Spatial and population genetic structure of microsatellites in white pine

PAULA E. MARQUARDT* and BRYAN K. EPPERSON†
*USDA Forest Service, North Central Research Station, Forestry Sciences Laboratory, 5985 Highway K, Rhinelander, WI 54501, USA; †Michigan State University, Department of Forestry, 126 Natural Resources Building, East Lansing, MI 48824, USA

Abstract

We evaluated the population genetic structure of seven microsatellite loci for old growth and second growth populations of eastern white pine (*Pinus strobus*). From each population, located within Hartwick Pines State Park, Grayling, Michigan, USA, 120–122 contiguous trees were sampled for genetic analysis. Within each population, genetic diversity was high and inbreeding low. When comparing these populations, there is a significant, but small (less than 1%), genetic divergence between populations. Spatial distance between populations or timber harvest at the second growth site were reasonable explanations for the observed minor differences in allele frequencies between populations. Spatial autocorrelation analysis suggested that, for the old growth population, weak positive structuring at 15 m fits the isolation by distance model for a neighbourhood size of about 100 individuals. In comparison, genotypes were randomly distributed in the second growth population. Thus, logging may have decreased spatial structuring at the second growth site, suggesting that management practices may be used to alter natural spatial patterns. In addition, the amount of autocorrelation in the old growth population appears to be lower for some of the microsatellites, suggesting higher numbers of rare alleles and that higher mutation rates may have directly affected spatial statistics by reducing structure.

Keywords: eastern white pine (*Pinus strobus*), Hartwick Pines State Park, microsatellite, population genetics, spatial autocorrelation.

Received 1 April 2004; revision received 3 August 2004; accepted 3 August 2004

Introduction

Eastern white pine (*Pinus strobus*) is a widespread tree species in the north-eastern USA and Canada, having both historical and economic value. Despite its importance, little is known about the mating system of white pine, its spatial genetic structure (the degree of relatedness between individuals separated by distance) and how its genetic diversity is affected by changes in the forest ecosystem. For example, it is largely unknown how management treatments alter the population level spatial patterns of genes, which in turn, influence other aspects of population genetic structure, such as inbreeding.

Generally, conifers have a high genetic load (reviewed by Ledig 1998) and can suffer from strong inbreeding depression, resulting in lower survival and slower growth (Johnson 1945; Fowler 1965; Wu et al. 1998). However, inbreeding is typically low in pine if the populations are not isolated. For example, inbreeding is close to zero for eastern Mediterranean pine (*Pinus brutia*; Panetsos et al. 1998), lodgepole pine (*Pinus contorta*; Epperson & Allard 1984), and eastern white pine (Beaulieu & Simon 1995; Epperson & Chung 2001). Although spatial genetic structure is expected to be weak in continuous populations such as white pine, it can be the most important factor controlling levels of inbreeding. For example, El-Kassaby et al. (1993) reported a mean outcrossing rate (single locus) of 94% for western white pine (*Pinus monticola*), indicating possible spatial family structure.

Few studies examine the spatial genetic structure within populations of tree species of temperate forests, which can be influenced greatly by seed dispersal and mating system (Wright 1943, 1946). Species in which seed dispersal is
limited often form family groups. A typical tree species with low seed dispersal is oak (Quercus), in which seeds are primarily dispersed by gravity. Berg & Hamrick (1995) reported individuals in a turkey oak (Quercus laevis) population as more strongly related at 10 m or less, than at greater distances, most likely owing to limited acorn dispersal. In another example, low, but significant spatial genetic structure at shorter distances is shown in white oak (Quercus petraea and Quercus robur; Streiff et al. 1998, 1999).

By comparison, conifer seed is winged and more widely dispersed by wind. Spatial genetic structure would be expected to be weaker in conifers because the higher gene flow of winged seed could disrupt the building of family structure. For example, the absence of family structure in the presence of high gene flow is suggested by near randomly distributed genotypes of lodgepole pine (Epperson & Allard 1989) and old-growth black spruce (Picea mariana; Knowles 1991). Nonetheless, structuring of conifer genotypes and inbreeding can still occur over time because of limited seed dispersal and mating between relatives, as shown by Epperson & Chung (2001) who provide evidence for weak correlations of alleles at 15 m in old-growth and seedling white pine populations.

Eastern white pine is quite genetically diverse among individual trees and across its geographical range. High heterozygosity is found for allozyme loci (Beaulieu & Simon 1994; Buchert et al. 1997; Rajora et al. 1998; Epperson & Chung 2001) and for DNA loci (Echt et al. 1996; Rajora et al. 2000). Few studies examine how changes in the forest ecosystem affect white pine genetic diversity. In one pair of studies, the genetic effects of harvesting were evaluated for two old-growth white pine populations in Ontario, Canada. Changes in genetic diversity were measured using microsatellite markers (Rajora et al. 2000) and allozyme markers (Buchert et al. 1997). For both studies, allele richness was affected negatively by harvesting, with a 26% decrease in allele richness from preharvest levels, but expected heterozygosity ($H_e$) showed less than a 3% reduction. Changes in diversity between harvested parental populations and naturally regenerating progeny, have not been studied in white pine.

This report is the first microsatellite study quantifying spatial genetic structure within mature populations of eastern white pine. Our objective was to learn more about the mating system of white pine, its spatial structure, and how white pine genetic diversity is affected by changes in the ecosystem. We used Moran’s $I$-coefficient (Moran 1950) to estimate Wright’s neighbourhood size (Wright 1943; Epperson et al. 1999) for evaluating inbreeding of natural populations of white pine and for estimating dispersal distance (Epperson & Chung 2001). In support of our hypothesis that spatial structure would be weak in both populations, but different, spatial autocorrelation analysis shows an absence of patch structure in the second growth population, possibly owing to logging. In comparison, the old growth population shows weak family structure fitting the isolation by distance model predicted by wind dispersal of pollen and seed in a continuous population. By evaluating the spatial genetic structure with equivalent analyses of allozymes and different classes of microsatellites, we demonstrate that marker choice can significantly affect both the magnitude of the structure measured, as well as the indirect measure of dispersal distance.

