

MICROSATELLITE ANALYSIS OF SPATIAL STRUCTURE AMONG SEEDLINGS IN POPULATIONS OF *PINUS STROBUS* (PINACEAE)¹

ROSEMARIE WALTER AND BRYAN K. EPPERSON²

Department of Forestry, Michigan State University, East Lansing, Michigan 48824 USA

In a detailed analysis of how limited seed dispersal can create spatial structuring of genetic variation, several nuclear microsatellites were assayed in seedlings from two forests of *Pinus strobus*, one old growth (OG) and the other (second site, SS) logged in ca. 1900. By using loci with a large number of alleles and new statistical methods on averaged spatial correlation coefficients, unusually precise estimates of spatial genetic structure were obtained, even though the structure was expected to be very weak. This high precision allowed the spatial patterns to be contrasted across loci and populations. At the OG site, the average spatial correlation coefficient for short distances (<15 m) exceeded its random expected value by 0.035, providing an indirect estimate of ca. 230 for Wright's neighborhood size. The value is similar to that estimated in a previous study of adult trees at OG and probably represents the natural level of spatial structure. A very similar value, 0.030, was obtained for seedlings at SS, despite the fact that unlike OG, genotypes of adults are randomly distributed, a likely result of logging. The results show that a single cycle of limited seed dispersal recreated the natural level of spatial structuring. In addition, one microsatellite, Rps50, had far greater amounts of allele variation, likely implicating it as having a higher mutation rate. The spatial structure of Rps50 also was significantly reduced, in a way that could be consistent with theoretical effects of high mutation rates (up to $\mu = 10^{-3}$). The choice of markers may influence estimates of spatial genetic structure. For example, if Rps50 is omitted the values are nearly doubled to 0.058 and 0.051 for SS and OG, respectively, both indicating a much smaller neighborhood size of ca. 100.

Key words: dispersal; eastern white pine; genetic structure; microsatellite; *Pinus strobus*; population genetics; spatial autocorrelation.

Spatial structure of genetic variation has several important influences on plant population genetics, including its interactions with other population genetic processes such as inbreeding depression and selection. Moreover, spatial distributions may be used to obtain indirect estimates of gene dispersal. Limits to dispersal of individuals or propagules results in non-random spatial distributions, and the strength of spatial structure, as measured for example by spatial autocorrelation statistics, is greater wherever distances of dispersal are more restricted. Repeated generations of limited dispersal causes the buildup of spatial structure in a process generically termed "isolation by distance." However, in plants, especially wind-pollinated outcrossing plants, the roles of pollen vs. seed dispersal can differ markedly. In particular, pollen may disperse much greater distances than does seed. Although some isolation-by-distance models indicate that the amount of spatial structure should be determined only by the weighted sum of dispersal variances for pollen and seed (Crawford, 1984), there can be differences in the roles of the two, especially if either is extreme. For example, reproductive adults could be distributed randomly in space in a given generation, and pollen dispersal may be effectively infinite; hence, the sum of seed and pollen dispersal variances is infinite. The theory then dictates there should be random spatial distribution of genotypes. However, if seed dispersal is somewhat limited, then in the following generation, the seed and seedlings may have considerable spatial structure. The spatial distribution of seedlings per se can have particular importance especially in wind-pollinated long-lived perennials (Neale and Adams, 1985; Yaz-

dani et al., 1985). Under some conditions, such spatial structure in seedlings can persist into the next adult population (Epperson, 2003).

Eastern white pine (*Pinus strobus* L.) is an outcrossing (Beaulieu and Simon, 1995), wind-pollinated forest tree, having wind-dispersed seed. In an earlier study using several isozyme loci (Epperson and Chung, 2001), we obtained estimates of the amount of spatial autocorrelation in two study populations located in northern Michigan. In one natural, old-growth population, weak but consistent structure was indicated in both adults and seedlings. In a second population, which was logged around the turn of the 20th century, the genotypes of adults were nearly randomly distributed (possibly as a direct result of logging), but the seedlings appeared to have spatial autocorrelation similar to that in the natural condition. This situation provided an opportunity to study the effects of a single generation of limited seed dispersal. However, it was not possible to make statistical contrasts in the results, because of an insufficient number of alleles and a lack of available proper statistical tests in the literature.

In this article, we examine the spatial structures in seedlings in detail, using more than 20 alleles. We examine a number of nuclear microsatellite loci, which typically have much larger average numbers of alleles than do isozyme loci (e.g., Rajora et al., 2000). The increase in information contained in spatial distributions with increasing numbers of alleles has been characterized recently (Epperson, 2003). As long as there are more than three or four alleles at a locus, the spatial distribution of each allele represents a nearly independent realization of the processes of limited dispersal. Whereas the isozyme data had 5–8 separate spatially informative allele patterns, the microsatellite loci had 21–24, thus representing a three- to fivefold increase in information. In addition, new results on the correlations of spatial coefficients for different

¹ Manuscript received 23 May 2003; revision accepted 20 November 2003.

This research was supported by a McIntire-Stennis (Project #1774) grant to B. K. E.

² E-mail: Epperson@msu.edu.

alleles (Epperson, 2003) allow proper statistical tests and standard errors to be constructed for multiallele–multilocus averages. It should be possible to obtain very precise estimates of spatial structure, which may allow statistical contrasts among the samples, even though the amounts of spatial autocorrelation are small.

Moreover, microsatellites and isozymes have rarely been studied in the same populations and almost never in studies of spatial structure within populations (an exception is Streiff et al., 1998). Microsatellite loci are appealing markers in part because typically they are hypervariable owing to unusually high rates of mutation. However, they also typically have somewhat different arrays of allele frequencies, often having relatively large numbers of low-frequency alleles. In samples with moderate numbers of sampled individuals, alleles that have very low frequencies (e.g., when an allele occurs in only one sampled individual) are not useful for studies of spatial structure, although they can be useful for other types of studies, such as those directed at parentage assignment. Moreover, some microsatellites mutate at such high rates that we may begin to question if mutation rates can have *direct* influences on spatial structure (Epperson, 1990), in addition to their indirect effects operating through numbers and frequencies of alleles. We examine some possible effects of high mutation rate in our data.

