Spatial Autocorrelation Analysis of the Distribution of Genotypes Within Populations of Lodgepole Pine

B. K. Epperson[1] and R. W. Allard

Department of Genetics, University of California, Davis, California 95616

Manuscript received March 21, 1988
Accepted for publication November 5, 1988

ABSTRACT

Spatial autocorrelation analyses of point samples within two populations of lodgepole pine (*Pinus contorta* ssp. *latifolia*) indicate that single-locus mature tree and pollen genotypes are distributed in a nearly random fashion for most of the allozyme loci assayed. This lack of structure in the distributions of most genotypes is consistent with outcrossing rates that are very nearly 1.0 and with estimates indicating that both pollen and seed are dispersed over long distances in lodgepole pine. However, spatial autocorrelation of genotypes for a few loci suggests that genotypes at these loci may be under natural selection.

THE structure of spatial distributions of genotypes in populations is an integral part of population genetic processes. Structure interacts with a number of factors, including: (1) microenvironmental heterogeneity that favors different genotypes in different areas (BRADSHAW 1984); (2) mortality due to stochastic events, not directly dependent on individual genotypes (*e.g.*, due to destruction of populational areas) (WRIGHT 1978); and (3) mating systems featuring limited dispersal of seed or pollen, and hence excesses of some matings due to proximity. Such interactions can encourage "an effective, continuing process of trial and error, among possible [genic] interaction systems" and thus when combined with interdemic selection increase the long-term evolutionary potential of populations (WRIGHT 1978).

Because structure is influenced by gene flow and natural selection, analyses of spatial patterns of genetic variation can be used to study these factors (SLATKIN 1987; BRADSHAW 1984). One particularly powerful method is the "multivariable approach" of SOKAL and WARTENBERG (1981), which compares the spatial patterns among multiple loci. Deviations for measures for one locus may imply selection at that locus because all neutral loci should have similar spatial patterns. Examples illustrating the utility of multivariable approaches include the study of large-scale patterns by SOKAL, SMOUSE and NEEL (1986) and the study of fine scale patterns within each of several populations by EPPERSON and CLEGG (1986).

Studies of spatial structure itself are also important because structure affects empirical estimates of many population genetic parameters (EPPERSON 1989). For example, estimates of the rate of outcrossing can be biased by assortative matings due to mating by proximity and positive autocorrelations of genotypes (ENNOS and CLEGG 1982; SHAW, KAHLER and ALLARD 1981; BIJLSMA, ALLARD and KAHLER 1986; RITLAND 1985).

In this paper, we report the results of an investigation of spatial distributions of genotypes for fourteen allozyme loci within two populations of lodgepole pine (*Pinus contorta* ssp. *latifolia*). Both of the sample sites are located within large, continuous, outcrossing populations (EPPERSON and ALLARD 1984); also pollen and seed dispersal of lodgepole pine are over relatively long distances compared to previously studied plant populations (reviews in LEVIN 1981; and CLEGG and EPPERSON 1985). Thus, the present analysis of structure in lodgepole pine populations extends the range of neighborhood sizes (WRIGHT 1943) of populations studied for small-scale structure. Measures of single-locus sample distributions are used to investigate the effects of natural selection. Moreover, single- and multiple-locus data address the common view that there are substantial familial relationships in the spatial distributions of genotypes in populations of conifers (LIBBY, STETTLER and SEITZ 1969; WRIGHT 1962).

We employed spatial autocorrelation statistics for nominal (genotypic) data (SOKAL and ODEN 1978a) in analyzing the distributions of genotypes in the present study. These statistics, which measure the statistical dependence of the genotypes at each location on the genotypes at other locations, have been used previously to study genotypic distributions by SOKAL and ODEN (1978b) and EPPERSON and CLEGG (1986). Spatial autocorrelation measures do not require the neutrality and other assumptions that are necessary for measures of genetic identity or kinship (EPPERSON 1989).

[1] Present address: Department of Botany and Plant Sciences, University of California, Riverside, California 92521.

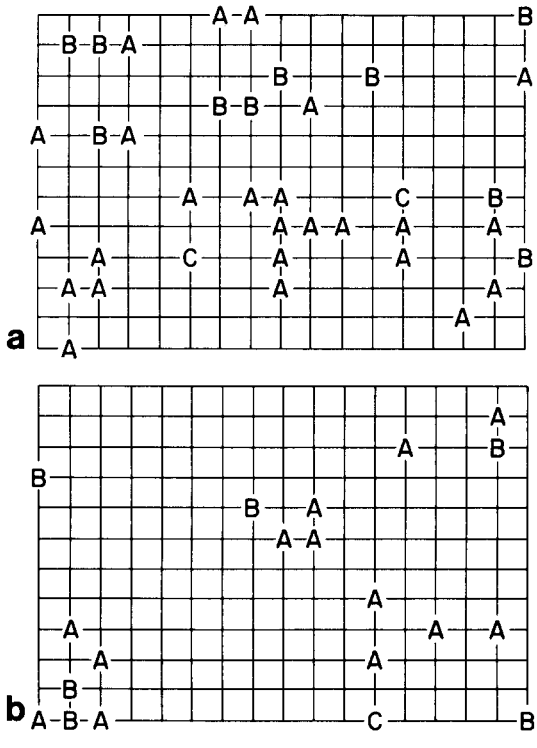

FIGURE 1.—Distributions of rare genotypes of locus *GOT I* in the Scar Mountain lattice of sample points. a, Heterozygotes for rare alleles in the sample of maternal (tree) genotypes. b, Rare pollen types in the sample of pollen genotypes. Letters *A*, *B* and *C*, denote the rare alleles *1A*, *1B* and *1C*, respectively.