Materials and methods

Study sites

In 1998, we selected two mature populations of eastern white pine for genetic analysis from Hartwick Pines State Park, Crawford County, Michigan (44°39’ N/84°42’ N) (Epperson & Chung 2001; Marquardt 2002). Old growth (sensu Hunter & White 1997) and second growth populations were evaluated. Hartwick Pines includes 16 ha of virgin old growth established c. 250 years ago following a wind disturbance (Wackerman 1924). Dominant trees were primarily white pine with some red pine (Pinus resinosa). Sampled trees at the old growth site had a mean diameter at breast height and standard deviation of 62 cm ± 12 cm (from 33 to 99 cm). No successful pine regeneration was present in the understory. The 6 ha second-growth population regenerated naturally after logging c. 100 years ago and is located 2 km east of the old growth population. Eastern white pine was the dominant species with red pine also present in the upper canopy. We sampled trees larger than 20 cm in diameter at breast height. This approach allowed us to sample contiguous adult trees on the same scale (i.e. size of area sampled and density) as the old growth site, while avoiding smaller trees not contributing pollen and seed. Mean diameter at breast height of the trees sampled at the second growth population was 48 ± 18 cm (from 22 to 94 cm). Sampled areas are c. 1.2 ha for the old-growth site (100 × 120 m) and second-growth site (110 × 110 m), with densities of 100 adult trees/ha for both populations.

Sampling and marker analysis

We chose sample size and porosity (i.e. proportion of individuals sampled) to optimize statistical analyses of fine-scale spatial structure for a wind-pollinated species with limited seed dispersal (Epperson et al. 1999). Needle and bud tissues were collected using a high powered rifle or pole pruners during late August to November, 1998, from contiguous adult trees, 122 old-growth and 120 second-growth white pine trees, for a total of 242 individuals sampled. Each sampled tree was mapped, and x and y coordinates were used when quantifying spatial structure. Live tissue
was transported on ice to the laboratory at Michigan State University, and stored at −20 °C until DNA extraction.

Genomic DNA was purified from either 20–40 mg bud tissue or 70 mg needle tissue, with Qiagen Plant DNAeasy isolation kits (Qiagen Inc., Valencia, CA, USA). Yields ranged from 20 to 160 µg DNA; stocks were diluted to 2.5 ng/µL in T_{0.0} (10 mM Tris-Cl pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA)) for amplification by the polymerase chain reaction. We surveyed genomic simple sequence repeat variation with seven-microsatellite markers (Rps1b, Rps2, Rps6, Rps39, Rps50, Rps84 and Rps127; Echt et al. 1996). The amplification reaction for each primer pair was conducted separately. Reaction mixtures contained 2 ng/µL DNA template in 10 µL of reaction buffer. The reaction buffer consisted of 20 mM Tris-Cl pH 8.75, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% Triton X-100, 100 µg/mL bovine serum albumin (BSA), 6% sucrose, 0.1 mM cresol red (Routman & Cheverud 1994), 200 µM each dNTP, 200–800 nM each primer, and 0.025 U/µL AmpliTaq Gold DNA Polymerase (PE Biosystems, Foster City, CA, USA). The final Mg concentration was adjusted to 4.5 mM with 25 mM MgCl₂. A touchdown amplification protocol (Echt et al. 1999) with a modified target annealing temperature of 55 °C was performed using PTC-100 or PTC-200 thermocyclers (MJ Research Inc., Watertown, MA, USA). Other minor adjustments to the protocol are detailed in Marquardt (2002). Amplified products were diluted in deionized water, and two to three loci were pooled for multiplexing on polyacrylamide gels. Amplified fragments were resolved using standard denaturing sequencing gels (Burst-Pac 6% sequencing gel; Owl Scientific, Portland, OR, USA), which consisted of 6% acrylamide/bis-acrylamide (19 : 1), 8.3 M urea, and 1X TBE (89 mM Tris, 89 mM Borate, 2 mM EDTA). Resolved fragments were sized by GENESCAN analysis software (PE Applied Biosystems 1996); CXR fluorescent ladder (Promega Corp., Madison, WI, USA) was used as the internal size standard. In addition, 13% of the total single-locus genotypic scores were designated for random re-testing, and there were zero errors of genotyping. Polymerase chain reactions and gel electrophoresis were conducted at the USDA Forest Service, Forestry Sciences Laboratory, Rhinelander, WI, USA.

Genetic diversity and population genetic structure

Six indices were used to measure genetic diversity: observed number of alleles per locus, allele richness (per-locus mean number of alleles), allele frequency (θ), Nei’s (1973) genetic diversity or expected heterozygosity (Hₑ), observed heterozygosity, and the effective number of alleles per locus (uₑ; Kimura & Crow 1964). Excesses or deficits in genetic diversity were measured for each locus by Wright’s (1921, 1922) fixation index (F). These indices were tested for significant differences from zero or Hardy–Weinberg equilibrium by a χ² statistic (Li & Horvitz 1953). Relative divergence of the two populations was measured by Wright’s Fₛₑ statistic (Wright 1943, 1951, 1965). At each locus, a χ² test determined if there were significant differences in allele frequencies between populations (Workman & Niswander 1970). In addition, divergence was measured by Weir’s θ (Weir & Cockerham 1984; Weir 1996). Bootstrapping over loci provided confidence intervals for θ.

