishment of a zone of secondary contact in Nova Scotia and another larger one in central Québec (Fig. 3). These ancestral populations would have been located in the east and in the southeast of North America: a southeastern population likely located east of the Appalachian Mountains, which would have been characterized by a high abundance of haplotype V and the presence of haplotypes III and VI, as observed today in populations #8, #11, #13, #17, #18, and #19, and an eastern coastal refuge, which would have been dominated by haplotype XIII, as seen in populations #5 and #6.

The hypothesis of a distinct eastern coastal refuge has also been suggested for *P. resinosa*, another northern hard pine (Walter & Epperson 2001, 2005), and also for black spruce (Jaramillo-Correa *et al.* 2004). This refuge could have been located on the emerged coast on the continental shelf along Nova Scotia and Magdalen Island in the Gulf of St Lawrence (Dyke *et al.* 2002). These areas were exposed during the last glaciation because of the lower sea level at LGM, which was approximated at -120 m (Piper *et al.* 1986).

According to this hypothesis, the glacial population from the eastern coastal refuge would have initially expanded inland from the coastal line, and establishing itself in Prince Edward Island, New Brunswick, and Nova Scotia (Fig. 3). It would have been followed by a second wave of migrants bearing haplotypes V and related haplotypes from the more southern glacial population presumably located east of the Appalachian Mountains. The expansion of this southeastern population in southern Québec and Ontario through the Adirondack Mountains, and in Nova Scotia following the coastal line, would have been favoured by higher summer insolation between 10 000 and 7500 BP (Guiles Ellis *et al.* 2004), which would have increased the frequency of fires (Carcaillet & Richard 2000) and enhance the release of jack pine seeds. In Nova Scotia and in the easternmost part of Maine (populations #1 to #4), a zone of secondary contact would have occurred between these populations and the eastern coastal populations dominated by haplotype XIII.

A less parsimonious scenario, not requiring the existence of a genetically distinct glacial population in the east, would imply the colonization of southern and central Québec through a northward-advancing migration front carrying simultaneously the major haplotypes V and XII from the United States, together with a long-distance colonization event from the southeast and ensuing expansion leading to the presence of haplotype XIII in the Maritime provinces of Canada. This scenario appears less likely,

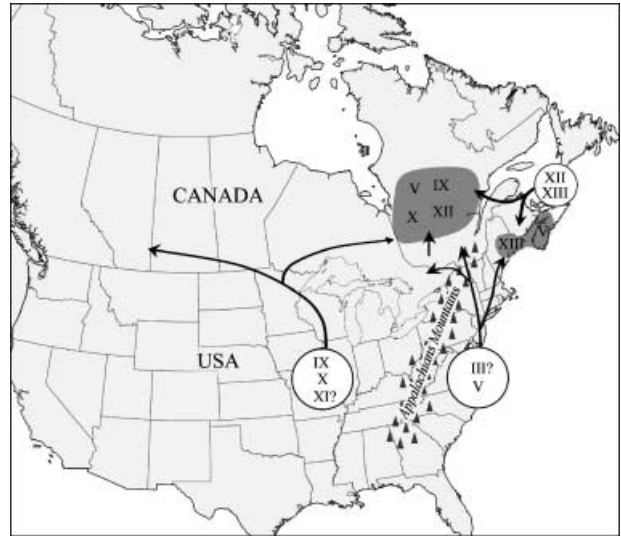


Fig. 3 Summary of the major inferred phylogeographical processes that led to the current distribution of mtDNA diversity of jack pine across its natural range. Glacial populations (circles) and postglacial recolonization routes were inferred from fossils, palynological, and genetic data. The grey colour in central Québec and Nova Scotia indicates the zones of secondary contact between ancestral populations.

given that neither trace of haplotype XII nor haplotype XIII was found in populations from southern Québec and Ontario (populations #11, #13, #17, #18, and #19).

A zone of secondary contact is also evident in central Québec where haplotype diversity was the highest and population differentiation was the lowest (Table 4). Because of the remnants of the ice mass over central and northern Québec up to 6000 BP (Matthews *et al.* 1989), central Québec was among the last Canadian regions colonized by jack pine. Two postglacial colonization fronts, one from the south with haplotype V dominating and one from the east with haplotype XII dominating, presumably entered the region early. The Champlain Sea, which was well in place by 11 000 and up to ~10 000 BP (Dyke & Prest 1987), could have restricted the progression of the trees carrying haplotype V in their northward advance. These two migration fronts would have been followed later by a third front of migration dominated by haplotypes IX/X from the west, after the drying of the Holocene inland sea south of James Bay (Lake Ojibway) at approximately 8500 BP (Dyke & Prest 1987; Matthews *et al.* 1989).