MATERIALS AND METHODS

Cones were collected from trees at two sites, Scar Mountain and Indian Creek, located about 7 miles apart in the Colville National Forest in the northeastern part of Washington state. Each site is occupied by dense (*ca.* 1000 trees per acre), even-aged (about 50 yr), and monospecific lodgepole pine forest. Both sites are located within an undisturbed continuous and densely forested area of at least 1 square mile; one site is located near the lower and the other near the upper altitudinal limit for lodgepole pine in the region.

The sample trees at each site were chosen by proximity to the points of a regularly spaced lattice with a separation of 50 feet between nearest-neighbor sample points (Figure 1). The dimensions of the lattices were 12 × 17 and 14 × 14 for the Scar Mountain and Indian Creek sites, respectively. This resulted in a sampling intensity of approximately 1 out of every 50 trees within the sample lattice areas at each site. Procedures for seed treatment and electrophoresis are reported elsewhere (EPPERSON and ALLARD 1987). Briefly, the tree genotypes for 14 allozyme loci were inferred from the electrophoretic phenotypes of at least seven (haploid) megagametophytes of seeds from each tree. Seven of the allozymes (*GOT I, ACO, ACPH, ADH, 6PGD I* and *II*, and *IDH*) were also scored in one seed embryo per tree. The haplotype of the fertilizing pollen was inferred by comparing the megagametophyte (maternal contribution to embryo genotype) and embryo phenotypes. For each allozyme, the electromorphs (up to eight) were combined into three allelic classes in the maternal data, but combinations were sometimes different in the pollen data (EPPERSON 1983). Thus, the data set consists of 201 fourteen-locus tree genotypes and seven-locus pollen types with recorded locations within the Scar Mountain site, and 195 fourteen-locus tree genotypes and seven-locus pollen types within the Indian Creek site (seed collected from a few trees at each site were inviable). In the spatial autocorrelation analyses, the few missing genotypes were replaced with the most common genotype for each locus; this should have negligible effects on the estimates which were made from the data. Nearly all ovules are fertilized by outcross pollen in both populations (EPPERSON and ALLARD 1984); thus the pollen data are essentially samples of the outcross pollen genotypes for seven loci at every sample point.

Tests for nonrandom distributions of the sample genotypes were done for each locus and for each type of data, maternal or pollen, using spatial autocorrelation statistics for nominal data. The number of times, n_i, each genotype i occurs in a sample of site n locations was recorded for a locus and all pairs of points in a sample were used in computing the statistics. The pairs of individuals sampled [total number of pairs, $n\ (n-1)$] were also classified according to the Euclidian distance of separation, d, on the sample lattice, into multiples, k, of the distance (50 feet) separating nearest neighbor locations. Then, for each distance class k, the observed number of the appropriate subset of pairs or joins of points (individuals) for genotypes i and j were counted for all i and j, using a FORTRAN program. Thus, $n_{ij}\ (k)$ is the number joins or of pairs of individuals observed with genotypes i and j, where these individuals are separated by a distance $\geq k - 0.5$, but $< k + 0.5$ units ($50k - 25$ to $50k + 25$ feet). Estimates of the expected number, $u_{ij}\ (k)$, and standard error, $SE_{ij}(k)$, of joins between genotypes i and j, for each distance class k, were calculated under the null hypothesis of nonfree sampling (sampling without replacement) from a random distribution of sample genotypes. This hypothesis is the appropriate one for tests for spatial associations of nominal data (CLIFF and ORD 1981), and here it means that the locations of sample genotypes are randomized. Thus, (suppressing k), $u_{ij} = Wn_in_j/n(n-1)$ for joins between unlike genotypes; and $u_{ii} = Wn_i\ (n_i - 1)/2n(n-1)$ for joins between identical genotypes, where W is twice the total number of joins for the distance class. The test statistic of the null hypothesis, $SND_{ij} = (n_{ij} - u_{ij})/SE_{ij}$, has an asymptotic standard normal distribution (SOKAL and ODEN 1978a) and hence values exceeding 1.96, 2.58 and 3.27 are significant at probability levels 0.05, 0.01 and 0.001. Distance class one ($0.5 < d < 1.5$ units, *i.e.*, distances of 25 to 75 feet), which includes not only strict nearest neighbors (*i.e.*, $d = 1.0$ unit, distance = 50 feet) but also diagonal neighbors ($d = 1.41$, *ca.* 70 feet), proved to be by far the most informative; consequently, we here present data for only this distance class for some of the results.

Test statistics that compare the observed *total* number of joins between unlike genotypes for a given locus and distance class to the estimated expected numbers and variances, assuming sampling without replacement from a randomized distribution, were also important (SOKAL and ODEN 1978a). In still additional analyses, multilocus genotypes were treated as the nominal types, and analogous expected values, variances, and test statistics on joins between multilocus types were calculated, based on the frequencies of the multilocus types in the samples. This method ignores linkage disequilibrium; however, for most pairs of loci, linkage disequilibrium is near zero in both the maternal data and pollen data (EPPERSON and ALLARD 1987).

Tables 1 and 2 list frequencies of single-locus genotypes of the maternal trees and allele frequencies in the pollen samples, respectively. Frequencies are important because frequency may affect the power of the statistics.