Spatial genetic structure

Spatial autocorrelation was measured by calculating Moran’s I-statistic (Sokal & Oden 1978; Cliff & Ord 1981) separately for each allele. Individual genotypes were first converted to allele frequencies (Dewey & Heywood 1988) where 1.0 was assigned to homozygotes for the allele, 0.5 to heterozygotes for the allele and 0.0 to all other genotypes. We measured the Euclidean distance between trees, assigning all pairs or joins of individuals to one of 10 mutually exclusive distance classes. The upper bound of the first distance class was set at 15 m, with upper bounds increased by 10 m for each successive class. Spacing set equal to the inverse of density (Epperson 1990) can be used to estimate the typical distance between nearest-neighbour individuals. To ensure that most pairs of nearest neighbours were included in the particularly important first distance class, we set its upper bound equal to approximately 1.5 times the square root of the inverse of the sample density (Epperson & Chung 2001). The bounds for each of the nine remaining distance classes were chosen so that each included c. 10% of the total joins because this tends to make statistical power similar for different distance classes (Epperson 1990). The set of I-values for the 10 distance classes is a correlogram, and correlograms were calculated for each allele.

After omitting individuals with missing genotypes, Moran’s I indices were calculated for each allele in each distance class. For each I-value, a standard normal deviate determined if the study population differed significantly from the expected value of I = −1/n − 1 under the null hypothesis of no spatial structure (Cliff & Ord 1981), where n is the number of genotypes remaining in the sample. Bonferroni’s approximation was applied to the correlogram for an allele to adjust the experiment-wise error rate (Sakai & Oden 1983). Alleles...
present in less than five copies were excluded from further analysis because these contain insufficient information for spatial analyses. Also excluded were corelograms for the second allele of a biallelic locus because it contributes redundant information. The remaining I-corelograms were averaged over all alleles for each locus separately, and further averaged over all loci. Such averages are similar but not necessarily identical to related measures such as the multivariate approach by Smouse & Peakall (1999; Epperson 2004).

There are correlations between distance classes for the same allele, which may be strong and dependent on the spatial distribution itself. Any test that assesses the significance of an entire spatial corelogram must take into account these correlations (Oden 1984). One example is the Q-test developed by Oden (1984) for the significance of I-corelograms, but the Q-test is conservative and prohibits proper tests for I-corelograms averaged over alleles of a locus. Therefore for locus-by-locus comparisons, additional statistical procedures (described in detail below) can be applied to sets of I indices for a single distance class (Epperson 2004). We focused on the first distance class for two reasons: (1) it was the only distance class in which we found substantial autocorrelations for more than a few alleles; and (2) the first distance class is the best one as a single indicator of global (correlogram) autocorrelation, and it is in fact a remarkably powerful test of global autocorrelation under isolation by distance processes (Oden 1984). Moreover, pairwise statistics also provide an efficient indirect estimator of gene flow.

In addition to correlations between distance classes for an allele, there are also correlations among I-indices for different alleles of a locus for the same distance class, but recently these have been characterized (Epperson 2004). Ignoring the correlations among alleles would under estimate the variance of the average I over alleles of a locus, but a simple correction can be made in most cases. Rather than divide the sum of variances for individual alleles by \( n^2 \), where \( n \) is the number of alleles as would be appropriate if they were independent, a remarkably close correction can be made by dividing by \( n(n - 1) \) (Epperson 2004). We calculated the approximated (corrected) variances of averages over alleles of a locus using this method and further averaged over loci under the assumption that loci are independent (i.e. little or no linkage disequilibrium). Standard normal deviates were constructed to test the null hypothesis of randomly distributed tree genotypes for each locus and for all loci combined (Walter & Epperson 2004).

### Results

#### Genetic diversity

All loci evaluated in the study were polymorphic (Table 1). Two to 14 alleles were detected for the loci assayed for the 242 individual trees sampled, with an average of 7.7 ± 3.9 alleles per locus. The locus with the largest number of alleles (14) was \( \text{Rps50} \). In contrast, the \( \text{Rps127} \) locus had only two alleles. There were a total of 54 alleles for all loci combined. The effective number of alleles ranged from 1.2 (\( \text{Rps1b} \)) to 5.7 (\( \text{Rps50} \)), with an average of 2.4 ± 1.5. Allele sizes ranged from 144 to 213 bp. Locus \( \text{Rps50} \) had the largest range of allele sizes (32 bp) among the seven, microsatellite markers and \( \text{Rps127} \) had the smallest (2 bp). An eighth primer pair (\( \text{Rps54b} \)) was dropped from the final analyses because it amplified the same microsatellite as \( \text{Rps6} \) and contained null alleles.

Diversity measures were similar for the two populations, with 47 and 51 alleles detected in the old growth population and second growth population, respectively. Average respective values for allele richness and the effective number of alleles, are 6.7 ± 3.6 and 2.4 ± 1.5 for the old growth population, and 7.3 ± 3.3 and 2.3 ± 1.4 for the second growth population. Average heterozygosity was moderately high with mean observed heterozygosities 2% to 6% lower than the probabilities expected under Hardy–Weinberg equilibrium. Average respective values for observed and expected heterozygosity are 0.47 ± 0.21 and 0.48 ± 0.20 for the old growth population and 0.46 ± 0.16 and 0.49 ± 0.18 for the second growth population.

#### Population genetic structure

The mean fixation indices are 0.01 ± 0.16 (old growth) and 0.05 ± 0.10 (second growth), indicating little inbreeding in both populations (Table 2). For the old-growth genotype frequencies, 29% differ significantly from Hardy–Weinberg expectations. Locus \( \text{Rps127} \) had an excess of heterozygotes (\( F = -0.23, P = 0.05 \)) while locus \( \text{Rps39} \) showed a heterozygote

