New Forests **6**: 257–278, 1992.

Review paper

Spatial structure of genetic variation within populations of forest trees

B. K. EPPERSON

Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

Received 8 October 1990; accepted 24 May 1991

Key words: genetic structure, natural selection, population genetics, autocorrelation

Application. New spatial statistical methods are introduced for studies of genetic structure within stands. These methods measure levels of inbreeding and can detect natural selection. This information may be useful for improving the genetic quality of stand regeneration.

Abstract. The spatial pattern and structure of genetic variation are important aspects of the population genetics of forest stands. Combined with limits to seed and pollen dispersal, spatial structure affects the level of inbreeding and the action of natural selection. The genetic constitution of stand regeneration, following different forestry practices, is also influenced by spatial structure. For example, natural regeneration with seed trees involves sampling seed trees from a stand that may be genetically nonhomogeneous. This paper reviews theoretical and empirical results on spatial patterns of genetic variation, produced under limited gene flow and selection, in terms of recently developed spatial statistics (e.g., spatial autocorrelation). Genetic correlations in samples from spatially structured populations are also described, as well as how spatial samples can be used to characterize the structure of genetic variation, and how inferences can be made about (spatially distributed) components of fitness and yield.

Introduction — the nature of spatial structure

Knowledge of structure of genetic variation is fundamental to understanding the population genetics of forest trees and other plants. Structure influences and is influenced by virtually every aspect of the population genetics of forest trees, including mating system and the action of natural selection. *Spatial* structure is the distribution of genotypes over the two dimensional space of a stand. It can be characterized through the physical locations and the genetic or genealogical relationships between individual trees. Summary measures of spatial structure may be used to improve our understanding of the population genetics of forest trees. In addition, forest

management and gene conservation practices usually include a component of spatial sampling procedures. For example, the spatial distribution of genotypes of seed trees following logging affects the genetic constitution of regeneration. Thus spatial structure is relevant to the measurement and management of the genetic composition of stands. Considerable advances have recently been made in relating spatial patterns of genetic variation within populations to different types of natural selection and amounts of dispersal (Sokal and Wartenberg 1983; Sokal, Jacquez and Wooten 1989; Epperson 1989, 1990).

Statistical measurements of genetic structure within stands have primarily involved either measures of polymorphism and heterozygosity or labor intensive multilocus parentage analysis. Parent-offspring data or similar genealogical data can reveal some but not all of the kinship relationships among individuals (Morton 1973a, b). More generally, in a spatial context, kinship is averaged among individuals or groups of individuals, based on their orientation in space, usually in terms of distances of separation. For selectively neutral loci, spatial patterns of genetic variation reflect the average total kinship between individuals, which usually is a decreasing function of distances of separation (Malecot 1948; Morton 1973a, b; Barbujani 1987). Spatial structure combined with limits to pollen dispersal results in consanguinous matings and local inbreeding. Moreover, where selection is operating, the population dynamics of genetic variation may depend importantly on the spatial distribution of genotypes. Conversely, spatial pattern analyses can be useful in detecting and characterizing natural selection, especially microenvironmental selection.

There have recently been major advances in spatial statistics that can be used to measure the important features of spatial patterns of genetic variation precisely (Sokal and Oden 1978; Cliff and Ord 1981; Upton and Fingleton 1985). Few studies have applied spatial correlation statistics to the population genetics of forest trees (Epperson 1983; Epperson and Allard 1989; Wagner et al. 1989) or to other plant species (Epperson and Clegg 1986; Dewey and Heywood 1988; Schoen and Latta 1989).

The present paper examines the utility of various spatial statistics for the study of *spatial structure* of genetic variation for allozyme and other loci, within forest stands. First, some general features of genetic structure within stands of forest trees and microenvironmental selection are presented. Following sections discuss spatial statistics and how measures of spatial structure are influenced by dispersal and selection.