TABLE 1

Single locus genotypic frequencies in the maternal samples

Locus	Genotype 11	12	13	22	23	33
A. Scar Mountain (201 trees)						
GOT I[a]	0.000	0.050	0.134	0.045	0.279	0.493
GOT II	0.000	0.040	0.000	0.886	0.075	0.000
ACO	0.030	0.189	0.059	0.418	0.239	0.065
ADH	0.368	0.055	0.423	0.000	0.015	0.139
ACPH	0.149	0.164	0.289	0.025	0.164	0.209
6PGD I	0.886	0.020	0.035	0.000	0.010	0.000
6PGD II	0.945	0.025	0.025	0.000	0.005	0.000
IDH	0.000	0.010		0.990		
PGI II	0.000	0.109	0.000	0.866	0.025	0.000
PGI I	0.000	0.010	0.000	0.965	0.025	0.000
ALAP I	0.030	0.080	0.075	0.234	0.408	0.174
ALAP II	0.925	0.015	0.055	0.000	0.000	0.005
PER I	0.030	0.169	0.005	0.721	0.065	0.010
PER II	0.020	0.079	0.015	0.462	0.343	0.079
B. Indian Creek (195 trees)						
GOT I[a]	0.000	0.031	0.097	0.051	0.246	0.574
GOT II	0.000	0.062	0.005	0.831	0.097	0.005
ACO	0.010	0.149	0.077	0.451	0.267	0.046
ADH	0.292	0.026	0.508	0.000	0.035	0.138
ACPH	0.149	0.139	0.272	0.032	0.210	0.200
6PGD I	0.872	0.108	0.020	0.000	0.000	0.000
6PGD II	0.985	0.005	0.010	0.000	0.000	0.000
IDH	0.000	0.051		0.949		
PGI II	0.005	0.026	0.000	0.954	0.015	0.000
PGI I	0.000	0.000	0.000	0.979	0.020	0.000
ALAP I	0.021	0.205		0.744		
ALAP II	0.908	0.046	0.046	0.000	0.000	0.000
PER I	0.010	0.144	0.005	0.831	0.010	0.000
PER II	0.015	0.041	0.010	0.544	0.318	0.072

[a] For locus *GOT I* the composite allele *1* includes the three electromorphs, *1A*, *1B*, and *1C*.

TABLE 2

Allele frequencies in the pollen samples

Locus	Allele[a] 1	2	3	4
A. Scar Mountain				
GOT I	0.064	0.902	0.034	
ACO	0.064	0.637	0.235	0.064
ADH	0.544	0.020	0.426	0.010
ACPH	0.441	0.206	0.353	
6PGD I	0.931	0.049	0.020	
6PGD II	0.975	0.025		
IDH	0.010	0.990		
B. Indian Creek				
GOT I	0.046	0.934	0.020	
ACO	0.102	0.617	0.250	0.031
ADH	0.495	0.026	0.469	0.010
ACPH	0.444	0.200	0.357	
6PGD I	0.939	0.041	0.020	
6PGD II	0.985	0.015		
IDH	0.046	0.954		

[a] For locus *GOT I*, the table heading genotypes "*1*," "*2*," and "*3*" are for alleles *1A*, *2* plus *3*, and *1B* plus *1C*, respectively.

RESULTS

Spatial autocorrelation analysis of tree genotypes: Counts of each type of join (combination of genotypes at a single locus) for each locus separately, for each distance class of separation, were tested for significant deviation from random expectations by calculating a Standard Normal Deviate (SND) test statistic. For each distance class, there were a maximum of 21 SNDs for a locus with six genotypes and 184 and 158 for all loci combined for the Scar Mountain and Indian Creek samples, respectively. Clearly, with 1840 and 1550 SND statistics for distance classes 1 through 10 there may be many cases of spurious significance. However, stronger weight may be applied to statistics for $k = 1$ because most forms of population structure are likely to exhibit greater autocorrelation at $k = 1$ than for $k > 1$.

For distance class one, only 12 test statistics among a total of 184 were significant (at the 0.05 level) in the Scar Mountain sample, and only 11 of 158 in the Indian Creek sample for all genotype combinations for all loci combined. Because so few of these statistics were significant, they are not listed. [We note that very similar statistical values were obtained for tests based on a subset of the pairs of data points in distance class one, namely, pairs of points that are strict nearest neighbors (EPPERSON 1983).] One conservative method for distinguishing structure-caused statistical significance from spurious significance among the test statistics is the SIDAK procedure (SIDAK 1967). In this procedure, a modified Bonferroni probability level α' for individual tests is calculated by $1-(1-\alpha)^{1/l}$ where α is set as the combined or experiment-wise Type I error rate (here $\alpha = 0.05$) and l is the number of statistics under simultaneous consideration. None of the individual SNDs exceeded the critical value corresponding to α' when l is taken as the total number of statistics for $k = 1$ in each sample.

It is also informative to consider the SNDs for $k = 1$ separately for each locus, because genotypes of different loci may not have the same structure. Differences are predicted in particular if loci are subject to different forms of selection. The number of SNDs for $k = 1$ for each locus ranged from 3 to 21. For the Scar Mountain sample, one SND exceeded the per locus SIDAK critical value (values read from ROHLF and SOKAL 1981); joins for 23 × 23 genotypes of *PER I* were in excess (SND = 3.35, SIDAK critical value is 3.02). This suggests that these rare heterozygotes (Table 1A) are clustered. For the Indian Creek sample, none of the SNDs exceeded the SIDAK critical values. It is worth noting, however, that 3 of the 11 nominally significant SNDs for this sample involved large values for *GOT I* (−2.60, 2.02 and 2.41 for joins 22 × 33, 23 × 23 and 33 × 33). All of these suggest positive autocorrelations of *GOT I* genotypes. None

of the other loci showed convincing evidence of structure in either sample except the *PER I* locus, and possibly *PER II* in the Scar Mountain sample, on a similar basis.

Combining alleles into classes for a few loci might have led to some loss of resolution in the above analyses. Negative autocorrelation is expected to develop between genotypes having different alleles when gene flow is limited or microenvironmental differentiation is relatively coarse-grained; thus, combining counts of joins between different genotypes with joins between truly like genotypes, lowers the power of the test statistics under positive autocorrelation alternative hypotheses. The bias is particularly noticeable for loci *GOT I* and *PER II*; for the other loci there was either no combining of electromorphs, or all except one of the electromorphs which were combined into a class were very rare (frequency less than 0.02, 0.04 in the case of one electromorph of *ACO*). For *PER II* (a cathodal migrating isozyme), eight electromorphs were combined into three classes, each containing one major allele and one or two rare alleles which had similar electrophoretic migration.