<table>
<thead>
<tr>
<th>Rps locus</th>
<th>Repeat sequence</th>
<th>Number of alleles per locus</th>
<th>( n_e )</th>
<th>Allele size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>(AC)(_{11})</td>
<td>8</td>
<td>1.2</td>
<td>193–213</td>
</tr>
<tr>
<td>2</td>
<td>(AC)(_{15})</td>
<td>10</td>
<td>2.2</td>
<td>151–173</td>
</tr>
<tr>
<td>6</td>
<td>(AC)(_{14})</td>
<td>9</td>
<td>2.0</td>
<td>159–186</td>
</tr>
<tr>
<td>39</td>
<td>(AC)(<em>{12})TC(AC)(</em>{3})</td>
<td>5</td>
<td>2.0</td>
<td>167–179</td>
</tr>
<tr>
<td>50</td>
<td>(AC)(_{17})</td>
<td>14</td>
<td>5.7</td>
<td>153–185</td>
</tr>
<tr>
<td>84</td>
<td>(CT)(<em>{10})(AC)(</em>{11})</td>
<td>6</td>
<td>1.8</td>
<td>144–162</td>
</tr>
<tr>
<td>127</td>
<td>(AC)(<em>{10})AT(</em>{5})</td>
<td>2</td>
<td>1.6</td>
<td>192–194</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>7.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>3.9</td>
<td>1.5</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2004 Blackwell Publishing Ltd, Molecular Ecology, 13, 3305–3315
Table 2  Population indices measured for two populations of eastern white pine (Pinus strobus): inbreeding by fixation index (F) and population divergence by $F_{ST}$ and $\theta$.

<table>
<thead>
<tr>
<th>Population</th>
<th>Old growth</th>
<th>Second growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>$F_{ST}$</td>
<td>$\theta$</td>
</tr>
<tr>
<td>Rps1b</td>
<td>0.06</td>
<td>0.16</td>
</tr>
<tr>
<td>Rps2</td>
<td>0.15</td>
<td>-0.02</td>
</tr>
<tr>
<td>Rps6</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Rps39</td>
<td>0.26*</td>
<td>0.17</td>
</tr>
<tr>
<td>Rps50</td>
<td>-0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Rps84</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Rps127</td>
<td>-0.23*</td>
<td>-0.11</td>
</tr>
<tr>
<td>Mean</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>SD</td>
<td>0.16</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*Significant from zero at $P = 0.05$.

deficit ($F$ of 0.26; $P = 0.05$). Second growth population genotype frequencies did not vary significantly from expected values. Although statistically significant, there is little divergence between the two populations (mean $F_{ST} = 0.0054$; mean $\theta = 0.0065$; $P = 0.05$). Differences in allele frequencies between populations (as measured by $F_{ST}$) are significant statistically for five of seven loci ($P = 0.05$).

Spatial genetic structure

Spatial genetic structure within both populations was weak. For each of two populations, Moran’s spatial autocorrelation coefficients ($I$) were calculated for each of 10 distance classes (Table 3; mean indices for individual loci and the overall mean values are shown). The theoretical expected value under the null hypothesis of no spatial structure is $-0.008$. Comparing the overall weighted (by numbers of alleles) mean $I$-values for distance class one, the structure at the old growth population ($I = 0.015$) is 15-fold greater than that for the second growth population ($I = 0.001$). In addition, inferences can be formed about a population’s structural pattern through the shape of its correlogram. The shape or pattern of spatial structure in the old growth population differed from that in the second growth population (Fig. 1). The old growth population had weak positive genetic correlation or similarity between individuals separated by 15 m or less, followed by negative correlations or dissimilarity at longer distances. This spatial pattern was in contrast to the second growth distribution, which appeared completely random.

The correlograms suggest that loci differ in the amount of spatial autocorrelation at the old growth site, seen almost exclusively in the first distance class (< 15 m). To test this further, we calculated standard errors for the

Table 3  Mean spatial autocorrelation coefficients (Moran’s $I$) for microsatellite loci in two mature populations of eastern white pine (Pinus strobus) for 10 distance classes