MATERIALS AND METHODS

Sampling—The sampled forests have been described in detail elsewhere (Epperson and Chung, 2001). Briefly, seedlings from two natural populations of eastern white pine (*Pinus strobus*) were sampled in Hartwick Pines State Park near Grayling in northern Michigan. Samples from one population (old growth, OG) were taken from an area of about 1 ha in an undisturbed old growth stand. This forest consists almost exclusively of very large individuals (ca. 35–55 m tall), lacking seedlings more than a year old and smaller, younger individuals less than ca. 35 m in height. The density of adult trees is ~120 trees/ha. One-year-old seedlings were sampled by collecting the one closest to each of the adult trees, which were mapped in the previous study. Originally, in 1998, 90 seedlings were sampled. However, because the seedlings were small and sampled nondestructively, sufficient tissue was available only for 46 of the seedlings. Consequently, in 1999 we sampled 43 new 1-yr-old seedlings to replace those that were lost, by sampling again from near the corresponding adult trees. Casual inspection of several hundred seedlings in the entire sample plot area revealed that nearly all were 1 yr old.

A similar size area was sampled in 1998 from a second population (second site, SS), which was logged approximately 100 yr ago, but undisturbed since. The SS population has regenerated and is well represented in all age classes. Seedlings of all sizes were present, in contrast to the OG site. Each of the sampled seedlings (121) was chosen as the one that was less than 30 cm in height and closest to each adult tree, as reported in Epperson and Chung (2001). Adult trees were considered as those having diameters >20 cm. The density of seedlings was obviously much greater at SS than at OG, although we did not measure the densities. Both populations are not pure stands, and other species include red pine, maple, beech, hemlock at OG and these plus black spruce and balsam fir at SS.

DNA isolation—DNA was isolated using the QIAGEN DNeasy Plant Kit (QIAGEN, Santa Clarita, California, USA). Needle tissue (~70 mg) was homogenized in a FastPrep FP120 instrument (BIO 101, Savant Instruments, Holbrook, New York, USA). A ceramic ball was placed at the bottom of a 2-mL screw cap centrifuge tube, the tissue was added, a ceramic cylinder was placed on top of the tissue, and then 600 μ L of lysis buffer was dispensed. Samples were ground for 45 s at the 4.5 setting. Remaining steps were performed according to the QIAGEN DNeasy protocol. DNA concentrations were measured by fluorometry, and the average yield was 8 μ g/extraction.

Nuclear microsatellite markers—Nuclear microsatellite markers were developed for eastern white pine (*Pinus strobus*) by Echt et al. (1996) and are commercially available as primer pairs from Research Genetics (Huntsville, Alabama, USA). In this study, the following primer pairs were used: Rps1b, Rps2, Rps34, Rps39, Rps50, and Rps127. Forward primers were labeled with one of three different fluorescent phosphoramidite dyes, either 6-FAM, HEX, or TET (Applied Biosystems, Foster City, California, USA). Rps2 was scored only for the OG sample.

The PCR amplification protocol has been described elsewhere (Walter and Epperson, 2001). Multiplexing was carried out by co-loading multiple PCR reactions with different fluorescent dyes in the same lane of the gel. This allowed simultaneous analysis and detection of polymorphic loci, even when they overlap in size. Electrophoresis was performed on a 0.40-mm 6% denaturing polyacrylamide gel (7 mol/L urea) at 2500 V/30 W for 5.5–6 h. The GeneScan sequencer (ABI 373; Applied Biosystems), with a 350-bp internal standard was used to determine the fragment sizes. In addition, we used two lanes with a known external standard on each gel to control for any potential gel-to-gel variations (which in fact did not occur). This procedure resolved fragment size to a single base pair. Repeatability of one-base-pair resolution was further verified by repeating the PCR amplifications and GeneScan gel runs for heterozygous individuals. Alleles are denoted by their nominal sizes in numbers of nucleotides; these should not be considered necessarily as absolute rather as relative sizes.

Statistical analysis—The distribution of allele frequencies among populations was assessed using the θ estimator (Cockerham and Weir, 1993) of F_{ST} . Observed values of θ were compared to the null hypothesis of the absence of differentiation by bootstrapping over loci. Measures of Wright's (1922) inbreeding coefficients F_{IS} were also calculated. All of these statistics and their significance levels were analyzed using the GDA program of Lewis and Zaykin (2001).

Spatial autocorrelation analysis was calculated separately for each allele of each locus (Sokal and Oden, 1978). Alleles that occurred in only a single tree in a sample were considered to be spatially noninformative and omitted from further analysis. For diallelic loci, calculations were done for only one allele because the information provided by the second allele would be completely complementary. For the remaining alleles, spatial autocorrelation analysis proceeded by treating nominal (join-count) correlations of genotypes converted into gene frequencies. Allele frequency values were assigned to genotypes as 1.0, 0.5, or 0.0 for homozygotes, heterozygotes, or no copies of that allele, respectively. For each allele, Moran's I statistics (Sokal and Oden, 1978) was calculated. The individuals were paired as joins (using Euclidean distances) representing one of 10 distance classes. Joins were formulated in a way that formed a continuous network and that ensured adequate numbers of nearest neighbors. Special attention was paid to the first distance class, by choosing its upper limit as ~1.5 times the square root of the inverse of sample density (Epperson, 2003). By doing this, it could be guaranteed that the majority of near-neighbor pairs are included in the first distance class. Spatial autocorrelation coefficients (Moran's I) were calculated for each of the 10 distance classes, separately for both populations. For each coefficient, a two-tailed test of significance was applied, again separately for OG and SS. In addition, a test for significance for an entire set of 10 distance classes for an allele for a population was assessed using a Bonferroni-type approximation for multiple tests. These statistics for spatial structure analysis were all computed using Wartenberg's SAAP program (version 4.3; Wartenberg, 1989). For OG, separate analyses were done for the two years of sampling. In both years, all collected seedlings were 1 yr old. Because seed-cone production at OG is sparse and highly erratic, seedlings collected in 1999 likely came from quite different maternal sources than those collected in 1998.