For *GOT I*, composite allele 1 included three rare electromorphs which were easily distinguished on gels, alleles here denoted *1A*, *1B* and *1C*. Spatial autocorrelation statistics calculated for heterozygotes for these alleles (numbers of heterozygotes for alleles *1A*, *1B*, *1C* were 24, 10, 2 in the Scar Mountain samples and 21, 3, 1 in the Indian Creek sample; there were no homozygotes) revealed correlations that were not detected in calculations based on heterozygotes for the composite allele *1*. In particular, correlations were positive among heterozygotes for allele *1A* for $k = 1$ (SNDs = 1.65 and 1.30 in the two samples, respectively). The SND for joins between heterozygotes for allele *1B* was 1.18 in the Scar Mountain sample. These statistics are consistent with much stronger evidence for clusters of rare alleles of *GOT I* found in the pollen data. Statistics for joins between like heterozygotes for alleles *1B* or *1C* in the Indian Creek sample, and for allele *1C* in the Scar Mountain sample, were small but positive; however, each of these heterozygotes were rare so that sampling errors may have affected the estimates for these genotypes. In addition, recall that the common genotypes *23* and *33* were each autocorrelated in the Indian Creek sample.

Correlations at longer distances, $k = 2$ to $k = 10$ were not large for pairs of genotypes for any loci; and most of the values that were significant at the experiment-wise error rate were for $k = 2$ for *GOT I* in the Indian Creek sample. Most of the significant test statistics for $k = 3$ and higher distance classes involved excesses of joins between rare genotypes. Such excesses may often result from the effects of chance sampling because the expected value of the number of joins is small (often less than one), and one or two joins can cause nominal statistical significance (see also CLIFF and ORD 1981). In addition, no obvious trends were found in the statistics for similar larger distance classes.

Test statistics for the total number of joins between unlike genotypes for a locus also indicated that genotypes are nearly randomly distributed for most loci. For the first distance class, only the SND for *GOT I* in the Indian Creek sample was significant at either the usual individual error rate ($\alpha = 0.05$) or the SIDAK error rate for 14 test statistics (one for each locus). This statistic (SND = −3.05) indicated an overall clustering of genotypes for *GOT I* at the Indian Creek site. Test statistics for the total number of joins between unlike genotypes for $k = 2$ to $k = 10$ were only rarely significant. In models with severely limited gene flow, it is expected that the number of joins between unlike genotypes will be in deficit for short distances and increase as distances between individuals increase. In summary, the above results indicate that there were no large patches of individuals with similar genotypes in the areas sampled, and hence that little if any large-scale genetic structuring has developed in either the Scar Mountain or the Indian Creek population.

Spatial autocorrelation analysis of pollen genotypes: The SNDs for joins among single locus pollen genotypes are near zero (Tables 3 and 4) for most loci, indicating that pollen genotypes were randomly distributed in both populations. Only three values out of 44 were significant at the usual individual test error rate of $\alpha = 0.05$ for distance class one in each sample. All three significant pairwise values in the Scar Mountain sample were for *GOT I* pollen genotypes. Two of these SNDs indicated negative correlations of the rare allele *1A*, or *1B* plus *1C*, with the two common alleles (alleles *2* plus *3* using now the maternal classification) for $k = 1$. These were also significant at the per locus SIDAK combined error rate (α'), with an experiment-wise or combined error rate of $\alpha = 0.05$. One of these was significant for the SIDAK value for all 44 statistics. Moreover, these correlations were significant up to $k = 3$. Further, the SNDs for the total number of joins between unlike pollen types separated by short distances are significant (at the individual error rate of $\alpha = 0.05$) only for locus *GOT I* in the Scar Mountain sample (Table 4), although this was not significant for α' for 14 loci. There was little evidence for structure in the results for other loci in both samples. These results strongly suggest that there are aggregations of rare alleles of *GOT I* in the pollen cloud. There is little evidence of structure for loci other than *GOT I* in the Scar Mountain sample. However, this does not immediately discount the possibility that there is structure smaller than 50 feet in scale, and/or that there is

TABLE 3

Spatial autocorrelation statistics for the number of joins for distance class one between pairs of pollen genotypes

Locus	Pairs of genotypes									
	1 × 1	*1 × 2*	*1 × 3*	*1 × 4*	*2 × 2*	*2 × 3*	*2 × 4*	*3 × 3*	*3 × 4*	*4 × 4*
A. Scar Mountain sample										
GOT I[a]	−0.80	2.81**	−1.38		0.31	−3.40***		2.46*		
ACO	−0.48	−1.56	0.72	0.45	−0.57	−0.22	0.12	0.42	0.23	0.16
ADH	−0.41	1.75	−0.30	0.54	−0.47	−0.45	−0.54	0.25	0.42	−0.19
ACPH	0.58	0.17	−1.14		−0.09	0.27		0.68		
6PGD I	−1.55	1.91	0.40		−1.32	−0.36		−0.47		
6PGD II	−0.17	−0.04			1.12					
IDH	−0.19	0.86			−0.86					
B. Indian Creek sample										
GOT I[a]	−0.29	0.61	1.69		−0.48	0.04		−1.58		
ACO	−0.40	0.45	−0.39	0.30	0.62	−0.17	−1.23	−0.41	0.76	0.63
ADH	0.85	1.86	−1.40	−1.48	1.08	−1.19	1.07	0.78	2.04*	−0.20
ACPH	0.59	0.62	0.12		−2.09*	−0.41		0.19		
6PGD I	−1.58	0.46	1.01		2.03*	0.79		−0.48		
6PGH II	−1.08	1.10			−0.34					
IDH	−0.29	1.86			−1.89					

*, **, *** Significant at the 5%, 1%, and 0.1% levels, respectively (two-tailed test).
[a] For locus *GOT I* the table heading genotypes "*1*," "*2*," and "*3*" are for alleles *1A*, *2* plus *3*, and *1B* plus *1C*, respectively.