Similar areas of great diversity resulting from a suture zone between two migration fronts were also observed in European trees (Petit *et al.* 2003), such as for white oaks (Dumolin-Lapègue *et al.* 1997) and for other organisms (see Taberlet *et al.* 1998). Such a region of suture was also observed in northeastern Québec between three genetically

distinct migration fronts in black spruce (Jaramillo-Correa *et al.* 2004). Contrary to documented beliefs that colonization process may incur loss of genetic diversity (Hewitt 1996), the present data suggest that it could also act as a factor of increased genetic diversity in particular areas. These zones of increased genetic diversity should be taken into consideration in programs aimed at conserving genetic resources.

The current observation of genetically distinct glacial populations in jack pine parallels other observations from studies on quantitative traits. Variability was reported to increase in all directions from Lake States to maximum phenotypic variability on the east coast, in the southern Lake States, and at the western border of the species (Schoenike 1976). Abrupt boundaries were also observed in Michigan (Arend *et al.* 1961) and Ontario (Skeates 1979). These boundaries were interpreted as a consequence of different postglacial history. Consistently with our observations from mtDNA, all these studies have proposed a distinct lineage for the populations located west of the Great Lakes.

The presence of genetically distinct glacial populations east and west of the Appalachian Mountains and the Great Lakes has also been suggested for black spruce from the study of mtDNA variation (Jaramillo-Correa *et al.* 2004). These observations together with ours suggest a correlation between species in the vicariance factors favouring genetic isolation and drift. It could be proposed that the northern part of the Appalachian Mountains and Adirondacks in the United States could have constituted a long-lasting physical barrier and hampered seed gene flow between glacial populations located east and west of the mountain range.

Evidence is also accumulating for a correlation between species for an eastward glacial refuge along the coast or on now submerged lands, but the exact location for such a refuge seems variable from one species to the other. The possibility of a glacial population along the coast of northern Newfoundland or Labrador has been inferred for black spruce (Jaramillo-Correa *et al.* 2004). Such a glacial population would have been located much more north than that proposed for jack pine in the present study. Contrary to black spruce, jack pine never reached Newfoundland or Labrador during the present interglacial. Therefore, the presumed eastern glacial population of jack pine would have to be located on emerged lands presumably off the coast of Cape Breton Island of Nova Scotia or in the Gulf of St Lawrence near perhaps Magdalen Islands. Limited refugia are also thought to have existed along the continental shelf on the eastern and western coasts of Nova Scotia, which were occupied by arctic and boreal plants and animals during the ice age (Pielou 1992). Botanists and entomologists consider the existence of such refugia to be essential to explain the present-day distributions of arctic

and boreal species in Nova Scotia (Anderson 1985; Hamilton & Langor 1987).

Acknowledgements

We are especially grateful to Roger Beaudoin (Ministère des Ressources Naturelles et de la Faune du Québec), Steve D'Eon, Peter Arbour, Daniel Plourde, Jean-Paul Bilodeau, and Marie Deslauriers (Canadian Forest Service), as well as Denis Mimeault and Sébastien Giguère (CRBF, Université Laval and Canadian Forest Service) for their help with population sampling. We also thank Isabelle Giguère and Sauphie Senneville (CRBF, Université Laval) for laboratory assistance, as well as Rémy Petit (INRA, Bordeaux, France) and two anonymous reviewers for their valuable comments on a previous draft of this manuscript. This research was made possible by grants from Fonds Québécois de recherche sur la nature et les technologies, the National Sciences and Engineering Research Council of Canada, and the Canadian Forest Service to J. Bousquet and J. Beaulieu.