Spatial structure of genetic variation, life history, and mating system

Spatial patterns of genetic variation within stands are expected to vary widely among forest tree species, which have a wide variety of dispersal mechanisms, reproductive biologies, and mating systems. Conifers generally exhibit wind dispersal of pollen and seed, high rates of outcrossing, and high levels of inbreeding depression. Many angiosperm trees, including some economically important trees in tropical regions, have quite different pollination mechanisms and other mating system features. The effects of mating system on levels of heterozygosity within populations are well established (see Wright 1946; Hamrick and Godt 1989). Plant populations with low to moderate distances of dispersal of pollen and seed (relative to tree density) are generally expected to build up substantial genetic isolation by distance and spatial autocorrelations of genotypes for selectively neutral loci (Wright 1943; Malecot 1948; Sokal and Wartenberg 1983; Sokal, Jacquez and Wooten 1989; Epperson 1989). In such populations there develops a distinctive structure of patches, or large areas where one homozygous genotype predominates. Moreover, many forests, particularly tropical forests, have a high level of species diversity (e.g., Sakai 1985). Wherever a population has low densities, pollen and/or seed may disperse long physical distances, yet there could be considerable isolation by distance and spatial correlations of genotypes. There is some evidence of marked isolation by distance in temperate hardwoods (e.g., Merzeau et al. 1989).

Spatial structure of genetic variation has been implicated in many conifer populations. "Structure-sensitive" single locus estimates of outcrossing rates are commonly lower than the relatively structure-insensitive multilocus measures (Shaw, Kahler and Allard 1981; Ritland 1985). A rare exception occurred in my own studies of populations of lodgepole pine (*Pinus contorta* ssp. *latifolia* Dougl. ex Loud.), which in fact had little or no spatial structuring (Epperson and Allard 1984). It is worth noting that spatial and/or temporal heterogeneity of gene frequencies in pollen are the proximal causes of deflated single locus estimates. Such heterogeneity could occur even if the spatial distribution of reproducing individuals is random, but individuals differ in timing or amount of pollen production (Epperson and Allard 1989).

Studies of spatial distributions and genealogical relationships between individual progeny in forest stands have revealed considerable family clumping in both temperate conifers and tropical angiosperms (Mitton 1983; Sakai 1985). The degree of family clumping varies in studies of progeny of remaining trees in seed tree or shelterwood stands (Neale and

Adams 1985; Yazdoni, Lindgren and Rudin 1985). Much of the spatial correlations caused by familial clumping of seedlings may be subsequently removed by intense competition that usually leaves only one or a few survivors per clump at reproductive maturity. It is worth noting that the structure among the genotypes of adult trees should continue to be governed primarily by the amount of dispersal per generation, as long as the survival of seedlings is random. However, when self progeny and other highly inbred progeny are less likely to survive (Sorensen and Miles 1982), the development of local patches of homozygotes may be retarded, because inbred progeny are more likely to be homozygous than are outbreds (Bennet and Binet 1956).

Theoretical studies of structure under different kinds of selection may help to resolve the controversy surrounding interpretations of the common observation that deficits of heterozygosity (which may be caused by spatial structure and/or selfing) for allozyme and other loci tend to disappear as regeneration ages (Phillips and Brown 1977; Brown and Albrecht 1980; Moran and Brown 1980; Farris and Mitton 1984; Yazdoni, Muona, Rudin and Szmidt 1985, Plessas and Strauss 1986; Muona et al. 1987). Although some recent studies reveal a lack of heterozygote advantage among outbred progeny, suggesting that inbreeding per se is the cause (Strauss 1986; Strauss and Libby 1987), recent review articles continue to call for more work in this important area (Muona 1989). In general, it is not clear how low levels of inbreeding in progeny from mildly consanguinous mating affects survival. A potentially great problem in forest regeneration practices is that slightly inbred seedlings may survive but show inferior qualities as mature trees. It is important to study how inbreeding caused by spatial structure behaves in natural stands, and in manipulated stands and samples.