TABLE 4

Spatial autocorrelation statistics for the total number of joins between unlike pollen genotypes as a function of distance class

Locus	Distance Class									
	1	2	3	4	5	6	7	8	9	10
A. Scar Mountain sample										
GOT I	−2.53*	−2.02*	−1.40	−0.99	−0.97	−0.71	−0.64	−0.53	−1.16	−2.03*
ACO	−0.60	−0.28	0.90	1.86	0.69	0.30	0.05	−0.33	−0.27	0.56
ADH	0.14	−0.54	1.01	0.71	−0.48	−0.35	0.29	1.30	−1.33	−0.49
ACPH	−0.72	0.24	−1.16	−0.65	−1.84	−0.28	0.39	−0.80	0.42	1.39
6PGD I	1.81	−0.39	−0.49	0.28	−0.82	−0.24	0.41	0.70	0.15	−0.44
6PGD II	−0.04	0.36	0.61	1.03	1.72	1.73	0.98	1.64	−1.37	−2.57*
IDH	0.86	0.43	0.68	0.03	0.09	0.20	1.09	0.88	0.24	−1.48
B. Indian Creek sample										
GOT I	1.61	1.04	0.83	1.01	0.91	1.26	2.28*	0.76	−0.15	0.36
ACO	−0.23	−0.63	−0.95	−0.37	−0.50	−0.38	−0.09	−0.28	−1.42	0.91
ADH	−1.08	0.94	0.67	0.48	0.90	0.06	1.69	−0.36	−0.69	−0.27
ACPH	0.28	1.05	−0.28	−1.99*	0.88	−0.63	−0.24	−1.03	−0.54	−0.51
6PGD I	1.16	1.92	0.83	0.58	0.79	−0.20	−0.46	1.47	0.78	−0.35
6PGH II	1.10	0.22	−0.07	−0.59	−0.03	−0.09	−0.06	−0.70	0.12	0.49
IDH	1.86	0.20	−0.19	−0.34	−0.15	−0.24	−0.21	−0.01	0.49	0.42

* Significant at the 5% level (two-tailed test).

limited pollen movement from very small clusters of heterozygotes.

No trends were observed in the statistics for pairs of genotypes of a locus for the longer distance classes (results not shown). The statistics for total number of unlike joins (Table 4) also indicate that the pollen cloud was unstructured spatially at distances of $k = 2$ or greater for most loci. These test statistics, in part because they are for haploid data, are expected to provide a sensitive summary measure of positive autocorrelation of like genotypes. Thus, for most loci, the absence of significant negative values at short distances for unlike pollen genotypes is inconsistent with the pollen cloud heterogeneity that is expected if there is a high degree of genetic isolation by distance. An exception is the previously discussed evidence for aggregation of *GOT I* pollen types up to $k = 3$ in the Scar Mountain sample. It is worth noting that the lack of significant autocorrelation statistics for pollen types for *GOT I* in the Indian Creek sample is not necessarily inconsistent with evidence that there may have been limited fine-scale structure of the common genotypes of *GOT I* among adults in this sample. The lack of significant values for *GOT I* pollen types in the Indian Creek sample for $k = 1$ to 3 might be due to the combination of alleles *2* and *3* of *GOT I*

in the pollen data; also further homogenization from pollen movement is expected.

Spatial autocorrelation statistics for multilocus pollen genotypes are expected to be sensitive measures of pollen cloud heterogeneity that might be generated by other causes such as limited pollen movement and variations in pollen production among individuals. All of the significant SNDs for joins between unlike two-locus genotypes for distance class one in the Scar Mountain sample were negative and involve *GOT I*; these probably simply reflect the autocorrelation of *GOT I* (single locus) pollen types. In the Indian Creek sample, the only significant SNDs, those between *IDH* and *6PGD I* or *6PGD II*, were positive; it seems that these cases result from sampling error, however, because each of these three loci are nearly fixed for one pollen genotype (Table 2). Few cases of significance were observed at greater distances. Only 8 among 189 test statistics for the total number of unlike joins, computed for $k = 2$ to 10, were significant in the Scar Mountain sample and only three were significant in the Indian Creek sample. Moreover, most of the significant values in the Scar Mountain sample involve *GOT I*. Those which occurred in distance classes 9 and 10 might well be artifacts of this sample; the dimensions of the lattice were 12×17, hence pairs of points 9 or 10 units apart are polarized in the *17* unit direction. We conclude that the two-locus pollen genotype distributions are mostly at random, and where not, they apparently resulted from autocorrelation at one of the loci (*GOT I* in the Scar Mountain sample).

Finally, SNDs were computed for each combination of observed five-locus pollen types for loci *GOT I*, *ACO*, *ADH*, *ACPH*, and *6PGD I*, for the first distance class. The results are easily summarized: in the Scar Mountain sample, only 49 of the 1035 SNDs, representing different combinations of 45 genotypes observed (out of a maximum of 162 genotypes, based on observed numbers of alleles) were significant, and in the Indian Creek sample, only 61 among 946 combinations of 43 genotypes were significant. Thus, the number of significant test statistics at the 5% level are near the number expected with random sampling. In addition, only one or two statistics among the 45 or 43 such tests in each sample were significant for like genotypes in both samples. Thus, these results indicate that the spatial distribution of multilocus pollen types is random or nearly so.

