MICROSATELLITE ANALYSIS REVEALS GENETICALLY DISTINCT POPULATIONS OF RED PINE (PINUS RESINOSA, PINACEAE)\textsuperscript{1}

JACQUELYN BOYS,\textsuperscript{2} MARILYN CHERRY,\textsuperscript{3} AND SELVAURAI DAYANANDAN\textsuperscript{2,4}

\textsuperscript{1}Forest and Evolutionary Genomics Laboratory, and the Center for Structural and Functional Genomics, Biology Department, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec H4B IR6, Canada; \textsuperscript{2}Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, Oregon 97331-5752 USA

Red pine (\textit{Pinus resinosa} Ait.) is an ecologically and economically important forest tree species of northeastern North America and is considered one of the most genetically depauperate conifer species in the region. We have isolated and characterized 13 nuclear microsatellite loci by screening a partial genomic library with di-, tri-, and tetrancleotide repeat oligonucleotide probes. In an analysis of over 500 individuals representing 17 red pine populations from Manitoba through Newfoundland, five polymorphic microsatellite loci with an average of nine alleles per locus were identified. The mean expected and observed heterozygosity values were 0.508 and 0.185, respectively. Significant departures from Hardy-Weinberg equilibrium with excess homozygosity indicating high levels of inbreeding were evident in all populations studied. The population differentiation was high with 28–35% of genetic variation partitioned among populations. The genetic distance analysis showed that three northeastern (two Newfoundland and one New Brunswick) populations are genetically distinct from the remaining populations. The coalescence-based analysis suggests that “northeastern” and “main” populations likely became isolated during the most recent Pleistocene glacial period, and severe population bottlenecks may have led to the evolution of a highly selfing mating system in red pine.

Key words: genetic bottleneck; genetic diversity; microsatellites; \textit{Pinus}; Pleistocene refugia; population genetics; Postglacial colonization.

Red pine, \textit{Pinus resinosa} Ait., is one of the most extensively planted trees in the northern United States and Canada and is widely used as a source of wood for poles, cabin logs, and fuel. It is distributed throughout northeastern North America, from southeastern Manitoba, eastward through the Great Lakes/St. Lawrence region to Newfoundland, and south to West Virginia (Rudolf, 1990). Despite its wide geographic distribution, red pine shows high morphological uniformity and is considered to be one of the most genetically depauperate conifer species in North America (Fowler and Lester, 1970). Its genetic uniformity may have resulted from passage through a genetic bottleneck during the last glacial period (Fowler and Morris, 1977; Walter and Epperson, 2001, 2005). The fossil data, dating from 16 000 to 18 000 years BP (before present), suggest that red pine populations persisted in refugia located in the southern Appalachian Mountains (Jackson et al., 2000). Red pine may also have survived during the glacial period in refugia off the present coastline of the eastern seaboard in nonglaciated islands and extensions of the mainland (Pielou, 1991). The highly fragmented population structure and self-compatible mating system (Fowler, 1964, 1965) may also have contributed to the loss of genetic variation through inbreeding and genetic drift during the post-glacial expansion from these refugia. The genetic uniformity of red pine contrasts strikingly with the high genetic variability found in most conifer species (Hamrick and Godt, 1990). However, low genetic diversity has also been reported in several conifer species including torrey pine (\textit{Pinus torreyana} Parry ex Carr; Ledig and Conkle, 1983), red spruce (\textit{Picea rubens} Sarg; Hawley and DeHayes, 1994), and western red cedar (\textit{Thuja plicata} Donn. Ex D. Don; Glaubitz et al., 2000).

Population bottlenecks are known to reduce genetic variability, purge deleterious mutations, and lower inbreeding depression, thereby promoting evolution of self-fertilization (Charlesworth and Charlesworth, 1979; Lande and Schemske, 1985). The highly self-compatible mating system in red pine (Fowler, 1965) may have evolved as a consequence of historical population bottlenecks. The low genetic variability of red pine may compromise its ability to respond to selective pressures such as changes in environmental conditions and may increase its risk of extinction from threats such as pests and disease. It has been predicted that red pine will be extirpated from the USA within the next century because of increases in temperature during the growth season (Prasad and Iverson, 1999; Cherry, 2001). Further decline in natural populations of red pine has been caused by harvesting without regeneration and by suppressing the fire that is necessary for seedling recruitment (Mosseler et al., 1992). As a mitigating measure against potential threats, genetically distinct populations of red pine need to be identified and conserved to ensure long-term survival of the species.