References

- Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Anderson TW (1985) Late-Quaternary pollen records from eastern Ontario, Quebec and Atlantic Canada. In: *Pollen Records of Late-Quaternary North American Sediments* (eds Bryant VM, Holloway RG), pp. 281–326. American Association of Stratigraphic Palynologist Foundation, Dallas, Texas.
- Arend JL, Smith NF, Spurr SH, Wright JW (1961) Jack pine geographic variation — 5-year results from Lower Michigan test. *Michigan Academy of Sciences, Arts, and Letters Papers*, **46**, 219–238.
- Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.
- Bastien D, Favre JM, Collignon AM, Sperisen C, Jeandroz S (2003) Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]. *Theoretical and Applied Genetics*, **107**, 574–580.
- Bray RJ, Curtis JI (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.
- Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology*, **12**, 1487–1495.
- Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.
- Carcaillet C, Richard PJH (2000) Holocene changes in seasonal precipitation highlighted by fire incidence in eastern Canada. *Climate Dynamics*, **16**, 549–559.
- Clark JS, Fastie C, Hurtt G *et al.* (1998) Reid's paradox of rapid plant migration — dispersal theory and interpretation of paleoecological records. *Bioscience*, **48**, 13–24.
- Cozzolino S, Cafasso D, Pellegrino G, Musacchio A, Widmer A (2003) Fine-scale phylogeographical analysis of Mediterranean *Anacamptis palustris* (Orchidaceae) populations based on chloroplast minisatellite and microsatellite variation. *Molecular Ecology*, **12**, 2783–2792.

- Crandall KA, Templeton AR (1993) Empirical tests of some predictions of the coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.
- Critchfield WB (1985) The late Quaternary history of lodgepole and jack pines. *Canadian Journal of Forest Research*, **15**, 749–772.
- Cruzan MB, Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution*, **15**, 491–496.
- Dancik BP, Yeh FC (1983) Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. *Canadian Journal of Genetics and Cytology*, **25**, 57–64.
- Danzmann RG, Buchert GP (1983) Isozyme variability in central Ontario jack pine. In: *Proceedings of the 28th Northeastern Forest Tree Improvement Conference* (ed. Eckert RT), pp. 232–248. Institute of Natural and Environmental Resources, University of New Hampshire, Durham, New Hampshire.
- Davis MB (1983) Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, **70**, 550–563.
- Delaunay B (1934) Sur la sphère vide. *Bulletin of the Academy of Sciences of the USSR*, **7**, 793–800.
- Delcourt HR (1979) Late Quaternary vegetation history of the eastern Highland Rim and adjacent Cumberland Plateau of Tennessee. *Ecological Monographs*, **49**, 225–280.
- Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129–131.
- Dong J, Wagner DB (1993) Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta*. *Theoretical and Applied Genetics*, **86**, 573–578.
- Duff RJ, Nickrent DL (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany*, **86**, 372–386.
- Duminil J, Pemonge M-H, Petit RJ (2002) A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. *Molecular Ecology Notes*, **2**, 428–430.
- Dumolin-Lapègue S, Demesure B, Fineschi S, Le Come V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.
- Dyke AS, Andrews JT, Clark PU *et al.* (2002) The Laurentide and Innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews*, **21**, 9–31.
- Dyke AS, Prest VK (1987) *Paleogeography of northern North America, 18000–5000 years ago*. Geological Survey of Canada, Ottawa, Canada. Map 1703A, scale, 1: 12 500 000.
- Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003) Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. *Molecular Ecology*, **12**, 891–901.
- Gauthier S, Simon JP, Bergeron Y (1992) Genetic structure and variability in jack pine populations: effects of insularity. *Canadian Journal of Forest Research*, **22**, 1958–1965.
- Griffin SR, Barrett CH (2004) Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: inferences from phylogeography. *American Journal of Botany*, **91**, 465–473.
- Gugerli F, Sperisen C, Büchler U *et al.* (2001) Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial recolonization of the western Alps. *Molecular Ecology*, **10**, 1255–1263.
- Guiles Ellis K, Mullins HT, Patterson WP (2004) Deglacial to middle Holocene (16,600–6,000 calendar years BP) climate change in the northeastern United States inferred from multi-proxy stable isotope data, Seneca Lake, New York. *Journal of Paleolimnology*, **31**, 343–361.
- Hamilton KGA, Langor DW (1987) Leafhopper fauna of Newfoundland and Cape Breton Islands (Rhynchota: Homoptera: Cicadellidae). *Canadian Entomologist*, **119**, 663–695.
- Heuertz M, Fineschi S, Anzidei M *et al.* (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Molecular Ecology*, **13**, 3437–3452.
- 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.
- Jackson ST, Overpeck JT, Webb T, Keatts SE, Anderson KH (1997) Mapped plant-macrofossils and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**, 1–70.
- Jackson ST, Webb RS, Anderson KH (2000) Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews*, **19**, 489–508.
- Jacobson GL Jr, Webb T III, Grimm EC (1987) Patterns and rates of vegetation change during the deglaciation of eastern North America. In: *North America and Adjacent Oceans During the Last Deglaciation. The Geology of North America* (eds Ruddiman WF, Wright HE Jr), pp. 277–288. DNAG v. K3, Geological Society of America, Boulder, Colorado.
- Jaramillo-Correa JP, Bousquet J, Beaulieu J *et al.* (2003) Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theoretical and Applied Genetics*, **106**, 1353–1367.
- Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, **13**, 2735–2747.
- Jeandroz S, Bastien D, Chandelier A, Du Jardin P, Favre JM (2002) A set of primers for amplification of mitochondrial DNA in *Picea abies* and other conifer species. *Molecular Ecology Notes*, **2**, 389–392.
- Kumar R, Lelu M-A, Small I (1995) Purification of mitochondria and mitochondrial nucleic acids from embryogenic suspension cultures of a gymnosperm, *Larix × leptoeuropaea*. *Plant Cell Reports*, **14**, 534–538.
- Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. *Proceedings of the National Academy of Sciences, USA*, **94**, 5722–5727.
- Latta RG, Linhart YB, Fleck D, Elliot M (1998) Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution*, **52**, 61–67.
- Lu M-Z, Szmidt AE, Wang XR (1998) RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *coxI* gene. *Plant Molecular Biology*, **37**, 225–234.
- Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Molecular Ecology*, **11**, 1441–1455.
- Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