DISCUSSION

Gene flow prevents formation of large genetically isolated demes: Clearly, the results show that the distribution of sample genotypes is random, or very nearly random, for most loci in both the Scar Mountain and Indian Creek samples. Moreover, the structure is small-scaled for the few exceptional loci in which the spatial distribution of sample genotypes indicates significant autocorrelation. Thus there is little or no large-scale structure within either of the study sites. The data set does not allow structure on a scale smaller than 50 feet (the minimum distance between sample locations) to be directly discounted. However, theoretical results suggest that mating by proximity alone is unlikely to have led to the development of patches smaller than 50 feet in diameter (*i.e.*, patches consisting of *ca.* 50 individuals, on average) in the two populations of lodgepole pine which were sampled. Simulated populations in which S. WRIGHT's neighborhood sizes, N_e, ranged from about 5 to 25, develop large aggregations of several hundred homozygotes (SOKAL and WARTENBERG 1983; TURNER, STEPHENS and ANDERSON 1982; B. K. EPPERSON, unpublished simulations). Also, EPPERSON and CLEGG (1986) found that populations of the bumblebee-pollinated morning glory, *Ipomoea purpurea*, are structured into large patches of homozygotes for an apparently selectively neutral locus (EPPERSON and CLEGG 1986). Moreover, estimates of patch sizes based on spatial autocorrelation statistics are larger in populations having larger neighborhood sizes among the simulated populations (SOKAL and WARTENBERG 1983).

Measurements of pollen dispersal distances in populations of lodgepole pine and other conifers typically have standard deviations (σ_p) in the range of about 50 to 200 feet (reviewed by WRIGHT 1953, 1962; STRAND 1957). Horizontal distances of dispersal for seeds of lodgepole pine and other conifers follow a negative exponential distribution, and the majority of seeds of lodgepole pine fall within about 200 feet of their source (*e.g.*, LOTAN 1975; review by CRITCHFIELD 1980). This implies that the standard deviation of seed dispersal (σ_s) is in the range of 50 to 100 feet. However, the above studies were based either on isolated trees, or trees located at an edge of a stand, and hence they may have overestimated the extent of seed dispersal in a closed-canopy forest. Calculation of neighborhood area, A, based on CRAWFORD's (1984) method [$A = 4\pi(\sigma_p^2/2 + \sigma_s^2)$] and conservative estimates of $\sigma_p = 50$ feet, and $\sigma_s = 50$ feet, indicate that A is 1.1 acres. Present densities, which in the study populations are about 1000 overmature trees (almost all are breeders) per acre (EPPERSON 1983), are typical for overmature stands in the region (CRITCHFIELD 1980). Thus N_e probably has been at least 1000 individuals historically, and perhaps several times larger, in the study populations. The spatial variance of allele frequencies and WRIGHT's F_{ST} should be near zero for neutral loci in an outcrossing population with neighborhood sizes of 1000 (WRIGHT 1946; CRAWFORD 1984; see also SOKAL and WARTENBERG, 1983).

Several other results support the conclusions based on the present autocorrelation analysis of pollen and adult genotypes that development of patch structure is prevented for most loci by overwhelming gene flow. The outcrossing rate, t, is very nearly 1.0 in both of the study populations (EPPERSON and ALLARD 1984), which, in the absence of spatial structure, should lead to values of Wright's fixation indices near zero (WRIGHT 1965). This was true in both the maternal trees and seed embryos for most loci (EPPERSON and ALLARD 1984, 1987). Moreover, single-locus estimates of t were statistically homogeneous (EPPERSON and ALLARD 1984), which also suggests that population subdivision is negligible (SHAW, KAHLER and ALLARD 1981; RITLAND 1985).

The lack of genetic isolation by distance observed in the present study contrasts with results of most previous studies of plant populations (EPPERSON and CLEGG 1986; reviews by LEVIN 1981; CLEGG and EPPERSON 1985; EPPERSON 1989). This difference is probably attributable to the dispersal of seed and pollen over longer distances in lodgepole pine (see *e.g.*, LEVIN 1981). Another important factor may be that stand densities were typically high in the populations of *P. contorta* ssp. *latifolia* studied. It should be noted, however, that near panmictic conditions prevailed in other lodgepole pine populations in which stand densities were several times lower than in the study populations and that genotypic frequencies in different contiguous sample quadrats in four such populations were statistically homogeneous (KNOWLES 1984). Fixation indices, F_{IS}, within quadrats were also small (average F_{IS} was 0.03) (see also DANCIK and YEH 1983; PERRY and DANCIK 1986).

Previous studies of conifer populations have not employed spatial autocorrelation methods. No evidence of structure was found in pitch pine populations (GURIES and LEDIG 1977), or in Douglas-fir populations (SHAW and ALLARD 1982; NEALE and ADAMS 1985). In contrast, in whitebark pine, substantial familial structure was observed, and this was attributed to the unusual dependency for seed dispersal in this species on a bird (Clark's nutcracker) and the birds' habit of caching the cones (FURNIER *et al.* 1987).

Small scale autocorrelations of the genotypes of some loci: Levels of gene flow sufficiently high to randomize genotypes of neutral loci will not necessarily randomize the genotypes of loci under natural selection. In some cases, selection operating differently in closely adjacent microhabitats can maintain high levels of spatial variations of genotypic frequencies despite considerable seed and pollen dispersal (review in BRADSHAW 1984). In the present study gene flow was high enough to prevent genetic isolation by distance in the two lodgepole pine populations for most loci; hence, detection of small-scale autocorrelation of genotypes for a few loci may provide evidence that such loci are under selection.

Evidence of structure was found for genotypes of *GOT I*; genotypes with rare alleles at this locus are spatially aggregated in both populations. Rare pollen types for *GOT I* were spatially autocorrelated in the Scar Mountain sample. In addition, there are apparent overlaps in the distributions of tree and pollen genotypes for *GOT I* in the Scar Mountain sample (Figure 1), which also indicates several small aggregations of rare alleles. Some heterozygotes at *PER II* in the sample from the Indian Creek site and some genotypes at *PER I* also appear to be spatially autocorrelated. Analyses of structure for *PER I* and *PER II* were limited because alleles of *PER II* were combined; and because pollen genotypes could not be scored for *PER I* and *PER II*. In contrast, for the remaining eleven loci, there was little evidence of structure in either the adult or pollen samples.