Estimates of effective population sizes (Schoen and Brown,
1991; Beerli and Felsenstein, 2001), selling rates of populations, gene flow, and mean coalescent times (Slatkin, 1995) are needed to gain insights into the evolutionary history of red pine and to formulate genetically sound conservation plans. These parameters could be estimated using polymorphic genetic markers. However, because of low or no variability, the traditional genetic markers are of limited utility for assessing the genetic architecture of red pine populations. Lack of genetic variability in red pine has been reported in isozyme (Fowler and Morris, 1977; Allendorf et al., 1982; Simon et al., 1986; Mosseler et al., 1991) and RAPD analyses (Mosseler et al., 1997). A partial genomic library of red pine DNA fragments between 500 to 1000 bp was constructed and screened with (TC) 15, (AC) 15, (AT) 15, (AAAT) 10, (TG) 15, (AG) 15, (AGAT) 15, and (AAAG) 10 oligonucleotide probes following the protocol of Dayanandan et al. (1998). Either single-stranded DNA (isolated using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, California, USA).

Oligonucleotide primers complementary to regions flanking the repeat regions were designed. These primers were synthesized by Operon Technologies (Alameda, California, USA). The polymerase chain reaction (PCR), under optimized conditions, was used for amplifying microsatellites in the sampled individuals. Amplification reactions were performed in a total volume of 25 μl with 0.2 mM dNTP, 2.0 mM MgCl2, 2.5 μl 10× buffer, and 2.5 pmol of each primer. The thermal cycling profile of PCR reactions consisted of an initial denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 1 min (denaturation), 1 min annealing (at the optimal annealing temperature as given in Table 1), and 72°C for 1 min (extension). This was followed by 72°C for 4 min (final extension). The sizes of amplified products were determined

### MATERIALS AND METHODS

#### Samples collected

A total of 518 individuals were sampled from 17 different red pine populations distributed throughout its natural range (Table 1, Fig. 1).

#### Isolation and characterization of nuclear microsatellite markers

DNA was extracted from needle samples following the method described in Dayanandan et al. (1997). A partial genomic library of red pine DNA fragments between 500 to 1000 bp was constructed and screened with (TC) 15, (AC) 15, (AT) 15, (AAAT) 10, (TG) 15, (AG) 15, (AGAT) 15, and (AAAG) 10 oligonucleotide probes following the protocol of Dayanandan et al. (1998). Either single-stranded DNA (isolated using the Wizard M13 DNA purification system; Promega, Madison, Wisconsin, USA) or double-stranded DNA (isolated using the QA prep Spin Miniprep Kit; Qiagen, Mississauga, Ontario, Canada) of positive clones were sequenced using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, California, USA).
using an ABI 310 genetic analyzer and Genescan software (Applied Biosystems).

Genetic diversity and mating system analyses—For each polymorphic locus, the total number of alleles per locus and per population was determined, and the distribution of allele frequencies was calculated. The average number of alleles per locus (A) for each population over all loci and unbiased values of expected (H_e) and observed (H_o) frequency of heterozygotes were calculated using the Genetic Data Analysis (GDA) version 1.0 software (Lewis and Zaykin, 2001). Deviations from Hardy-Weinberg equilibrium and linkage disequilibrium were tested using Fisher exact tests with 3200 reshufflings. The heterozygote deficiency for each locus and population was tested using Genepop 3.4 software (Raymond and Rousset, 1995) following the method of Rousset and Raymond (1995). Population genetic structure was analyzed through hierarchical F coefficients (Weir and Cockerham, 1984) using the GDA software. The statistical significance of F coefficients at the 95% confidence level was tested through bootstrap analysis with 10,000 replicates. The selfing rate (S) in each population was calculated using the formula S = 2H/1 + F (Crow and Kimura, 1970).

The standard genetic distances, D_ST (Nei, 1978), were calculated using the GDA software, and R_ST (Slatkin, 1995) and delta μ^2 (Goldstein et al., 1995) were calculated using RSTCALC (Goodman, 1997). The F_ST and D_ST assume an infinite allele model (IAM), while R_ST and delta μ^2 assume a stepwise mutation model (SMM). The concordance between F_ST and R_ST values, as well as between D_ST and delta μ^2 distance values were tested through Mantel tests to determine if our results depended on the underlying mutation model of the analyses. The unweighted pair group method with arithmetic average (UPGMA) dendrograms based on D_ST and delta μ^2 were constructed and visualized with the TreeView software program (Page, 1996). The relationships between genetic distances and geographic distance were analyzed using Isolation By Distance (IBD) software (Bohonak, 2002). The IBD software program assesses the significance between a given distance matrix and the geographic distance through Mantel’s test and evaluates the strength of the relationship through regression of all pairwise genetic distances with their corresponding geographic distances. We tested the relationships between D_ST and delta μ^2 with geographic distance, Rousset (1997) distance with the logarithm of geographic distance, and the logarithm of Slatkin (1993) distance with the logarithm of geographic distance.