To obtain summary statistics for the overall amount of spatial structure, we calculated the average, \bar{I} , across alleles and loci for each sample. The average was not weighted, because theoretical results show that alleles should generally have very similar stochastic and statistical variances (Epperson et al., 1999), which was supported by inspection of the estimated variances under the null hypothesis of a random spatial distribution. Significance tests for the

TABLE 1. Number of genotypes scored (*n*), number of alleles, and estimates of fixation index, F_{IS} , within the old growth (OG) and second site (SS) stands, and the θ estimate of differentiation, for polymorphic SSR loci. Values in parentheses indicate the numbers used in spatial autocorrelation analyses. Also shown are the average values of F_{IS} and the overall value of θ .

Locus	OG			SS			θ
	<i>n</i>	No. alleles	F_{IS}	<i>n</i>	No. alleles	F_{IS}	
Rps1b	89	3	-0.033	97 (96)	5 (4)	-0.038	-0.002
Rps2	87	4	0.076	—	—	—	—
Rps34	88	3	0.231*	113	4	0.180*	-0.004
Rps39	88	5	0.036	109	4	0.084	-0.003
Rps50	88 (86)	11 (9)	-0.041*	104 (103)	12 (11)	-0.001	0.005
Rps127	88	2 (1)	-0.010	95	2 (1)	-0.002	0.025
Average/overall			0.043			0.051	0.005

* Statistically significant at the 5% level.

average values were developed by estimating the standard error under the null hypothesis, by the formula:

$$SE(\bar{I}) = \frac{\sqrt{\sum_j [SE(I_j)]^2}}{k} \quad (1)$$

where $SE(I_j)$ is the estimated standard error of I (output by SAAP) for allele j , under the null hypothesis, and k is the total number of alleles included in the average. Because individual I statistics are asymptotically normal (Cliff and Ord, 1981), the average has an approximate normal distribution. Hence, a standard normal deviate was formed for the average values in the customary way, by subtracting the averaged expected values [which is $-1/(n - 1)$, where n is the number of sample genotypes] of per-allele I -statistics and dividing by the standard error of the average from Equation 1. For short distances, spatial structure implies positive autocorrelation, thus we report one-tail probabilities for the averages. Equation 1 ignores the correlations among values; however, we were mindful of these. The value for the second allele of diallelic Rps127 was omitted. The correlations for I statistics for different alleles of a locus were recently characterized (Epperson, 2003). They are generally positive, thus Equation 1 would underestimate the true value. However, the correlations are large only for pairs of common alleles (of the same locus), and thus only a few pairs of alleles (e.g., for Rps39 alleles 171 and 173) are affected. Moreover, correlations among loci can be assumed to be near zero. With the large total number of alleles across loci in this study, Equation 1 should be very close to the true standard error. In addition, the standard errors (SE) from Equation 1 could be used directly to construct 95% confidence intervals (using ± 1.96 times the SE) for the average values, apart from the null hypothesis, because theoretical results indicate that the standard errors of individual I statistics under the null hypothesis are scarcely changed by the presence of spatial structure (Epperson et al., 1999), unless it is much stronger than that observed in this study. Because the correlations of I statistics among alleles also are similarly unaffected by structure (Epperson, 2003), the standard errors on the averages should be essentially unchanged by the presence of spatial structure.

RESULTS

Table 1 shows the numbers of alleles for each locus for the OG and the SS populations. Also shown are the estimates of fixation indices for six (OG site) and five (SS) polymorphic loci and the values of θ between the two populations for five of the loci. In the OG population, genotypes for two loci, Rps34 and Rps50, occurred in ratios that were statistically significantly different from Hardy-Weinberg expectations, in contrast to the SS population in which only locus Rps34 was different. In both populations, homozygotes for Rps34 were in excess. This is likely due to the presence of null alleles, which have been found in other unpublished assays of Rps34 in *Pinus strobus*. The presence of null alleles presumably would have marginal effects on the spatial analyses, because its largest effect is to overestimate the presence of homozygotes for

rare alleles. Rare heterozygotes have values of allele frequency (0.5) that are already far above the mean. Regarding measures of differentiation between the two populations, none of the per-locus θ values nor the overall value of θ were statistically significant.

In the spatial analyses of the 1999 year-old seedlings at OG, there were many large and significant values of Moran's I statistics (Table 2). The amount of autocorrelation varied widely, as might be expected for weak structure and fairly small samples. However, the average I correlogram, over alleles and loci, clearly indicated a weak isolation-by-distance pattern. For the first distance class, the excess (0.060) of the observed average value (0.035) over the expected value (-0.024) under the null hypothesis of a random distribution was large and statistically significant. For the 1998 seedlings, fewer significant values were observed (Table 3), and the averages values were small and none of those for short distances were significant. The average of the correlograms over the two years was significant. Values for the correlogram were 0.012, -0.029, -0.022, -0.022, -0.028, -0.053, -0.042, -0.024, 0.005, and -0.014 for distance classes 1 through 10, respectively. For the most important distance class 1, the excess, 0.035, of the observed value over the expected value (-0.024) is significant at the 2% level. The value for distance class 6 was also significant, and the shape of the correlogram fits an isolation by distance pattern. For completeness, we also spatially analyzed the genotypic data after combining them for both years. Among 250 statistics (not shown) only 7% were nominally significant at the 5% level, none of the per-allele correlograms were significant, nor was the correlogram averaged over alleles and loci. Most of the structure was masked, but some was still evident.