Other evidence of selection for genotypes of loci *GOT I*, *PER I* and *PER II*: Earlier analyses showed that allele frequencies for the Scar Mountain and Indian Creek samples were statistically homogeneous for most loci despite the fact that the two populations are separated by seven miles (EPPERSON and ALLARD 1987). This is consistent with studies that have shown that over 90% of the allozyme variability resides within populations of lodgepole pine separated over the vast range of the species (WHEELER and GURIES 1982; YEH and LAYTON 1979). However, two loci, *GOT I* and *PER I*, show relatively large and statistically significant differences in allele frequencies (up to 0.05 for *GOT I* and 0.06 for *PER I*) between the two sample sites (EPPERSON and ALLARD 1987). This may be evidence of differential selection at loci *GOT I* and *PER I* because differences in frequencies between the two samples should be similar for all selectively neutral loci. In addition only the genotypic frequencies for loci *PER I* and *PER II* have fixation indices significantly greater than zero, the value predicted for selectively neutral loci, as based on the nearly 100 percent outcrossing rates in the study populations (EPPERSON and ALLARD 1984). Thus loci *GOT I*, *PER I* and *PER II* differ from the other loci by several measures, each of which is influenced by stochastic spatial processes shared by all loci. In part, tight linkage among the three loci ($c \leq 0.02$; EPPERSON and ALLARD 1987) may account for some of the similarities of the spatial patterns and F-statistics, for these three loci. However, the values of these statistics also varied among the three loci, and this suggests that linkage is not the sole cause.

Evidence that epistatic selection for loci *GOT I*, *PER I* and *PER II* caused statistical correlations of the genotypes of these loci within individuals, *i.e.* linkage disequilibrium, in the Scar Mountain sample was pub-

lished earlier (EPPERSON and ALLARD 1987). The observed correlations are too large to have been caused by genetic drift and linkage only, but could have been caused by epistatic selection of weak to moderate intensity. We note that linkage disequilibrium can be produced between pairs of nonepistatic loci in structured populations. Although some structure was observed for loci *GOT I*, *PER I* and *PER II*, it apparently does not account for the observed disequilibrium (see APPENDIX).

Little is known about the effects of linkage and multilocus selection, either with or without a spatial component, on the spatial structure of outbreeding populations (WRIGHT 1978; EPPERSON 1989). It is worth noting that even epistatic selection (favoring combinations of genotypes) without a spatial component might, under conditions of relatively high levels of gene dispersal, produce limited temporary structure consistent with the statistics observed for loci *GOT I*, *PER I* and *PER II*. Thus, epistatic selection might be an alternative explanation to microhabitat selection, although it is also possible that both forms of selection are operating on these loci.

The authors thank MICHAEL T. CLEGG and R. R. SOKAL for helpful comments on this work, and AILEEN WIETSTRUK for the word processing of the drafts of the manuscript. This investigation was supported in part by National Science Foundation grants BM573-0113 and BSR-841-8381. Contribution from the Department of Genetics, University of California, Davis.

LITERATURE CITED

BRADSHAW, A. D., 1984 Ecological significance of genetic variation between populations, pp. 213–228 in *Perspectives on Plant Population Ecology*, edited by R. Dirzo and J. Sarukhán. Sinauer Associates, Sunderland, Mass.

BIJLSMA, R., R. W. ALLARD and A. L. KAHLER, 1986 Nonrandom mating in an open-pollinated maize population. Genetics **112:** 669–680.

CLEGG, M. T., and B. K. EPPERSON, 1985 Recent developments in population genetics. Adv. Genet. **23:** 235–269.

CLIFF, A. D., and J. K. ORD, 1981 *Spatial Processes: Models and Applications*. Pion, London.

CRAWFORD, T. J., 1984 The estimation of neighborhood parameters for plant populations. Heredity **52:** 273–283.

CRITCHFIELD, W. B., 1980 Genetics of lodgepole pine. USDA Forest Service Research Paper, WO-37, 57 pp.

DANCIK, B. P., and F. C. YEH, 1983 Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. Can. J. Genet. Cytol. **25:** 57–64.

ENNOS, R. A., and M. T. CLEGG, 1982 Effect of population substructuring on estimates of outcrossing rate in plant populations. Heredity **48:** 283–292.

EPPERSON, B. K., 1983 Multilocus Genetic Structure of Natural Populations of Lodgepole Pine. Ph.D. dissertation, Department of Genetics, University of California, Davis.

EPPERSON, B. K., 1989 Spatial patterns of genetic variation within plant populations, in *Population Genetics and Germplasm Resources in Crop Improvement*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer Associates, Sunderland, Mass. (in press).

EPPERSON, B. K., and R. W. ALLARD, 1984 Allozyme analysis of the mating system in lodgepole pine populations. J. Hered. **75:** 212–214.

EPPERSON, B. K., and R. W. ALLARD, 1987 Linkage disequilibrium between allozymes in natural populations of lodgepole pine. Genetics **115:** 341–352.

EPPERSON, B. K., and M. T. CLEGG, 1986 Spatial autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). Am. Nat. **128:** 840–858.

FELDMAN, M. W., and F. B. CHRISTIANSEN, 1975 The effect of population subdivision on two loci without selection. Genet. Res. **24:** 151–167.

FURNIER, G. R., P. KNOWLES, M. A. CLYDE and B. P. DANCIK, 1987 Effects of avian seed dispersal on the genetic structure of whitebark pine populations. Evolution **41:** 607–612.

GURIES, R. P., and F. T. LEDIG, 1977 Analysis of population structure from allozyme frequencies. In *Proceedings of the Fourth Southern Forest Tree Improvement Conference*, Gainesville, Fla.

KNOWLES, P., 1984 Genetic variability among and within closely spaced populations of lodgepole pine. Can. J. Genet. Cytol. **26:** 177–184.