Estimation of effective population sizes, migration rates, and coalescence times—Genetic distance analysis revealed two major clusters of populations, the Northeastern group consisting of three populations (Terra Nova and Rowells Brook in Newfoundland, and Geary in New Brunswick) and the other group consisting of the 14 remaining populations (Figs. 2, 3). Hereafter we refer to these two groups as Northeast and Main populations, respectively. We have estimated the effective population sizes, migration rates, and coalescence times for Northeast, Main, and total (Northeast + Main) groups separately. Coalescent-theory-based maximum likelihood estimation of effective population sizes was carried out using MIGRATE software (Beerli and Felsenstein, 2001). The value of θ was estimated using the MIGRATE software, and assuming an average mutation rate of microsatellites as 10^-3 per generation, we calculated N_e as θ/4 × 10^-3. The R_ST-based migration rates (M_ST) and coalescent times (τ) were estimated following Slatkin (1995). We calculated the average coalescent time T, following equation 19 of Slatkin (1995). The coalescent time in generations was calculated as τ = T_n × N, where N is the size of each group. We used the these estimates of effective population sizes as N, and assuming an average generation length of red pine as 50 yr, we estimated the absolute coalescent time as τ = τ × 50 yr.

RESULTS

Recovery of microsatellite loci and detection of polymorphism—A partial genomic library consisting of ~13,000 clones was screened using di-, tri-, and tetra-nucleotide repeat oligonucleotide probes. The sequencing of 35 positive clones identified through screening with TC/AG probes revealed 19
microsatellites (17 TC/AG, one AT, and one AT mixed repeat), and primers were designed for nine clones. Seven primers were designed based on 23 microsatellites (12 AC/TG, three AT, three TC/AG, three AC/TG and AT mixed repeats, one TCC/AGG repeat, and one 6-bp repeat of TCCACA) identified through the sequencing of 68 clones positive for AC/TG probes. The AT/AAT/AAAT probes identified 13 positive clones, but sequencing revealed microsatellites (AT/AAT/AAAT repeats) in nine clones, and primers were designed based on nine clones. A total of 25 oligonucleotide primer pairs complementary to flanking regions of microsatellites were designed. Optimization of PCR conditions resulted in successful amplification with 20 primer pairs. Thirteen of these 20 successfully amplified primer pairs produced amplification products consistent with a single-locus segregation pattern, and four primer pairs showed polymorphism. One of these polymorphic primer pairs had an amplification pattern consistent with two loci, and alleles of each locus were easily distinguishable for scoring. Of the remaining 12 primer pairs, seven primer pairs produced multiple amplification products, and amplification by the five primer pairs was inconsistent.

Levels of genetic variation and population genetic structure—In a total of 518 red pine individuals surveyed, 45 alleles were identified at five polymorphic microsatellite loci, with an average of nine alleles per locus. The highest diversity was found at the (AT)22 repeat (locus PRE16) with 15 alleles. The loci PRE10, PRE13, PRE24, and PRE24B had 13, 12, 3, and 2 alleles, respectively (Table 2).