<table>
<thead>
<tr>
<th>Distance bound*</th>
<th>15</th>
<th>25</th>
<th>35</th>
<th>45</th>
<th>55</th>
<th>65</th>
<th>75</th>
<th>85</th>
<th>95</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rps1b</td>
<td>0.050</td>
<td>-0.010</td>
<td>-0.035</td>
<td>0.050</td>
<td>0.010</td>
<td>0.045</td>
<td>-0.040</td>
<td>-0.065</td>
<td>-0.015</td>
<td></td>
</tr>
<tr>
<td>Rps2</td>
<td>0.033</td>
<td>-0.008</td>
<td>0.010</td>
<td>-0.045</td>
<td>-0.055</td>
<td>0.018</td>
<td>-0.005</td>
<td>0.023</td>
<td>0.060</td>
<td>-0.038</td>
</tr>
<tr>
<td>Rps6</td>
<td>0.003</td>
<td>-0.003</td>
<td>-0.005</td>
<td>-0.015</td>
<td>-0.018</td>
<td>0.008</td>
<td>0.025</td>
<td>-0.058</td>
<td>0.003</td>
<td>-0.013</td>
</tr>
<tr>
<td>Rps39</td>
<td>0.060</td>
<td>-0.010</td>
<td>0.000</td>
<td>-0.023</td>
<td>-0.030</td>
<td>-0.057</td>
<td>-0.010</td>
<td>0.027</td>
<td>0.070</td>
<td>-0.020</td>
</tr>
<tr>
<td>Rps50</td>
<td>-0.010</td>
<td>0.018</td>
<td>-0.042</td>
<td>-0.020</td>
<td>0.000</td>
<td>-0.016</td>
<td>0.006</td>
<td>-0.008</td>
<td>0.018</td>
<td>-0.006</td>
</tr>
<tr>
<td>Rps84</td>
<td>-0.015</td>
<td>-0.020</td>
<td>0.003</td>
<td>-0.033</td>
<td>-0.008</td>
<td>-0.008</td>
<td>-0.005</td>
<td>0.025</td>
<td>0.018</td>
<td>0.005</td>
</tr>
<tr>
<td>Rps127</td>
<td>0.030</td>
<td>-0.050</td>
<td>-0.010</td>
<td>0.040</td>
<td>-0.010</td>
<td>-0.020</td>
<td>0.020</td>
<td>-0.060</td>
<td>-0.050</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.015</td>
<td>-0.006</td>
<td>-0.015</td>
<td>-0.025</td>
<td>-0.014</td>
<td>-0.008</td>
<td>0.007</td>
<td>-0.015</td>
<td>0.019</td>
<td>-0.013</td>
</tr>
<tr>
<td>Second growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rps1b</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.023</td>
<td>-0.027</td>
<td>0.017</td>
<td>-0.013</td>
<td>0.003</td>
<td>-0.010</td>
<td>-0.023</td>
<td>0.010</td>
</tr>
<tr>
<td>Rps2</td>
<td>0.013</td>
<td>0.003</td>
<td>-0.035</td>
<td>-0.030</td>
<td>0.010</td>
<td>0.023</td>
<td>-0.008</td>
<td>-0.045</td>
<td>0.008</td>
<td>-0.003</td>
</tr>
<tr>
<td>Rps6</td>
<td>0.010</td>
<td>0.020</td>
<td>0.025</td>
<td>-0.015</td>
<td>-0.020</td>
<td>-0.003</td>
<td>-0.043</td>
<td>-0.020</td>
<td>-0.043</td>
<td>-0.013</td>
</tr>
<tr>
<td>Rps39</td>
<td>-0.005</td>
<td>-0.023</td>
<td>-0.005</td>
<td>-0.045</td>
<td>0.018</td>
<td>-0.008</td>
<td>0.018</td>
<td>0.000</td>
<td>-0.040</td>
<td>0.010</td>
</tr>
<tr>
<td>Rps50</td>
<td>-0.001</td>
<td>-0.012</td>
<td>-0.003</td>
<td>0.009</td>
<td>0.008</td>
<td>-0.037</td>
<td>-0.016</td>
<td>-0.004</td>
<td>-0.006</td>
<td>-0.030</td>
</tr>
<tr>
<td>Rps84</td>
<td>-0.013</td>
<td>0.033</td>
<td>-0.008</td>
<td>-0.025</td>
<td>-0.030</td>
<td>-0.005</td>
<td>-0.030</td>
<td>0.000</td>
<td>-0.008</td>
<td>0.015</td>
</tr>
<tr>
<td>Rps127</td>
<td>0.030</td>
<td>0.050</td>
<td>0.050</td>
<td>-0.040</td>
<td>0.030</td>
<td>-0.030</td>
<td>-0.060</td>
<td>-0.050</td>
<td>-0.140</td>
<td>0.030</td>
</tr>
<tr>
<td>Mean</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.005</td>
<td>-0.017</td>
<td>0.002</td>
<td>-0.013</td>
<td>-0.015</td>
<td>-0.013</td>
<td>-0.020</td>
<td>-0.006</td>
</tr>
</tbody>
</table>

Expected value of $I = -0.008$.

*Upper distance bound (meters). †Arithmetic mean over alleles. ‡Weighted mean over loci.

© 2004 Blackwell Publishing Ltd, Molecular Ecology, 13, 3305–3315
per-locus mean I-values (i.e. averaged over alleles) for distance class one (Table 4). For the old growth population, although only one locus (Rps39) showed statistical significance, several others (Rps1b, Rps2 and Rps127) have similar positive values and fairly large test-statistic values. The unweighted average (0.022) over all seven loci is highly significant (standard normal deviate, SND = 2.31; P-value = 0.011). The average for Rps1b, Rps2, Rps39 and Rps127 (0.043) is also highly significant (SND = 2.60), whereas the average (~0.007) for the other three loci, Rps6, Rps50 and Rps84, is almost exactly equal to the expected value (~0.008) under the randomization null hypothesis. Moreover, the autocorrelation for the first group (Rps1b, Rps2, Rps39 and Rps127) is statistically greater than the second group (Rps6, Rps50 and Rps84). The difference in means (0.043–0.007 = 0.050) is twice the estimated standard error of the difference (SE = 0.025; SND = 2.00; P = 0.02). For reasons that are detailed in the Discussion, Rps50 and possibly Rps84 may have less structure, and averages excluding these also are statistically significant (Table 4). In contrast, for the second-growth population, none of the per-locus means are significant (results not shown). The unweighted average over loci (0.004) is not statistically different from randomization, and it is significantly lower than the old growth average for Rps1b, Rps2, Rps39 and Rps127 (SND = 1.77, P = 0.04).

### Discussion

**Genetic diversity and heterozygosity**

Heavy logging occurred in the Hartwick region near the beginning of the twentieth century. While the second growth population was logged, the old growth population was left undisturbed. There is little genetic difference between populations, and the diversity measures (allele richness, effective number of alleles, expected heterozygosity and observed heterozygosity) are similar for the two populations. Since logging reduces the population size, it could be possible, under certain circumstances, for logging to decrease the diversity of the regenerating seedlings. A decrease in diversity would depend on many factors and would not necessarily occur every time a forest is logged. It seems reasonable that the logging-induced change in the effective population size was not severe enough to cause a decline in diversity of the second growth population. Many conifers combine several life history traits that promote high genetic diversity: large geographical range, long-lived perennial, late successional species, high chromosome number, sexual reproduction, high fecundity, outcrossing breeding system and wind pollination (Hamrick et al. 1979; Ledig 1998). As predicted, average expected heterozygosity is high for the two populations combined (Hₑ = 0.49) and 4% higher than the observed values (Hₒ = 0.47), suggesting little inbreeding. Similar genetic diversity values were reported by Rajora et al. (2000), for two old-growth white pine populations in Ontario, evaluated at 13 microsatellite loci, including five of the seven loci assayed in this study.