- Matthews JV Jr, Anderson TW, Boyko-Diakonow M *et al.* (1989) Quaternary environments in Canada as documented by paleobotanical case histories. In: *Quaternary Geology of Canada and Greenland. Geological Survey of Canada, Geology of Canada* (ed. Fulton RJ), vol. 1, pp. 483–539. Geological Survey of Canada, Ottawa, Canada. (Also published as Geological Society of America, *The Geology of North America*, vol. K-1. Geological Society of America, Boulder, Colorado).
- McCune B, Grace JB, Urban DL (2002) *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, Oregon.
- McLeod TK, MacDonald GM (1997) Postglacial range expansion and population growth of *Picea mariana*, *Picea glauca* and *Pinus banksiana* in the western interior of Canada. *Journal of Biogeography*, **24**, 865–881.
- Misenti TL, DeHayes DH (1989) Genetic diversity of marginal vs. central populations of pitch pine and jack pine. In: *Proceedings of the Thirty-First Northeastern Forest Tree Improvement Conference and the Sixth North Central Tree Improvement Association* (ed. Demeritt ME Jr), pp. 63–72. Pennsylvania State University, University Park, Pennsylvania.
- Mitton JB, Kreiser BR, Latta RG (2000a) Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. *Molecular Ecology*, **9**, 91–97.
- Mitton JB, Kreiser BR, Rehfeldt GE (2000b) Primers designed to amplify a mitochondrial nad1 intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*. *Theoretical and Applied Genetics*, **101**, 1269–1272.
- MjM Software Design (1999) *PC-ORD. Multivariate analysis of ecological data. Version 4 for Windows*. MjM Software Design, Gleneden Beach, Oregon.
- Monmonier MS (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis*, **3**, 245–261.
- Neale DB, Sederoff RR (1988) Inheritance and evolution of conifer organelle genomes. In: *Genetic Manipulation of Woody Plants* (eds Hanover J, Keathley D), pp. 251–264. Plenum Press, New York.
- Palmé AE, Su Q, Rautenberg A, Manni F, Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular Ecology*, **12**, 201–212.
- Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, **28**, 87–97.
- Perron M, Perry DJ, Andalo C, Bousquet J (2000) Evidence from sequence-tagged-site markers of a recent progenitor-derivative species pair in conifers. *Proceedings of the National Academy of Sciences, USA*, **97**, 11331–11336.
- Petit RJ, Vendramin GG (2005) Phylogeography of organelle DNA in plants: an introduction. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N). Kluwer, Amsterdam, The Netherlands, in press.
- Petit RJ, Pineau E, Demesure *et al.* (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences, USA*, **94**, 9996–10001.
- Petit RJ, Aguinagalde I, de Beaulieu J-L *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.
- Petit RJ, Bialozyt R, Garnier-Géré P, Hampe A (2004) Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. *Forest Ecology and Management*, **197**, 117–137.
- Petit RJ, Duménil J, Fineschi S, Hamp A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.
- Pielou EC (1992) *After the Ice Age. The Return of Life to Glaciated North America*. University of Chicago Press, Chicago, Illinois.
- Piper DJW, Mudie PJ, Letson JRJ, Barnes NE, Iuliucci RJ (1986) *The marine geology of the inner Scotian Shelf of the South Shore. Paper 85–19*. Geological Survey of Canada, Ottawa, Canada.
- Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered *versus* unordered alleles. *Genetics*, **144**, 1237–1245.
- Ritchie JC (1987) *Postglacial Vegetation of Canada*. Cambridge University Press, New York.
- Ross HA, Hawkins JL (1986) Genetic variation among local populations of jack pine (*Pinus banksiana*). *Canadian Journal of Genetics and Cytology*, **28**, 453–458.
- Rudolph TD, Yeatman CW (1982) *Genetics of Jack Pine*. Res. Pap. WO-38. Forest Service, Department of Agriculture, Washington D.C.
- Sáenz-Romero C, Guries RP, Monk AI (2001) Landscape genetic structure of *Pinus banksiana*: allozyme variation. *Canadian Journal of Botany*, **79**, 871–878.
- Schoenike Roland E (1976) *Geographic Variation in Jack Pine*. University of Minnesota Agriculture Experiment Station, Technical Bulletin 304, St. Paul, Minnesota.
- Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model — a computer simulation approach. *Genetics*, **134**, 983–993.
- Sinclair WT, Morman JD, Ennos RA (1999) The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. *Molecular Ecology*, **8**, 83–88.
- Skeates DA (1979) Discontinuity in growth potential of jack pine in Ontario: a 20-year provenance assessment. *Forestry Chronicle*, **55**, 137–141.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Sneath PHA, Sokal RR (1973) *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, California.
- Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome*, **42**, 158–161.
- Soranzo N, Alia R, Provan J, Powell W (2000) Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology*, **9**, 1205–1211.
- Sperisen C, Büchler U, Gugerli F, Matyas G, Geburek T (2001) Tandem repeat in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology*, **10**, 257–263.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Takahata N, Palumbi SR (1985) Extranuclear differentiation and gene flow in the finite island model. *Genetics*, **109**, 441–457.
- Tremblay NO, Schoen DJ (1999) Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Molecular Ecology*, **8**, 1187–1198.