Among the 17 populations surveyed, the highest allelic diversity with an average of 4.6 alleles per locus was found in Delta County (Michigan) and the lowest in Hardy County (West Virginia) with an average of 1.2 alleles per locus. Private alleles were identified in six of the 17 populations. One private allele per population was found in the Petawawa (Ontario), Terra Nova (Newfoundland), and Rowsells Brook (Newfoundland) populations, two were found in the Abitibi (Quebec) and South Junction (Manitoba) populations, and three were detected in the Delta County (Michigan) population. The $H_e$ values ranged from 0.242 at locus PRE24B to 0.789 at locus PRE16, while the $H_o$ values ranged from 0.067 at locus PRE13 to 0.317 at locus PRE16. The $H_e$ values across all loci were highest in the Beauceville (Quebec), Tracy1 (New Brunswick), and Geary (New Brunswick) populations, with values of 0.489, 0.486, and 0.483, respectively. The lowest $H_e$ values were found in the Hardy County (West Virginia), High Ledges (Massachusetts), and Pendleton County (West Virginia) populations, with values of 0.014, 0.142, and 0.211, respectively. Observed heterozygosity ranged from 0 in Hardy County (West Virginia) to 0.269 in Beaver Lake (Nova Sco-
Mean $H_o$ and $H_e$ values were 0.185 and 0.508, respectively (Table 2). Significant departure from Hardy-Weinberg equilibrium was observed at most loci in most populations, and significant heterozygote deficiency was evident in all populations at all loci except for PRE24B. No linkage disequilibrium was observed, indicating all five loci segregate independently of each other. The values of $F_{ST}$ were $-0.133$ to $0.456$, $0.221$ to $1.000$, and $0.211$ to $0.455$, respectively. The bootstrap analyses showed that overall $F$ values were significantly $(P = 0.05)$ greater than 0 (Table 3). The overall $F_{ST}$ value was 0.280, while the overall $R_{ST}$ value was 0.350. The $F$ values ranged from 0.384 to 1.000 across all populations and were significantly greater than 0 for all populations except for the two Newfoundland populations. The selting rates ($S$) ranged from 0.555 in the Beaver Lake (Nova Scotia) population to 1 in the Hardy County, West Virginia (Table 1).

Nei’s genetic distances $D_s$ (Table 4) ranged from 0 between Sioux Lookout (Ontario) and Delta County (Michigan) to 1.487 between Terra Nova (Newfoundland) and High Ledges (Massachusetts). The delta $\mu^2$ distances (Table 4) ranged from 0.010 between Sioux Lookout (Ontario) and Delta County (Michigan) to 3.887 between Rowsells Brook (Newfoundland) and Delta County (Michigan). The Mantel test results indicated no significant differences between pairwise estimates of $F_{ST}$ and $R_{ST}$ ($r = 0.91$, $P = 0.0001$), and genetic distance measures, $D_s$ and delta $\mu^2$, were highly congruent ($r = 0.94$, $P = 0.0001$) with each other.

The dendograms based on $D_s$ and delta $\mu^2$ (Figs. 2, 3) showed that the two Newfoundland populations and the Geary (New Brunswick) population were genetically distinct from the remaining populations. The topologies of the two dendrograms were similar except for the placement of the Beaueville (Quebec) population. This population grouped with the New Brunswick and Nova Scotia populations in the dendrogram based on $D_s$, but showed close affinity to the Petawawa (Ontario) population in the dendrogram based on delta $\mu^2$ distances.

Weak but significant positive correlations were found between geographic distance and all genetic distance measures tested ($D_s$: $r = 0.3812$, $P = 0.0072$; delta $\mu^2$: $r = 0.3323$, $P = 0.0144$; Roussel’s distance: $r = 0.3120$, $P = 0.0291$). Similarly, a negative correlation was found between geographic distance and the Slatkin’s (1993) measure of similarity ($r = -0.2500$, $P = 0.0214$). The effective population sizes, migration rates and coalescent times of Northeast Main, and total populations are given in Table 5. The effective population sizes of Northeast, Main, and total populations were 62, 222, and 615, respectively. The migration rate within Northeast and Main populations was 0.9, and the migration between Northeast and Main populations was 0.42. The average coalescent time in the Northeast population was 69 generations (3400 yr), and the corresponding value for the Main population was 246 generations (12300 yr). The coalescent time of the total population was 1463 generations (73154 yr).

**DISCUSSION**

The low recovery rate of microsatellites with a single-locus inheritance pattern in this study is typical of organisms with a large genome size such as conifers (Fischer and Bachmann, 1998) and is in agreement with several studies (Smith and Devey, 1994; Kostia et al., 1995; Pfeiffer et al., 1997; Fisher et al., 1998; Beaureb et al., 2003). The reason for this could be multiple sites for primer binding as a result of genome duplications, a common phenomenon in large genomes, or an increased probability of nonspecific binding sites for the designed primers (Garner, 2002).

The greater number of TC/AG repeats relative to AC/TG repeats in red pine is consistent with findings for several tree species, including *Populus tremuloides* (Dayanandan et al., 1998), *Pinus radiata* (Fisher et al., 1998), and *Picea glauca* (Dayanandan et al., 2001). An exception is white pine (*Pinus strobus*), in which AC/TG repeats were found to be more abundant than TC/AG (Echt et al., 1996). Although microsatellite primers based on longer repeats are expected to demonstrate higher levels of polymorphism (Weber, 1990), the greatest polymorphism in red pine was detected in locus PRE16 with 15 alleles.