Very similar levels of structure were observed for the SS population (Table 4). In total, the autocorrelation coefficients were in 31 (13%) of 240 instances significantly different from their expected values, and many of the correlograms were statistically significant. Most importantly, for the first distance class, six of 24 (25%) statistics were significant (for three alleles of Rps1b, two alleles of Rps34, and one allele of Rps50), and all of these were positive (i.e., no significant negative values). For correlograms averaged over alleles and loci, the excess 0.030 of the observed value (0.020) over the expected value (-0.010) is significant at the 1% level, and the excess is very close to (and not statistically different from) that found at OG.

DISCUSSION

Weak spatial genetic structure of seedlings was revealed by spatial autocorrelation analysis, generally in accordance with

TABLE 2. Spatial autocorrelation coefficients (Moran's *I*) for year-old seedlings collected in 1999 from the old-growth population of *Pinus strobus* for 10 distance classes.

Locus	Allele	Distance class ^a										<i>P</i> ^b	<i>q</i> ^c
		1 (15 m)	2 (25 m)	3 (35 m)	4 (45 m)	5 (55 m)	6 (65 m)	7 (75 m)	8 (85 m)	9 (95 m)	10 (144 m)		
Rps1b	204	0.40**	-0.09	-0.16*	-0.09	-0.08	-0.03	0.02	-0.06	-0.03	0.05	0.000	0.023
	210	0.14	-0.19*	-0.05	0.18**	-0.15	-0.06	0.00	-0.08	-0.07	0.00	0.028	0.953
	216	-0.05	-0.10	-0.02	0.08	-0.06	-0.02	-0.02	-0.01	-0.03	-0.05	0.506	0.023
Rps2	168	0.13	-0.17	-0.05	0.04	0.03	-0.02	-0.17	-0.09	0.01	0.07	0.576	0.756
	170	0.11	-0.16	-0.03	0.10	0.01	-0.01	-0.16	-0.11	-0.09	0.05	0.671	0.207
	172	-0.06	-0.05	-0.00	-0.00	-0.02	-0.08	-0.02	-0.06	0.38**	-0.15*	0.001	0.024
Rps34	146	0.06	-0.12	-0.18	0.16*	-0.02	-0.15	-0.06	0.11	0.05	-0.06	0.115	0.233
	147	0.14	-0.03	-0.14	0.12*	-0.06	-0.19*	-0.13	0.05	0.15	-0.01	0.340	0.733
	149	-0.05	0.13*	-0.03	-0.07	-0.06	-0.02	-0.02	0.01	-0.04	-0.04	0.103	0.035
Rps39	171	0.09	-0.08	-0.14	0.06	0.13*	0.04	-0.05	-0.22*	-0.13	-0.02	0.227	0.558
	173	0.07	0.03	0.06	0.01	0.06	0.12	-0.03	-0.25**	-0.12	-0.20**	0.092	0.314
	175	-0.01	-0.13	0.12	-0.06	-0.02	-0.08	-0.09	0.10	-0.24*	0.10*	0.288	0.035
	179	0.38**	-0.10	-0.05	-0.05	-0.03	-0.05	-0.04	-0.00	0.01	-0.10	0.000	0.023
	181	0.01	-0.08	-0.04	-0.09	0.13*	0.04	-0.05	-0.13	-0.02	-0.00	0.242	0.070
Rps50	170	-0.21	-0.10	0.13	-0.09	-0.14	0.06	0.02	0.04	0.05	-0.05	0.572	0.256
	172	0.06	-0.02	-0.11	-0.04	0.13*	-0.07	-0.01	-0.09	-0.14	-0.01	0.433	0.280
	174	0.04	0.02	0.03	-0.07	0.11	-0.00	0.15*	-0.18	-0.35**	-0.19*	0.083	0.134
	176	-0.08	-0.06	0.01	-0.08	0.12	-0.09	0.06	-0.06	-0.03	-0.07	0.634	0.122
	180	-0.28*	0.13	-0.07	-0.05	-0.04	-0.10	0.03	0.06	-0.11	0.05	0.349	0.159
	184	-0.07	-0.10	-0.02	-0.11	-0.05	0.19**	-0.03	-0.06	-0.09	0.03	0.015	0.024
Rps127	195	-0.08	0.03	-0.09	-0.06	-0.11	0.03	-0.06	0.06	0.05	0.03	1.000	0.762
Average		0.035	-0.059	-0.040	-0.005	-0.006	-0.023	-0.031	-0.046	-0.038	-0.027		
Average - Expected		0.060	-0.035	-0.015	0.019	0.019	0.001	-0.071	-0.022	-0.013	-0.003		
SE		0.027	0.021	0.020	0.017	0.018	0.019	0.018	0.020	0.027	0.017		
Z score		2.176*	-1.623	-0.772	1.158	1.030	0.054	-0.401	-1.076	-0.490	-0.170		

^aUpper bound for distance class in parentheses.^bSignificance level for entire correlogram.^cAllele frequency among alleles used in spatial autocorrelation analyses.* $P < 0.05$, ** $P < 0.01$.

TABLE 3. Spatial autocorrelation coefficients (Moran's *I* for year-old seedlings collected in 1998 from the old-growth population of *Pinus strobus* for 10 distance classes.