LEVIN, D. A., 1981 Dispersal versus gene flow in plants. Ann. Mo. Bot. Gard. **68:** 233–253.

LIBBY, W. F., R. F. STETTLER and F. W. SEITZ, 1969 Forest genetics and forest tree breeding. Annu. Rev. Genet. **3:** 469–494.

LOTAN, J. E., 1975 Regeneration of lodgepole pine forests in the northern Rocky Mountains, in *Management of Lodgepole Pine Ecosystems: Symposium Proceedings*. Cooperative Extension Service, Washington State University, Pullman.

NEALE, D. B., and W. T. ADAMS, 1985 The mating system in natural and shelterwood stands of Douglas-fir. Theor. Appl. Genet. **71:** 201–207.

PERRY, D. J., and B. P. DANCIK, 1986 Mating system dynamics of lodgepole pine in Alberta, Canada. Silvae Genet. **35:** 190–195.

RITLAND, K., 1985 The genetic-mating structure of subdivided populations. I. Open-mating model. Theor. Pop. Biol. **27:** 51–74.

ROHLF, J. F., and R. R. SOKAL, 1981 *Statistical Tables*. W. H. Freeman, San Francisco.

SHAW, D. V., and R. W. ALLARD, 1982 Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor. Appl. Genet. **62:** 113–120.

SHAW, D. V., A. L. KAHLER and R. W. ALLARD, 1981 A multilocus estimator of mating system parameters in plant populations. Proc. Natl. Acad. Sci. USA **78:** 1298–1302.

SIDAK, Z., 1967 Confidence regions for the means of multivariate normal distributions. J. Am. Stat. Assoc. **62:** 626–633.

SLATKIN, M., 1987 Gene flow and the geographic structure of natural populations. Science **236,** 787–792.

SOKAL, R. R., and N. L. ODEN, 1978a Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10:** 199–228.

SOKAL, R. R., and N. L. ODEN, 1978b Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol. J. Linn. Soc. **10:** 229–249.

SOKAL, R. R., and D. E. WARTENBERG, 1981 Space and population structure, in *Dynamic Spatial Models*, edited by D. Griffith and R. MCKINNON. Sijthoff and Moordhoff, Alphen aan den Rijn, The Netherlands.

SOKAL, R. R., and D. E. WARTENBERG, 1983 A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics **105:** 219–237.

SOKAL, R. R., P. E. SMOUSE and J. V. NEEL, 1986 The genetic structure of a tribal population, the Yanomama Indians. XV.

Patterns inferred by autocorrelation analysis. Genetics **114:** 259–287.

STRAND, L., 1957 Pollen dispersal. Silvae Genet. **6:** 129–136.

TURNER, M. E., J. C. STEPHENS and W. W. ANDERSON, 1982 Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proc. Natl. Acad. Sci. USA **79:** 203–207.

WHEELER, N. C., AND R. P. GURIES, 1982 Population structure, genic diversity and morphological variation in *Pinus contorta* Dougl. Can. J. For. Res. **12:** 595–606.

WRIGHT, J. W., 1953 Pollen-dispersion studies: some practical applications. J. For. **31:** 114–118.

WRIGHT, J. W., 1962 *Genetics of Forest Tree Improvement.* FAO Forestry and Forest Products Studies, No. 16, Rome, 399 pp.

WRIGHT, S., 1943 Isolation by distance. Genetics **28:** 114–138.

WRIGHT, S., 1946 Isolation by distance under diverse systems of mating. Genetics **31:** 39–59.

WRIGHT, S., 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19:** 395–420.

WRIGHT, S., 1978 *Evolution and the Genetics of Populations. Volume 4. Variability Within and Among Natural Populations.* University of Chicago Press, Chicago, 580 pp.

YEH, F. C., and C. LAYTON, 1979 The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia.* Can. J. Genet. Cytol. **21:** 487–503.

Communicating editor: B. S. WEIR

APPENDIX

Sampling from a spatially heterogeneous population produces levels of linkage disequilibrium between pairs of loci that depend directly on the covariance of allele frequencies in space (see, for example, FELDMAN and CHRISTIANSEN 1975). The levels of correlation between loci in the spatial distribution of genotypes for pairs of *GOT I, PER I* and *PER II* appear to be too small to account for disequilibria observed between those loci. Although the spatial autocorrelation statistics for some genotypes at these loci are significant, the spatial variations in genotypic frequencies are small: deviations of the observed from expected numbers of joins were usually less than 10%. In addition, the spatial structure of genotypes is probably largely erased during the mating cycle, because for most loci the mating system is random over large populational areas. The degree of covariance can be measured by comparing the observed number of joins between unlike two-locus genotypes with an estimate, μ, of the expected number of such joins in samples collected from a population in which single-locus spatial distributions are nonrandom, but uncorrelated among loci. Using T_i to represent number of joins between unlike single-locus genotypes for locus i, and W twice the sum of all types of joins for distance class one, $\mu = W/2 - [2(W/2 - T_i)(W/2 - T_j)/W]$. The observed values are very close to the values of μ (Table 5), indicating that there is essentially no correlation of joins between like genotypes for different pairs of loci.

TABLE 5

Numbers of joins between unlike two-locus genotypes for pairs of loci *GOT I, PER I* and *PER II*

Loci	Observed	Expected[a]	SND	Expected[b]
A. Scar Mountain sample				
GOT I PER I	595	599.6	−0.49	593.1
GOT I PER II	646	642.4	0.44	644.8
PER I PER II	592	595.1	−0.33	588.7
B. Indian Creek sample				
GOT I PER I	483	508.2	−2.49*	477.8
GOT I PER II	579	587.8	−0.99	578.2
PER I PER II	499	489.6	0.91	506.9

[a] Predicted value based on observed two-locus genotypic frequencies.

[b] Predicted value based on observed numbers of unlike joins for each locus separately, see APPENDIX.

* Significant at the 5% level (two-tailed test).