- Wagner DB, Furnier GR, Saghai-Maroo MA *et al.* (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proceedings of the National Academy of Sciences, USA*, **84**, 2097–2100.
- Walter R, Epperson BK (2001) Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology*, **10**, 103–111.
- Walter R, Epperson BK (2005) Geographic pattern of genetic diversity in *Pinus resinosa*: contact zone between descendants of glacial refugia. *American Journal of Botany*, **92**, 92–100.
- Watts WA (1970) The full-glacial vegetation of northwestern Georgia. *Ecology*, **51**, 17–33.
- Watts WA (1979) Late Quaternary vegetation of central Appalachia and the New Jersey coastal plain. *Ecological Monographs*, **49**, 427–469.
- Webb T III (1988) Glacial and Holocene vegetation history: eastern North America. In: *Vegetation History* (eds Huntley B, Webb T III), pp. 385–414. Kluwer Academic, Dordrecht, The Netherlands.
- Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.
- Whitehead DR (1964) Fossil pine pollen and full-glacial vegetation in southeastern North Carolina. *Ecology*, **45**, 767–777.
- Williams JW, Webb T III, Richard PH, Newby P (2000) Late Quaternary biomes of Canada and the eastern United States. *Journal of Biogeography*, **27**, 585–607.
- Williams JW, Shuman BNT, Webb III, Bartlein PJ, Leduc PL (2004) Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. *Ecological Monographs*, **74**, 309–334.
- Wright S (1978) *Evolution and the Genetics of Population, 4. Variability Within and Among Natural Populations*. University of Chicago Press, Chicago, Illinois.
- Wu J, Krutovskii KV, Strauss SH (1998) Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. *Genetics*, **150**, 1605–1614.

This collaborative project reflects the interests of the authors to gain a better understanding of genetic variation, biogeographical history, and evolutionary processes in trees. Julie Godbout is a PhD candidate studying the genetic diversity and population structure of pine species. Juan Pablo Jaramillo-Correa is a postdoctoral fellow studying various aspects of the molecular ecology of tree species. Jean Beaulieu is a research scientist interested in quantitative genetics and problems related to the biodiversity of tree species. Jean Bousquet is a professor involved in the application of molecular and genomic approaches to studying the evolutionary biology of plant and tree species.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.