Locus	Allele	Distance class ^a										<i>P</i> ^b	<i>q</i> ^c
		1 (15 m)	2 (25 m)	3 (35 m)	4 (45 m)	5 (55 m)	6 (65 m)	7 (75 m)	8 (85 m)	9 (95 m)	10 (144 m)		
Rps1b	204	-0.03	0.09*	-0.06	-0.06	-0.08	-0.06	-0.06	0.00	0.01	0.05	0.402	0.022
	210	-0.09	-0.06	-0.02	-0.11	0.15*	-0.07	-0.02	0.05	-0.08	0.03	0.219	0.946
	216	-0.06	0.03	-0.09	-0.04	-0.05	-0.01	-0.08	0.14	-0.09	0.04	0.535	0.033
Rps2	166	0.07	-0.08	-0.07	-0.05	-0.05	-0.04	-0.03	-0.04	0.12	0.06	1.000	0.033
	168	-0.15	-0.02	-0.02	0.08	-0.04	-0.04	-0.06	-0.00	-0.00	-0.02	1.000	0.717
	170	-0.15	-0.03	-0.11	0.08	0.06	-0.01	-0.06	-0.05	0.05	-0.02	0.979	0.339
Rps34	146	-0.03	-0.19*	0.02	0.08	-0.01	-0.09	-0.03	0.12	-0.06	-0.01	0.209	0.222
	147	0.13	-0.18*	0.09	0.07	-0.15	-0.17*	-0.15	0.09	-0.03	0.08	0.325	0.689
	149	0.13	0.10	0.08	-0.04	-0.18*	-0.31**	-0.21*	-0.07	0.00	0.19**	0.007	0.089
Rps39	171	0.03	0.15*	-0.03	-0.09	-0.06	-0.19*	-0.08	-0.08	0.11	0.06	0.220	0.611
	173	-0.01	0.06	0.02	-0.08	0.03	-0.25**	-0.03	-0.06	0.10	0.04	0.068	0.367
Rps50	156	0.00	0.02	-0.08	-0.02	0.08	-0.10	0.10	-0.04	0.03	-0.14*	0.404	0.033
	170	0.00	-0.08	-0.06	-0.09	0.05	0.01	0.03	0.10	-0.09	-0.03	1.000	0.233
	172	0.03	-0.05	-0.05	-0.06	-0.08	-0.07	0.13	-0.05	0.24*	-0.09	0.198	0.156
	174	-0.15	-0.04	0.11*	-0.13	0.03	-0.06	-0.05	0.06	0.02	-0.02	0.457	0.189
	176	0.01	0.10	0.07	-0.14	-0.13	-0.12	-0.11	0.02	0.41**	-0.12	0.003	0.100
	180	0.07	0.02	-0.09	0.02	-0.08	-0.01	-0.16	-0.02	-0.07	0.05	0.936	0.189
	184	0.10	0.04	-0.00	-0.06	-0.16*	-0.12	-0.09	0.03	-0.01	0.06	0.412	0.033
	186	-0.04	0.16**	-0.09	0.03	-0.05	-0.10	-0.08	-0.04	-0.06	-0.03	0.059	0.033
Rps127	188	-0.09	-0.03	0.15**	-0.09	-0.05	0.05	-0.06	-0.07	-0.10	-0.04	0.053	0.033
	195	-0.01	0.01	0.12*	-0.12	-0.29**	0.02	0.01	-0.12	0.51**	-0.14	0.000	0.075
Average		-0.011	0.001	-0.005	-0.039	-0.051	-0.083	-0.052	-0.001	0.048	0.000		
Average - Expected		0.011	0.024	0.017	-0.016	-0.028	-0.060	-0.029	0.021	0.071	0.023		
SE		0.021	0.017	0.016	0.017	0.019	0.019	0.022	0.024	0.027	0.016		
Z score		0.523	1.356	1.088	-0.988	-1.476	-3.260**	-1.338	0.898	2.638**	1.440		

^a Upper bound for distance class in parentheses.

^b Significance level for entire correlogram.

^c Allele frequency among alleles used in spatial autocorrelation analyses.

* $P < 0.05$, ** $P < 0.01$.

TABLE 4. Spatial autocorrelation coefficients (Moran's I) in the logged population of *Pinus strobus* for 10 distance classes.

Locus	Allele	Distance class ^a										P^b	q^c	
		1 (15 m)	2 (25 m)	3 (35 m)	4 (45 m)	5 (55 m)	6 (65 m)	7 (75 m)	8 (85 m)	9 (95 m)	10 (129 m)			
Rps1b	196	0.22**	-0.04	-0.05	-0.04	-0.02	-0.03	-0.02	-0.02	-0.03	-0.03	0.01	0.000	0.016
	208	0.12**	-0.03	-0.04	-0.03	-0.04	-0.02	-0.03	-0.00	0.02	0.02	0.02	0.002	0.010
	210	0.14**	0.04	-0.05	-0.02	0.03	-0.04	-0.08*	-0.04	-0.01	-0.01	-0.02	0.010	0.943
	216	-0.02	-0.03	-0.03	-0.01	0.05	-0.01	0.03	-0.06	0.04	0.04	-0.06	0.655	0.031
Rps34	145	-0.01	0.05*	-0.00	-0.00	-0.01	-0.04	-0.04	-0.05	-0.00	-0.00	0.03	0.103	0.018
	146	0.13**	0.04	-0.00	-0.00	-0.08*	0.03	-0.05	0.02	-0.12**	-0.10*	0.006	0.006	0.257
	147	0.12**	0.00	-0.03	-0.03	-0.04	0.04	-0.05	0.06*	-0.11*	-0.03	0.022	0.022	0.686
	149	-0.06	-0.06	0.06*	-0.03	0.00	-0.02	-0.00	0.02	-0.01	-0.02	-0.02	0.118	0.040
Rps39	171	0.00	-0.06	0.02	0.01	-0.06	-0.02	-0.04	0.08**	0.05	0.05	-0.07	0.092	0.615
	173	0.02	-0.07*	0.01	0.02	-0.06	-0.02	-0.05	0.08**	0.08*	0.08*	-0.09*	0.088	0.317
	175	0.00	-0.01	-0.01	-0.02	-0.04	-0.02	0.06**	-0.02	-0.01	-0.01	-0.03	0.039	0.009
	181	-0.05	0.01	-0.01	-0.02	-0.03	-0.03	0.05	0.01	-0.00	-0.00	-0.03	0.626	0.060
Rps50	156	0.10**	-0.04	-0.02	-0.01	-0.02	-0.02	-0.03	-0.01	0.00	0.00	0.01	0.012	0.010
	164	-0.01	-0.03	0.01	-0.03	-0.02	-0.04	0.05*	-0.03	-0.03	-0.03	0.02	0.292	0.019
	170	0.02	0.05	0.07*	0.04	0.00	0.01	-0.03	-0.09*	-0.08	-0.08	-0.20**	0.000	0.194
	172	-0.05	-0.02	0.03	-0.06	0.03	-0.00	0.02	-0.04	-0.02	-0.02	-0.02	0.684	0.316
Rps127	174	0.00	0.00	0.02	0.05	-0.01	0.02	-0.01	-0.04	0.04	0.04	0.07	0.882	0.151
	176	-0.04	0.03	0.06*	-0.02	-0.04	-0.02	0.02	-0.07	-0.05	-0.05	-0.01	0.217	0.063
	178	-0.04	0.07*	-0.02	0.02	-0.03	-0.01	-0.00	-0.06	-0.02	-0.02	-0.05	0.103	0.024
	180	-0.08	0.03	-0.03	-0.03	-0.01	-0.02	0.02	0.04	-0.01	-0.01	-0.04	0.749	0.146
Average	182	-0.06	0.00	-0.00	-0.03	-0.03	-0.02	0.05*	0.00	0.00	0.01	-0.05	0.430	0.034
	186	-0.01	-0.04	-0.03	0.01	0.00	0.01	-0.04	-0.06	0.07*	0.01	0.01	0.352	0.019
	188	0.04	-0.03	-0.02	0.07*	-0.06	-0.06	-0.00	-0.00	-0.00	-0.01	-0.01	0.113	0.024
	195	0.01	-0.03	-0.01	0.13**	0.05	-0.07	0.01	-0.02	-0.02	-0.02	-0.26**	0.000	0.642
Average	0.0204	-0.0071	-0.0029	-0.0054	-0.0183	-0.0167	-0.0067	-0.0125	-0.0088	-0.0088	-0.0442	-0.0442	-0.0442	-0.0442
Average - Expected	0.0302	0.0027	0.0069	0.0044	-0.0085	-0.0069	0.0031	-0.0027	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
SE	0.0090	0.0074	0.0066	0.0067	0.0072	0.0070	0.0072	0.0076	0.0090	0.0090	0.0090	0.0090	0.0090	0.0090
Z score	3.316**	0.370	1.047	0.661	1.173	-0.983	0.438	-0.349	0.116	0.116	0.116	0.116	0.116	0.116