### Table 2. Primer sequences and PCR annealing temperatures ($T$), repeat pattern, allele sizes, number of alleles detected ($A$), and observed ($H_o$) and expected ($H_e$) heterozygosity values for five polymorphic nuclear microsatellite loci in *Pinus resinosa*.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequences (5′–3′)</th>
<th>Repeat motif</th>
<th>$T$ (°C)</th>
<th>Size (bp)</th>
<th>$A$</th>
<th>$H_o$</th>
<th>$H_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE10</td>
<td>F:GGTTCCTGGCAAAAGAGTTGAGG&lt;br&gt;R:GTTGGGAGATAGTGATAGAGT</td>
<td>(AT)$_{31}$</td>
<td>55</td>
<td>325–355</td>
<td>12</td>
<td>0.242</td>
<td>0.683</td>
</tr>
<tr>
<td>PRE13</td>
<td>F:GATGTCCTTATGGCTTTGTC&lt;br&gt;R:AGGTTGAATACCGGCTTGT</td>
<td>(AT)$_{31}$</td>
<td>60</td>
<td>245–277</td>
<td>13</td>
<td>0.067</td>
<td>0.526</td>
</tr>
<tr>
<td>PRE16</td>
<td>F:TCTGGGATAGCTCTTTTGT&lt;br&gt;R:TCATTTTTTTTCTTTTCTTTAC</td>
<td>(AT)$_{32}$</td>
<td>50</td>
<td>334–369</td>
<td>15</td>
<td>0.317</td>
<td>0.789</td>
</tr>
<tr>
<td>PRE24</td>
<td>F:GTTTTCATGAGGCGAGG&lt;br&gt;R:CGTGGGGAGATAGTGATAGAGT</td>
<td>(AG)$_{29}$</td>
<td>50</td>
<td>227–231</td>
<td>3</td>
<td>0.105</td>
<td>0.302</td>
</tr>
<tr>
<td>PRE24B</td>
<td>Same as PRE24</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.192</td>
<td>0.242</td>
</tr>
<tr>
<td>All loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.185</td>
<td>0.508</td>
</tr>
</tbody>
</table>

**Notes:** $F_{ST}$ = Individual inbreeding coefficient relative to the subpopulation, $F_{RT}$ = Individual inbreeding coefficient relative to the total population, $F_{ST}$ = inbreeding coefficient of subpopulations relative to the total population.

### Table 3. Inbreeding ($F$) coefficients based on five polymorphic nuclear microsatellite loci of *Pinus resinosa*.

<table>
<thead>
<tr>
<th>Locus</th>
<th>$F_{ST}$</th>
<th>$F_{RT}$</th>
<th>$F_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE10</td>
<td>0.535</td>
<td>0.651</td>
<td>0.250</td>
</tr>
<tr>
<td>PRE13</td>
<td>0.840</td>
<td>0.874</td>
<td>0.211</td>
</tr>
<tr>
<td>PRE16</td>
<td>0.456</td>
<td>0.605</td>
<td>0.273</td>
</tr>
<tr>
<td>PRE24</td>
<td>0.381</td>
<td>0.663</td>
<td>0.455</td>
</tr>
<tr>
<td>PRE24B</td>
<td>-0.133</td>
<td>0.221</td>
<td>0.312</td>
</tr>
<tr>
<td>Overall</td>
<td>0.504</td>
<td>0.643</td>
<td>0.280</td>
</tr>
<tr>
<td>Bootstrap upper</td>
<td>0.690</td>
<td>0.770</td>
<td>0.361</td>
</tr>
<tr>
<td>Bootstrap lower</td>
<td>0.278</td>
<td>0.492</td>
<td>0.236</td>
</tr>
</tbody>
</table>

The dendograms were similar except for the placement of the Beaueville (Quebec) population. This population grouped with the New Brunswick and Nova Scotia populations in the dendrogram based on $D_s$, but showed close affinity to the Petawawa (Ontario) population in the dendrogram based on delta $\mu^2$ distances. Weak but significant positive correlations were found between geographic distance and all genetic distance measures tested ($D_s$: $r = 0.3812$, $P = 0.0072$; delta $\mu^2$: $r = 0.3323$, $P = 0.0144$; Roussel’s distance: $r = 0.3120$, $P = 0.0291$). Similarly, a negative correlation was found between geographic distance and the Slatkin’s (1993) measure of similarity ($r = -0.2500$, $P = 0.0214$). The effective population sizes, migration rates and coalescent times of Northeast Main, and total populations are given in Table 5. The effective population sizes of Northeast, Main, and total populations were 62, 222, and 615, respectively. The migration rate within Northeast and Main populations was 0.9, and the migration between Northeast and Main populations was 0.42. The average coalescent time in the Northeast population was 69 generations (3400 yr), and the corresponding value for the Main population was 246 generations (12300 yr). The coalescent time of the total population was 1463 generations (73154 yr).