Mean fixation indices (Wright 1965) indicated little inbreeding in both the old growth and second growth populations, as expected for a wind pollinated, outcrossing species. It is worth noting that, on average, for both populations, the mean microsatellite fixation index is 72% lower than the allozyme analysis of the same tree populations (Table 1; Epperson & Chung 2001). Mean fixation indices

### Table 4 Old-growth population (Pinus strobus) single-locus values

<table>
<thead>
<tr>
<th>Locus</th>
<th>Ave. I</th>
<th>I−EI</th>
<th>SE</th>
<th>SND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rps1b</td>
<td>0.050</td>
<td>0.058</td>
<td>0.047</td>
<td>1.23</td>
</tr>
<tr>
<td>Rps2</td>
<td>0.033</td>
<td>0.041</td>
<td>0.027</td>
<td>1.52</td>
</tr>
<tr>
<td>Rps6</td>
<td>0.003</td>
<td>0.011</td>
<td>0.027</td>
<td>0.41</td>
</tr>
<tr>
<td>Rps39</td>
<td>0.060*</td>
<td>0.068</td>
<td>0.033</td>
<td>2.06</td>
</tr>
<tr>
<td>Rps50</td>
<td>−0.010</td>
<td>−0.002</td>
<td>0.024</td>
<td>−0.08</td>
</tr>
<tr>
<td>Rps84</td>
<td>−0.015</td>
<td>−0.007</td>
<td>0.027</td>
<td>−0.26</td>
</tr>
<tr>
<td>Rps127</td>
<td>0.030</td>
<td>0.038</td>
<td>0.047</td>
<td>0.81</td>
</tr>
<tr>
<td>Ave.</td>
<td>0.022*</td>
<td>0.030</td>
<td>0.013</td>
<td>2.31</td>
</tr>
<tr>
<td>Ave−Rps50</td>
<td>0.025*</td>
<td>0.033</td>
<td>0.015</td>
<td>2.27</td>
</tr>
<tr>
<td>Ave−Rps50−Rps84</td>
<td>0.035*</td>
<td>0.043</td>
<td>0.017</td>
<td>2.59</td>
</tr>
<tr>
<td>Rps1b, 2, 39, 127</td>
<td>0.043*</td>
<td>0.051</td>
<td>0.020</td>
<td>2.60</td>
</tr>
<tr>
<td>Rps6, 50, 84</td>
<td>−0.007</td>
<td>0.001</td>
<td>0.015</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*Ave. I, single-locus values for Moran’s I for the first distance class, averaged over all alleles of a locus; I−EI, the excess of the average over the expected value under the null hypothesis; SE, standard error for the average; SND, test statistic for deviation of the average from the null hypothesis, which has an asymptotic standard normal distribution. Also shown are unweighted averages over various sets of loci: all (Ave.); all excluding Rps50 (Ave. − Rps50); all excluding Rps50 and Rps84 (Ave. − Rps50−Rps84); only Rps1b, Rps2, Rps39 and Rps127; and only Rps6, Rps50 and Rps84. Expected value of I = −0.008.

© 2004 Blackwell Publishing Ltd, Molecular Ecology, 13, 3305–3315
are 0.03 and 0.11 for the simple sequence repeat (SSR) and allozyme marker systems, respectively. This reduced inbreeding measured for microsatellite analysis, compared with allozyme analysis, is consistent with similar findings in the Ontario white pine genetic study (Buchert et al. 1997; Rajora et al. 2000). Using reported observed and expected heterozygosity values, \( F = 1 - (H_G / H_E) \), leads to a microsatellite fixation index of 0.141 (Rajora et al. 2000), which is 21% lower than the corresponding allozyme value (0.179; Buchert et al. 1997). On average, less inbreeding occurred at the Michigan sites compared with the Canadian sites. When averaging the mean fixation indices over both marker sets (allozyme and microsatellite), the mean fixation index of 0.07 for Hartwick Pines is 56% lower than the mean value for the Canadian studies \( (F = 0.16) \).

Twenty-nine per cent of the old growth loci differ significantly from Hardy–Weinberg expectations. Inbreeding from mating relatives or as a result of population subdivision from geographical or temporal isolation (Wahlund effect), is an unlikely explanation for the observed heterozygote deficiency for \( R_{ps39} \). In theory, inbreeding affects all loci, but the observed significant deviation from Hardy–Weinberg equilibrium for the old growth population occurs at just 1 of 14 fixation indices, for the populations combined. The presence of a null allele could account for this decrease in heterozygosity. Null alleles result from mutations in the sequence at the priming site; mispriming prevents amplification of the target sequence (Callen et al. 1993). Therefore, alleles would no longer be visible through gel electrophoresis, making it impossible to distinguish a null allele heterozygote from a non-null allele homozygote. Increases in reported homozygosity levels could result in artificially significant inbreeding coefficients because the population would appear to have more homozygotes than expected at Hardy–Weinberg equilibrium. Moreover, null homozygotes are one explanation for the failed amplification reactions \( (R_{ps39}) \) for two trees in the old growth population. An individual with two null alleles would appear to be a failed reaction. The frequency of the \( R_{ps39} \) null allele can be estimated from the observed deficiency of heterozygotes (Brookfield 1996). According to Brookfield (1996), individuals with no visible bands are treated as observations with the null allele frequency \( (r) = (H_E - H_G) / (1 + H_G) \). The frequency of null alleles is estimated to be 0.07 in both populations combined. If assuming the two failed polymerase chain reaction (PCR) reactions were the result of double null alleles, then the observed null homozygote frequency is 0.008 with a null allele frequency of 0.09, which is close to the frequency estimated using Brookfield’s (1996) measure.

The observed population divergence at less than 1% \( (F_{ST} = 0.0054, \theta = 0.0065) \) is negligible and consistent with values reported for the allozyme analysis of the same populations \( (G_{ST} = 0.008; \text{Epperson \& Chung 2001}), \) and for other species with similar wind dispersed seed and pollen. For example, the allozyme-derived mean \( G_{ST} \) values reported for \( Gleditsia trianths \) (Schnabel & Hamrick 1990), and \( Q. laevis \) (Berg & Hamrick 1995) are 0.059 and 0.001, respectively. Because statistical power is very high for this study, the small differences in allele frequencies are statistically significant for 71% of loci. Spatial distance between populations or perhaps timber harvest at the SS, are reasonable explanations for the observed minor differences in allele frequencies between populations.