^a Upper bound for distance class in parentheses.^b Significance level for entire correlogram.^c Allele frequency among alleles used in spatial autocorrelation analyses.* $P < 0.05$, ** $P < 0.01$.

other studies of tree species. The values for the shortest distances (distance class one, up to 15 m) at the SS site averaged ca. 0.02. Some alleles exhibit large autocorrelations, others have spatial distributions that were not distinguishable from random. In general terms, this variability is consistent with the theoretical stochasticity for populations with high dispersal. The overall level of autocorrelation is in broad terms consistent with the dispersal characteristics of *Pinus strobus* as well as with the autocorrelations observed for isozymes, averaging 0.04 (Epperson and Chung, 2001). The theoretical degree of autocorrelation for different total amounts of dispersal, as measured by Wright's neighborhood size, was analyzed by Epperson et al. (1999). Those were measured in simulated large populations ($n = 10000$); hence, the expected value under the null hypothesis, $-1/(n - 1)$, was very near zero. In comparing our results, with samples sizes n near 50 or 100, it appears to be more appropriate to compare the observed value minus the expected value. In the SS sample, this amounts to $0.02 - (-0.01)$, or 0.03, corresponding to a neighborhood size of about 230 (Epperson et al., 1999).

Adults of the SS population have essentially random spatial genetic distributions (Epperson and Chung, 2001), and the lack of spatial structure may have been caused by logging at the site prior to the establishment of what is now the adult cohort. The results show that a single generation of limited seed movement has reestablished the natural level of spatial genetic structure among seedlings, because very similar amounts of structure were found for the seedlings at the OG population. At OG, for the year-old seedlings collected in 1999 the average value for distance class 1 was 0.035, larger by 0.060 than the expected value under the null hypothesis of a random distribution. Not as much structure was indicated in the seedlings collected in 1998, but for the average of the two years, a clear pattern of isolation by distance was found. For the first distance class, the average autocorrelation across the two years is 0.012, which is statistically greater than the average expected value, -0.023 (Z score = 2.03), and the excess, 0.035, is similar to that observed at the SS population. There was also some evidence of positive autocorrelation at the second distance class, 15–25 m, for some alleles. This is consistent with the distance that seeds may disperse in *Pinus* species (e.g., Govindaraju, 1988). It is reasonable to expect to find (mostly half-sib) seeds from the same tree concentrated within a "seed shadow" having diameter of ca. 30–50 m. However, the degree of concentration also depends on the degree of overlap of seed shadows of different trees.

An interesting cautionary point is that if the OG seedling data for the two years had been blindly combined and then analyzed, much of the structure would have been masked. This can be related to the low and sporadic production of cones. At the OG site in 1998, we found mature cones on only 13 trees. In 1999, again few trees had cones, and most of these cone-bearing trees did not have mature cones in 1998. The cones that matured in 1998 probably contributed substantially to the (year-old) seedlings collected in 1999, because the only other potential sources could be seeds in any seed bank or seeds from outside the immediate area (although casual inspection of trees in same also indicated few had cones). It is reasonable to expect that similarly low cone production occurred in 1997, probably the year that most 1998 year-old seedlings came from. It was not feasible to analyze seed banks, but the difference between 1998 and 1999 itself suggests that

a seed bank contributed little to the seedlings, because it would tend to make the yearly cohorts similar.