**DISCUSSION**

The low recovery rate of microsatellites with a single-locus inheritance pattern in this study is typical of organisms with a large genome size such as conifers (Fischer and Bachmann, 1998) and is in agreement with several studies (Smith and Devey, 1994; Kostia et al., 1995; Pfeiffer et al., 1997; Fisher et al., 1998; Beaureb et al., 2003). The reason for this could be multiple sites for primer binding as a result of genome duplications, a common phenomenon in large genomes, or an increased probability of nonspecific binding sites for the designed primers (Garner, 2002).

The greater number of TC/AG repeats relative to AC/TG repeats in red pine is consistent with findings for several tree species, including *Populus tremuloides* (Dayanandan et al., 1998), *Pinus radiata* (Fisher et al., 1998), and *Picea glauca* (Dayanandan et al., 2001). An exception is white pine (*Pinus strobus*), in which AC/TG repeats were found to be more abundant than TC/AG (Echt et al., 1996). Although microsatellite primers based on longer repeats are expected to demonstrate higher levels of polymorphism (Weber, 1990), the greatest polymorphism in red pine was detected in locus PRE16 with 15 alleles.
which was designed based on a clone with (AT)22 repeat, whereas the locus PRE24, which was designed based on a clone with longest repeat, (AG)29, had the least polymorphism, with only three alleles at the locus PRE24 and two alleles at locus PRE24B. This pattern may not be unusual for forest trees, however, because other studies have indicated that loci based on longer repeats do not necessarily have higher levels of polymorphism (Echt et al., 1996; Dayanandan et al., 1998). The microsatellite locus based on the shortest cloned repeat demonstrating polymorphism in red pine was (AT)21, whereas all primer pairs based on clones with repeats shorter than 21 were monomorphic.

The observed heterozygosity ($H_o$) values of 0.067–0.317 with a mean of 0.185 reported here for red pine are lower than those reported for other pine species using microsatellite markers. For example, $H_o$ values for $P. radiata$ ranged from 0 to 0.85 with a mean of 0.625 (Smith and Devey, 1994), and those for $P. strobus$ ranged from 0.125 to 0.812 with a mean of 0.515 (Echt et al., 1996). The $H_o$ values for red pine are much lower than expected heterozygosity ($H_e$) values (Tables 1, 2), and all populations as well as four of the five loci showed significant heterozygote deficiency. Despite the lower levels of $H_o$ and relatively few polymorphic loci used, the average of nine alleles per locus across the five polymorphic loci characterized for red pine was comparable to or greater than microsatellite-based genetic studies of other pines, including six alleles per locus observed in $P. radiata$ (Smith and Devey, 1994), 5.4 alleles per locus in $P. strobus$ (Echt et al., 1996), and 6.7 alleles per locus in $P. sylvestris$ (Soranzo et al., 1998). Similar results of relatively high genetic diversity but low heterozygosity have been reported in $Elymus alaskanus$, and this pattern has been attributed to a high level of inbreeding (Sun and Salomon, 2003).

Genetic diversity across red pine populations was highly variable, with $H_o$ values ranging from 0.489 in Beauceville (Quebec) to 0.014 in Hardy County (West Virginia). The highest allelic diversity was found in the Delta County (Michigan) population, and the lowest in the Hardy County (West Virginia) population. The presence of unique or private alleles can be considered a measure of genetic distinctiveness. Private alleles were found in several populations in both the western and eastern extremes of the red pine range. The lower number of alleles detected in the Terra Nova (Newfoundland) population, however, may be due to smaller sample size. From a conservation perspective, the distribution pattern of alleles, the presence of unique alleles in several of the populations examined, and the distribution of genetic variation in red pine could be used as a genetic criterion to protect as many distinct populations as possible throughout its range.
The $F_{ST}$ and $F_{IT}$ values measure the deviation of genotypic frequencies from Hardy-Weinberg equilibrium within subpopulations and in the total population, respectively. The overall $F_{ST}$ and $F_{IT}$ values were significantly higher than zero and suggest a departure from Hardy-Weinberg equilibrium with deficiency of heterozygotes. The increased levels of homozygosity and the departure from Hardy-Weinberg equilibrium could be attributable to a mating system with a high level of inbreeding through self-fertilization (Fowler, 1965) or through mating between closely related individuals. Several factors, including the presence of null alleles due to sequence variation at the primer sites that prevent amplification during PCR, and selection against heterozygotes, can also result in heterozygote deficiency. However, the pattern detected here was consistent across all loci, and therefore it is unlikely that the high levels of observed homozygosity result from null alleles or selection against heterozygotes.