Spatial genetic structure

Barriers to gene flow are not always physical or temporal; distance itself can be isolating (Wright 1946). Dispersal is distance dependent with individuals proximal to one another related more closely through local inbreeding, if they are not removed by inbreeding depression (Wright 1943). Wright (1946) used the concept of neighbourhood to describe how the isolating effects of distance increase spatial genetic structure through inbreeding. We observed \( c.50 \) alleles for the seven loci analysed in this study, for a total sample size of \( c.6000 \) genes per population sample. Statistical power is high for detecting spatial structure at this level of sampling, with the total number of sample genotypes being the main factor controlling experiment-wise statistical power for spatial analysis (Epperson & Li 1996). We expected the spatial genetic structure for the adults to be weak because white pine pollen and seed are dispersed highly by wind. Although seed dispersal can be much shorter than pollen dispersal, extensive pollen flow or gene migration would limit clustering of related trees from low seed dispersal and the development of neighbourhood structure at short distances. In support of this premise of weak spatial structure, Moran’s \( I \)-analysis indicated weak positive spatial structure at short distances for the old-growth population. The randomly distributed second-growth population suggests that logging decreased spatial structure. The same contrast was observed for the isozyme analysis of the same two adult populations (Epperson & Chung 2001).

The mean positive autocorrelation at 15 m for old growth \( (I = 0.02) \) demonstrates that the first distance class is crucial for detecting spatial structure (Epperson & Chung 2001). Autocorrelation is consistent with values reported for other species with similar wind dispersed pollen and seed, i.e. 0.05 for \( Maclura pomifera \) (Schnabel et al. 1991), 0.06–0.09 for \( G. trianths \) (Schnabel & Hamrick 1990; Schnabel et al. 1991), and 0.05 for \( Q. laevis \) (Berg & Hamrick 1995).

Contrast and spatial structure among markers

Although broadly consistent, the microsatellite spatial autocorrelation value \( (I = 0.02) \) is only 30% of the isozyme
value \((I = 0.06)\) for the same old growth population (Epperson & Chung 2001). It is possible that high mutation rates, such as those found at microsatellite loci, directly decrease spatial structure. Epperson (1990) simulated the effects of random replacement of genes with opposite alleles, e.g. through immigration or mutation, on the spatial correlations of gene frequencies among quadrats. Referring to Fig. 1, Epperson illustrated the effect of gene replacement on spatial structure. Replacement genes introduced at a lower rate \((\mu)\) of 0.001 reduces autocorrelation by 33\%, whereas a faster replacement rate of 0.01 causes a greater decrease in spatial structure of 55\%. Moreover, new simulations by Epperson (forthcoming) showed that by reducing identity by descent (e.g. correlations), a high mutation rate of \(10^{-2}\) causes a direct c. 40\% reduction in the Moran’s \(I\) for converted individual genotypes at the shortest distance class. In contrast, a lower mutation rate of \(10^{-3}\) causes an average of only 7\% reduction in autocorrelation, which would not be noticeable in this study. Since both of these values are within the mutation range of microsatellites, the Moran’s \(I\)-values in Hartwick Pines could have been directly reduced by mutation, especially for the more variable loci such as \(Rps50\).

Our results showed that \(Rps50\) has significantly less spatial structure than the other loci as a set. The same contrast was found in the genotypes of seedlings at the old growth site (Walter & Epperson 2004). In the seedlings, essentially no structure was observed for \(Rps50\), whereas the other assayed microsatellites — \(Rps127\), \(Rps1b\), \(Rps2\), \(Rps39\) and \(Rps34\) (which amplified fragments overlapping with \(Rps6\) but with fewer alleles) — as a set exhibited structure very similar to isozymes analysed in the seedling and adult populations (Epperson & Chung 2001; Walter & Epperson 2004). (It should be noted that the nominal microsatellite fragment sizes for various alleles reported in Walter & Epperson (2004) differ slightly from those listed in Table 1 because of slightly differing gel conditions.) In the adults, if \(Rps50\) is excluded, the average \(I\)-statistic for distance class one increases to 0.025. \(Rps84\) and \(Rps50\) (which were not scored in the seedlings) are also more variable and may have higher mutation rates, and if excluded, the mean \(I\)-statistic further increases to 0.043. This mean \(I\)-value is very similar to that found in the seedlings for loci other than \(Rps50\) \((I = 0.048;\) Walter & Epperson 2004\), and in the smaller number of alleles scored for isozymes in the seedlings \((I = 0.04)\) and adults \((I = 0.06)\) (Epperson & Chung 2001).

It appears that the microsatellite loci with greater numbers of alleles exhibit lesser spatial structure, and there are two potential reasons for this. One explanation is based on the fact that \(Rps6\), \(Rps50\) and \(Rps84\) have many more low-frequency alleles than the other loci (data not shown). Theoretical results have shown that although allele frequency generally has little or no effect on spatial autocorrelation measures, when an allele’s frequency is less than about 0.02–0.05, some reduction in short distance correlation results, of the order of c. 15\% for populations with dispersal similar to \(P. strobus\) (Epperson 2003). Several alleles with such frequencies were retained in the spatial analyses for these loci. However, this would explain only part of the difference, and only for the rarest alleles retained; it does not explain the reductions observed for higher-frequency alleles.