It should also be noted that although the old, mature trees in the OG population have been "undisturbed" and the ecology that created them was natural, fire suppression in the twentieth century has made the present situation at ground level far from natural. There is a very dense understory canopy of mature individuals of beech, maple, and hemlock, all of which are highly shade tolerant. Year-old seedlings of relatively shade-intolerant white pine are suffering near-complete mortality. Similarly, no seedlings, not even 1-yr-olds, of red pine (*Pinus resinosa*) were found despite very large seed production. Similar mortality of *P. strobus* and *P. resinosa* seedlings was observed at the Hartwick OG forest some 20 years ago (Rose, 1984). Low light availability and deep leaf litter at the forest floor are likely responsible for lack of regeneration, and both are largely attributable to the understory of deciduous trees, most of which would not be present if fires had gone through the site. Moreover, the adult trees at OG should be considered an overmature stand, and most of the trees had crowns only near their tops (e.g., for many trees the distance from the ground to their lowest live branch was >30 m). Leaf biomass and net primary productivity are low (Rose, 1984).

Year of establishment varied naturally among the seedlings sampled at the SS population. Only seedlings less than 30 cm in height were collected, but within this class all sizes were well represented. Hence, in this population we were also measuring over multiple years. Because we did not age the seedlings, we cannot separate the sample by yearly cohorts. However, in our observations cone productivity was much greater than at OG, and the presence of autocorrelation at apparently natural levels in the total sample itself suggests that the seed sources were similar over generative years.

The amount of structure greatly differed among loci, and it can be demonstrated that the differences are greater than expected by chance. In particular, almost all alleles of Rps50 have very little or no spatial autocorrelation. Rps50 has by far the greatest numbers of alleles. The allele frequencies of Rps50 may partially explain the cause. Although allele frequency generally has little or no effect on spatial autocorrelation measures, when an allele's frequency is less than about 0.02 to 0.05, some reduction in short distance correlation results, on the order of ca.15% for populations with dispersal levels like those in *Pinus strobus* (Epperson et al., 1999; Epperson, 2003). Several alleles with such frequencies were retained in the spatial analyses especially for Rps50 (Tables 2–4). However, this would explain only part of the difference and only for the rarest alleles.

The fact that Rps50 has higher number of alleles likely means that it has a higher rate of mutation than the other microsatellites. Recent work has shown that differences over a certain range of mutation rates can cause differences in the amount of spatial structure (B. K. Epperson, unpublished manuscript). For species with dispersal levels similar to those for *Pinus* species, more precisely corresponding to a neighborhood size of 115, a mutation rate of 10^{-2} resulted in an average value of 0.0211 for distance class 1 (nearest neighbors), which is 40% smaller than the value of 0.035 produced in the same model without mutation. Large reductions were also generally observed for other distance classes and dispersal levels. In contrast, when mutation rates are less than or equal to 10^{-3} , spatial structure is essentially unchanged from that produced by models with no mutation. Thus the range of 10^{-2} – 10^{-3} is

TABLE 5. Average estimates of Wright's fixation indices (F_{IS}) for progeny and adult populations.

Population	Genetic marker	F_{IS}		Species	Reference
		Adult	Progeny		
Old growth site	Isozyme	0.08	0.051	<i>Pinus strobus</i>	Epperson and Chung (2001)
Second site	Isozyme	0.14	0.006	<i>Pinus strobus</i>	Epperson and Chung (2001)
Schyan	Isozyme	-0.034	0.099	<i>Pinus strobus</i>	Beaulieu and Simon (1995)
Lemieux	Isozyme	0.005	0.024	<i>Pinus strobus</i>	Beaulieu and Simon (1995)
McKay Lake	Isozyme	-0.105	0.011	<i>Pinus monticola</i>	El-Kassaby et al. (1987)

critical for direct effects of mutation on spatial structure in general. Microsatellite markers often have high mutation rates, on the order of 10^{-3} (Jarne and Lagoda, 1996) and up to 10^{-2} (Bruford et al., 1992), in comparison to allozymes which have rates on the order of 10^{-6} (Voelker et al., 1980).

Although the actual rates of mutation of the SSRs used in this study are not known, it appears that the range of relative mutation rates is consistent with a differential mutation effect on spatial structure, using the well-known relationship of effective number of alleles n_e to the effective population size (not to be confused with the neighborhood size) N and μ , $N\mu = (n_e - 1)/4$, under the genetic drift-infinite alleles mutation model (Ewens, 1979). A total of 12 alleles of Rps50 were present in the samples, and the average effective number of alleles was 5.7, compared to an average of 3.8 alleles and average effective number 1.7, for the remaining loci. This leads to estimates of $N\mu$ of approximately seven (6.7) times larger for Rps50 (1.175) than for the remaining loci (0.175). Because the effective population size N should be essentially constant over loci, the seven-fold increase should hold true for the mutation rate for Rps50.

The average Moran's I statistic for distance class 1 is -0.0118 for Rps50 in the SS seedlings, indistinguishable from the expected value, -0.0098 , under the null hypothesis of a random distribution. The average among alleles at all other loci is 0.0477, which is statistically greater than both the average for Rps50 and the expected value under the null hypothesis. Moreover, this average is quite close to that observed for allozymes, 0.04 (Epperson and Chung, 2001). In comparison to theoretical models, the excess [$0.0477 - (-0.0098) = 0.0575$] corresponds to a neighborhood size of ca. 100. This is probably a more accurate estimate than the value of 230 obtained when Rps50 is included. For the OG seedlings sampled in 1999, the average value of Moran's I statistic for distance class 1 is -0.0187 for Rps50, again indistinguishable from the null hypothesis, whereas for all other alleles combined the average has a quite large value of 0.0681. As may be expected, for the seedlings collected in 1998 both the average for Rps50 (-0.0078) and that for all other alleles (-0.0134) were small. Nonetheless, the average for all loci other than Rps50 averaged over the two sample years is 0.0274, which represents an excess of 0.0508 over the average expected value. This value is very similar to the excess, 0.05, observed for isozymes (Epperson and Chung, 2001).

Interestingly, another study of small-scale autocorrelation, that assayed both isozymes and microsatellites, in *Quercus robur* and *Q. petraea*, found more structure for the latter. However, the authors attributed this to statistical noise, especially because of the much smaller numbers of alleles assayed for isozymes relative to microsatellites (Streiff et al., 1998).