The level of evolutionary divergence of populations is influenced by the opposing effects of migration, which tends to homogenize populations, and genetic drift or mutation, which leads to population differentiation. The overall $F_{IT}$ value of 0.280 and overall $R_{ST}$ value of 0.350 indicate that 28–35% of the genetic variation present in red pine is partitioned among subpopulations. This indicates that red pine populations are highly differentiated, an unusual feature of many conifers mainly because of long-distance pollen flow facilitated by wind pollination. For example, allozyme studies demonstrated low differentiation ($F_{ST} \leq 0.02$) among northern populations of European Scots pine (Gullberg et al., 1985), and the proportions of genetic diversity detected among populations of several widespread conifer species are usually less than 10% (Ledig, 1998). The high genetic differentiation among populations in red pine could be attributable to a high rate of selfing in red pine, because predominantly selfing plant species tend to show higher population differentiation than outcrossing species (Loveless and Hamrick, 1984; Hamrick and Godt, 1990).

The high selfing rate observed in red pine (Table 1) is unusual among members of Pinaceae, in which most species possess a primarily outcrossing mating system. The evolution of self-fertilization is promoted by population bottlenecks (Lande and Schemske, 1985), and selfing is known to have evolved multiple times from outcrossing lineages (Stebbins, 1957). Furthermore, Lande and Schemske (1985) demonstrated that selfing is promoted if inbreeding depression is less than 50%. In general, the inbreeding depression values of the majority of conifers are high (Williams and Savolainen, 1996), and therefore a selfing mating system is not favored. However, inbreeding depression in red pine is considerably low, and the lethal equivalents per diploid zygote of red pine is the lowest among pines (Fowler, 1964), suggesting that population bottlenecks may have purged deleterious mutations, thereby reducing inbreeding depression and promoting the evolution of self-fertilization in red pine.

The mutation model of microsatellites for calculating genetic distances and population differentiation remains controversial. The infinite allele model (IAM) assumes that each mutation creates a novel allele, whereas the stepwise mutation model (SMM) assumes a closer genetic relatedness between alleles of similar size. The $F_{ST}$ (Wright, 1965) and standard genetic distance $D_s$ (Nei, 1978) are based on the IAM, whereas the $R_{ST}$ (Slatkin, 1995) and delta $\mu^2$ are based on the SMM. Mantel test results showed no significant differences between $R_{ST}$ and $F_{ST}$ estimates, nor between $D_s$ and delta $\mu^2$ values, indicating that the genetic patterns detected in our study are not dependent on the underlying mutation model and confirming the robustness of our findings.

The genetic distance analyses showed that the Northeastern population group, comprising the two Newfoundland populations and Geary in New Brunswick, is distinct from the Main (western) population. The high gene flow rates within the Northeastern and Main populations and lower gene flow rate between the Northeastern and Main populations further support the genetic isolation between these two populations. The smaller effective population size of Northeastern populations compared to Main populations suggests that Northeastern populations may have experienced a more severe bottleneck than the western populations. The coalescent time estimate of 70,000 years suggests that the isolation of Northeastern from Main populations may have occurred during the most recent Pleistocene glacial period. Although the assumptions underlying these estimates (mutation rate of microsatellite loci as $10^{-3}$ and generation time of red pine as 50 yr) are debatable, these values agree with data from the current literature. In plant microsatellites, mutation rates of $1 \times 10^{-3}$ and $3.9 \times 10^{-3}$ in chickpeas (Udupa and Baum, 2001), $2.4 \times 10^{-4}$ in durum wheat (Thuillet et al., 2002), and $7.7 \times 10^{-3}$ in maize (Vigouroux et al., 2002) have been reported. Seed production in red pine is known to reach maximum levels at approximately 50 yr of age (Rudolf, 1990).