The second explanation for \(Rps6\), \(Rps50\), and \(Rps84\) exhibiting less spatial structure is higher mutation rate, as explained earlier. In particular, \(Rps50\) appears to have higher mutation rates. Although the actual rates of mutation of the microsatellites used in this study are not known, we can approximate the relative differences expected in mutation rate using the well-known relationship of effective number of alleles, \(n_e\), to the effective population size, \(N\), and \(N\mu\) \(= (n_e - 1)/4\), under the equilibrium genetic drift–infinite alleles mutation model (Ewens 1979). For \(Rps50\) in both samples, the average effective number of alleles was 5.7, compared with an average of 1.8 for the remaining seven loci (Table 1). This leads to estimates of \(N\mu\) of approximately six (5.9 times larger for \(Rps50\) (1.175) than for the remaining loci (0.200). In the seedlings sampled at both sites, a similar contrast (6.7 greater estimated \(N\mu\)) was observed between \(Rps50\) and the other microsatellites (Walter & Epperson 2004). The effective population size \(N\) should essentially be the same for all loci, and these results imply that \(Rps50\) mutates at a rate seven times greater than the other SSRs. It is possible that \(Rps50\) mutates at a rate of the order of \(10^{-2}\) and the other SSRs and the isozymes mutate at a rate an order of magnitude lower \((10^{-3})\), or more. Hence, differences in mutation rate could be consistent with our observations.

Estimated dispersal

When using \(I\)-statistics to indirectly infer dispersal levels by comparing them to theoretical values of autocorrelation for different total amounts of dispersal, as measured by Wright’s (1946) neighbourhood size (see Discussion below), it appears to be more appropriate to compare the excess of the average observed \(I\)-value over the expected value \((I - EI;\) Walter & Epperson 2004). Theoretical values are measured in simulated large populations \((n = 10000;\) Epperson et al. 1999), and the expected value under the null hypothesis, \(-1/(n - 1)\), is very near zero. In comparison, our sample sizes, \(n\), are on the order of 100 with an expected value of \(-0.008\). The excess of the average \(I\)-value over all alleles and loci is 0.030, corresponding to a neighbourhood size of c. 230, using line 1 of Table 4 of Epperson et al. (1999). However, it seems reasonable, given the potential biases discussed above, to consider only the group of loci \((Rps1b, Rps2, Rps39\) and \(Rps127)\) that do not have large numbers of
rare alleles or evidence of higher mutation rates. Among these loci, the excess of the average Moran’s I-values is 0.051, which is identical for similar loci in the seedlings ($I = EI = 0.051$; Walter & Epperson 2004). This adjusted I-value corresponds to a much smaller neighbourhood size of c. 100 (discussed in detail below), which is identical to the neighbourhood size estimated in the isozyme study in the adults (Epperson & Chung 2001). It appears that choice of markers, even among SSRs can affect I-statistics strongly as indirect measures of dispersal.

Wright’s (1946) neighbourhood size, or the number of mating individuals ($N$) drawn at random from within a circle of area $4\pi d^2$ and radius $2d$, can be used to estimate total dispersal distance. Hence, referring to the above discussion, the overall average excess of Moran’s I-value for all loci of $c. 0.03$ (Table 4) corresponds to a Wright’s neighbourhood size of approximately 230 trees for the old growth population (Epperson et al. 1999; Table 4, line 1). If only the more reliable microsatellite loci (Rps1b, Rps2, Rps39 and Rps127) are used, the average excess (0.05) corresponds to a more reasonable neighbourhood size of c. 100. Thus, standardized for density, the combined seed and pollen dispersal distance ($\sigma$) can be estimated using the neighbourhood formula for a monoeccious population, $N = 4\pi d^2 / (S. \text{Wright} 1946; J. \text{Wright} 1952)$, leading to a predicted value of approximately $795 \text{m}^2$ for $\sigma^2$ and $28 \text{m}$ for $\sigma$, using a population density ($d$) of 100 trees/ha. Seed dispersal distances can be considerably shorter than pollen dispersal. Crawford (1984) showed that for plants with seed dispersal ($\sigma_2$) and pollen dispersal ($\sigma_2p$), the total parent-offspring corresponds to $\sigma^2 = 1/2\sigma_2p + \sigma_2s$. Typical empirical measures of average dispersal distances for pine pollen range from 17 m (Wright 1952) to 69 m (Wang et al. 1960), and for pine seed from 15 m to 30 m (Epperson & Allard 1989). Estimating dispersal given the total parent variance of $795 \text{m}^2$, our results would correspond (for example) to an average pollen standard deviation of $30 \text{m}$, and an average seed dispersal distance of $15 \text{m}$, which is in agreement with typical empirical estimates.

Conclusions

In summary, gene flow is high both among and within populations. Divergence between the populations is negligible and levels of genetic diversity are similar for the two populations. Spatial distance between populations or timber harvest at the second growth site, are reasonable explanations for the observed minor differences in allele frequencies between populations. Within the old-growth population, autocorrelation fits predicted levels for white pine based on pollen and seed dispersal and density. Autocorrelations for some microsatellites are lower than those for other microsatellites and allozymes, and this appears to be due to their having larger numbers of low frequency alleles and possibly higher mutation rates. The small positive autocorrelation values at short distances observed for the old growth site are consistent with the isolation by distance model for predicted genetic differences in a continuous population, with a Wright’s neighbourhood size of c. 100 individuals. In contrast, the absence of spatial structure at the second growth population suggests that logging removed the natural level of spatial genetic structure.

Acknowledgements

This paper is part of a thesis submitted to Michigan State University in partial fulfilment of requirements for a degree of Master of Science. The authors thank Kenichiro Shimatani, Joe Zeleznik, Ann Stephens and Jon Gregorich for their help in the collecting of sample materials. The United States Department of Agriculture/Forest Service/North Central Station (Research Work Units 4152 and 4155), and the McIntire—Stennis Cooperative Forestry Research Program (project #1774), funded the research.

References


Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics, 49, 725–738.
Wright JW (1952) Pollen dispersion of some forest trees. USDA Northeastern Forest Experiment Station, Paper 46, 1–42.
Wright S (1921) Systems of mating. I. The biometric relations between parent and offspring. Genetics, 6, 111–178.


This study is part of Paula Marquardt’s graduate research studying population and spatial genetics of forest trees using molecular markers. She is also involved in global change research studying the effects of air pollution on tree physiology, with emphasis on the photosynthetic response. Bryan Epperson, Professor at Michigan State University, studies theoretical and statistical aspects of geographical genetics, and uses molecular markers to study the spatial population genetics of trees and other species.