The spatial autocorrelations in *Pinus strobus* can be compared to that found in other tree species. Average values for

short distance classes for juvenile populations of 0.052 and 0.06 for *Gleditsia triacanthos* (Schnabel and Hamrick, 1990) and 0.098, 0.063, and 0.032 (small scale intervals: 0–1 m, 0–2 m, and 0–5 m, respectively) for *Quercus laevis* (Berg and Hamrick, 1995) were obtained.

In our study, two loci, Rps34 and Rps50, in the OG stand and also Rps34 in the SS population, deviated significantly from Hardy-Weinberg expectations, but as noted, the deviations for Rps34 are probably due to null alleles. All of the isozymes were in Hardy-Weinberg proportions (Epperson and Chung, 2001). Beaulieu and Simon (1995) found in two filial eastern white pine populations that one of four polymorphic loci deviated. The mean values of the estimates of Wright's fixation indices (F_{IS}) for eastern and western white pine progeny and adult populations are listed in Table 5. The positive values reveal a slight deficit of heterozygotes, indicating some form of inbreeding, and negative values represent a slight excess of heterozygotes. However, this could be due to null alleles in considerable frequency.

LITERATURE CITED

- BEAULIEU, J., AND J.-P. SIMON. 1995. Mating system in natural populations of eastern white pine in Quebec. *Canadian Journal of Forest Research* 25: 1697–1703.
- BERG, E. E., AND J. L. HAMRICK. 1995. Fine-scale genetic structure of a turkey oak forest. *Evolution* 49: 110–120.
- BRUFORD, M. W., O. HANOTTE, J. F. Y. BROOKFIELD, AND T. BURKE. 1992. Multi- and single-locus fingerprinting. In A. R. Hoelzel [ed.], *Molecular analysis of populations: a practical approach*, 225–269. IRL Press, Oxford, UK.
- CLIFF, A. D., AND J. K. ORD. 1981. *Spatial processes: methods and applications*. Pion, London, UK.
- COCKERHAM, C. C., AND B. S. WEIR. 1993. Estimation of gene flow from F-statistics. *Evolution* 47: 855–863.
- CRAWFORD, T. J. 1984. The estimation of neighborhood parameters for plant populations. *Heredity* 52: 273–283.
- ECHT, C. S., P. MAY-MARQUARDT, M. HSEIH, AND R. ZAHORCHAK. 1996. Characterization of microsatellite markers in eastern white pine. *Genome* 39: 1102–1108.
- EL-KASSABY, Y. A., M. D. MEAGHER, J. PARKINSON, AND F. T. PORTLOCK. 1987. Allozyme inheritance, heterozygosity and outcrossing rate among *Pinus monticola* near Ladysmith, British Columbia. *Heredity* 58: 173–181.
- EPPELSON, B. K. 1990. Spatial autocorrelation of genotypes under directional selection. *Genetics* 124: 757–771.
- EPPELSON, B. K. 2003. *Geographical genetics*. Princeton University Press, Princeton, New Jersey, USA.
- EPPELSON, B. K., AND M. G. CHUNG. 2001. Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobus* (Pinaceae). *American Journal of Botany* 88: 1006–1010.
- EPPELSON, B. K., Z. HUANG, AND T.-Q. LI. 1999. Measures of spatial structure in samples of genotypes for multiallelic loci. *Genetical Research, Cambridge University Press* 73: 251–261.
- EWENS, W. J. 1979. *Mathematical population genetics*. Springer, New York, New York, USA.

- GOVINDARAJU, D. R. 1988. Life histories, neighbourhood sizes, and variance structure in some North American conifers. *Biological Journal of the Linnean Society* 35: 69–78.
- JARNE, P., AND P. J. L. LAGODA. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424–429.
- LEWIS, P. O., AND D. ZAYKIN. 2001. Genetic data analysis: computer program for the analysis of allelic data, version 1.0 (d16c). Computer program distributed by the authors, website: <http://lewis.eeb.uconn.edu/lewishome/software.html>.
- NEALE, D. B., AND W. T. ADAMS. 1985. The mating system in natural and shelterwood stands of Douglas-fir. *Theoretical and Applied Genetics* 71: 201–207.
- RAJORA, O. P., M. H. RAHMAN, G. P. BUCHERT, AND B. P. DANCİK. 2000. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology* 9: 339–348.
- ROSE, W. M. 1984. Biomass, net primary production and successional dynamics of a virgin white pine (*Pinus strobus*) stand in northern Michigan. Ph.D. dissertation, Michigan State University, East Lansing, Michigan, USA.
- SCHNABEL, A., AND J. L. HAMRICK. 1990. Organization of genetic diversity within and among populations of *Gleditsia triacanthos* (Leguminosae). *American Journal of Botany* 77: 1060–1069.
- SOKAL, R. R., AND N. L. ODEN. 1978. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society* 10: 199–228.
- STREIFF, R., T. LABBE, R. BACILIERI, H. STEINKELLNER, J. GLÖSSL, AND A. KREMER. 1998. Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology* 7: 317–328.
- VOELKER, R. A., H. E. SCHAFFER, AND T. MUKAI. 1980. Spontaneous allozyme mutations in *Drosophila melanogaster*: rate of occurrence and nature of mutants. *Genetics* 94: 961–968.
- WALTER, R., AND B. K. EPPERSON. 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.
- WARTENBERG, D. E. 1989. SAAP. A spatial autocorrelation analysis program, version 4.3. Exeter Software, Setauket, New York, USA.
- WRIGHT, S. 1922. Coefficients of inbreeding and relationship. *American Naturalist* 56: 330–338.
- YAZDANI, R., D. LINGREN, AND D. RUDIN. 1985. Gene dispersion and selfing frequency in a seed tree stand of *Pinus sylvestris* (L.). In H. R. Gregorius [ed.], Population genetics in forestry, 1139–1154. Lecture Notes in Biomathematics, vol. 60. Springer-Verlag, Berlin, Germany.