Among the red pine populations examined in this study, the area of highest genetic diversity based on expected heterozygosity values is centered in the New Brunswick–Nova Scotia region and extends to Quebec. This is consistent with the evidence from cpSSR data (Echt et al., 1998; Walter and Epperson, 2001, 2005). This pattern of genetic structure could have arisen from the postglacial colonization of this area by individuals from multiple refugia, including southern Appalachian and northeastern refugia, forming an admixed population through secondary contact (Walter and Epperson, 2001, 2005). The presence of highest allelic diversity and greatest number of private alleles in the Michigan population suggests the possibility of a glacial refugium in the Michigan area. Although no direct evidence exists for red pine in this area during the last glacial period, macrofossils of other conifers (e.g., *Picea* spp.) have been found in the area south of Michigan (Jackson et al., 2000). Thus, based on genetic evidence, red pine may have colonized its current range from as many as three glacial refugia.

Higher genetic variation is often found in southern populations relative to northern populations, as is evident from several taxa that survived in much reduced refugial populations in Europe during the Pleistocene glacial periods (Hewitt, 2001). A similar trend has also been demonstrated for plants in western (Solit, et al., 1997) and eastern (Lewis and Crawford, 1995) North America, suggesting low diversity in newly colonized northern areas. In our study, however, the lowest genetic diversity was found in the three most southern populations, a pattern inconsistent with post-glacial colonization from a single southern glacial refugium. This further supports the existence of multiple refugia. However, reduced genetic diversity in the smaller and more isolated population in West Virginia could result from lack of gene flow and genetic drift in marginal populations, because the effect of genetic drift may be two to 30 times greater near the edge of a species range because of smaller effective population sizes (Vucetich and Waite, 2003).
In summary, these polymorphic microsatellite markers are highly valuable to assess genetic diversity and identify genetically unique populations of red pine, one of the economically and ecologically important, but genetically depauperate tree species in North America. The geographical distribution and genetic structuring of red pine populations strongly suggest dispersal from multiple refugia with an intermediate admixture zone. The high levels of population differentiation maintained in this species may result from high levels of self-pollination and inbreeding. The information obtained in this study will be valuable not only for the theoretical advancement of our knowledge on the plant genetic diversity in relation to geological history and refugial habitats during the Pleistocene era, but also valuable for designing genetically sound conservation and management programs for red pine.

LITERATURE CITED


Charlesworth, B., AND D. CHARLESWORTH. 1979. The evolutionary geneti-


Cherry, M. 2001. Options for allocating afforestation stock in Ontario with


theory. Harper and Row, New York, New York, USA.


Dayanandan, S., O. P. RAJORA, AND K. S. BAWA. 1998. Isolation and character-

erization of microsatellites in trembling aspen (Populus tremuloides).

Theoretical and Applied Genetics 96: 950-956.


Firestone, K. B., B. A. HOULDEN, W. B. SHERWIN, AND E. GEFEN. 2000. Variability and differentiation of microsatellites in the genus Dasyurus and conservation implications for the large Australian carnivorous marsupi-


Fisch, D., AND K. BACHMANN. 1998. Microsatellite enrichment in organ-


WO-8, USDA Forest Service, Washington, D.C.


Goldstein, D. B., A. R. LINARES, L. L. CAVALLI-SFORZA, AND M. W. FELD-


GOODMAN, S. J. 1997. RQuadCalc: a collection of computer programs for cal-


Sinauer, Sunderland, Massachusetts, USA.


Hewitt, G. M. 2001. Speciation, hybrid zones and phylogeography—or see-
genes in space and time. Molecular Ecology 10: 537-549.

Jackson, S. T., R. S. WEBB, K. H. ANDERSON, J. T. OVERPECK, T. WEBB III,

J. W. WILLIAMS, AND B. C. S. HANSSEN. 2000. Vegetation and environment in Eastern North America during the last glacial maximum. Qua-

ternary Science Reviews 19: 489-508.


lution 37: 79-85.

LEWIS, P. O., AND D. J. CRAWFORD. 1995. Pleistocene refugium endemics exhibit greater allozymic diversity than widespread congeners in the ge-

genus Polygonylum (Polygonaceae). American Journal of Botany 82: 141-

149.

LEWIS, P. O., AND D. ZAYKIN. 2001. Genetic Data Analysis: Computer pro-

gram for the analysis of allelic data, version 1.0 (d16c). Distributed by the authors at website http://lewis.eeb.uconn.edu/lewishome/software.

LOVELESS, M. D. 1996. TREEVIEW: an application to display phylogenetic


PFEIFFER, A., A. M. OLIVIER, AND M. MORGANTE. 1997. Identification and
characterization of microsatellites in Norway spruce (Picea abies K.)