Experimental and theoretical evidence indicates that marked spatial patterns are produced in areas of species hybridization and introgression. Spatial autocorrelation analysis may improve our understanding of hybrid viability and introgression. Wagner et al. (1989) used spatial autocorrelation statistics to describe the distributions of chloroplast DNA genotypes in large samples from a hybrid swarm of lodgepole pine and jack pine (*Pinus banksiana* Lamb.). Many cpDNA variations showed strong autocorrelations across the contact region. Similar patterns may be found in zones of recent contact and admixture between populations of the same species.

Population biology and microenvironmental selection in forest trees

The potential for selection in conifers and many other forest trees is great.

Populations can tolerate tremendous genetic loads because of a great overabundance of pollen, seed, seedlings and finally saplings. For example, Campbell (1979) estimates that some 20,000 seeds are produced to replace one mature Douglas Fir. Moreover, there are remarkable levels of genetic variation maintained within a conifer population, even within a stand. Individual forest trees must tolerate a wide range of environmental conditions during a lifetime (Mitton and Grant 1984). Conditions can also differ greatly over very small distances. Differences in soil depth, various soil qualities, water availability, exposure, depth of winter snow pack, light availability, competing species, all can exist on a spatial scale as small as a few meters. On a somewhat larger scale are the classic contrasts, of conditions and often species compositions, between north and south facing slopes. Large intrapopulational differences in allozyme allele frequencies have been found between stands on north and south facing slopes (Mitton et al. 1977; Beckman and Mitton 1984), and between wet and dry sites (Mitton et al. 1989).

An understanding of microenvironmental selection would be of considerable direct importance to forest management. In order to utilize potential local adaptation of stands, regeneration methods must use materials that are derived from source materials from nearby areas (Brown and Moran 1981). However, important genetic variation is contained within populations and there is often little differentiation among populations (Hamrick and Godt 1989; Muona 1989). Given the potential for microenvironmental selection, it may be that populations have already arrived at near-optimal local solutions to microenvironmental heterogeneity within stands areas, the limits to dispersal, and constraints from avoidance of inbreeding depression. This could result in a near-optimal genotype for each potential tree site. In general, if microenvironmental selection within a stand is strong, then plus tree selection greatly loses its appeal. Moreover, planting of genetically superior seedlings becomes impractical if not impossible because it would require matching favorable genotypes of each planted seedling to identified microenvironments within the stand. It is important to characterize the theoretical structures within stands that result from different space-time or spatial evolutionary models of dispersal and selection. In general terms, each generation gene flow tends to reduce correlations, which are then reinforced by another round of selection favoring different genotypes in different microenvironments. With the high levels of dispersal for most conifers, selection must be strong in order for correlations to build up over time. More detailed discussions are presented below.

Microenvironmental selection may be detected from strong spatial correlations of genetic variation, and strong associations of genotypic frequencies with microenvironments (Bradshaw, 1984). Our ability to

attribute genetic differentiation to microenvironmental selection is complicated by the fact that several other factors, including limits to seed and pollen dispersal can also create differentiation. Spatial statistics may improve interpretations of associations of genetic variation with microenvironments. These methods measure spatial relations of genetic variation with microenvironmental factors more precisely. They may be modified to account properly for spatial correlations of genetic variation due to dispersal alone, and to the spatial patterns of the microenvironmental factors themselves. Discussions of various spatial statistics are presented in more detail below.

The most detailed study of spatial structure within stands is on populations of lodgepole pine (Epperson 1983; Epperson and Allard 1987, 1989). These populations are essentially 100 percent outcrossing and estimated values of Wright's neighborhood sizes exceed 1,000. However, two populations, separated by only 11 km, differed in allele frequencies for some allozyme loci as much as populations at different ends of the vast geographical range of the species (Yeh and Layton 1979; Wheeler and Guries 1982). Within each study population, there were spatial autocorrelations of genotypes for some loci but not others, possibly indicating microenvironmental selection for some loci.

Spatial autocorrelation statistics for genetic variation

In this section, two important spatial statistical methods appropriate for genetic data are outlined. More detailed discussions can be found in Epperson (1989). The first method considered below is for point samples of genetic values that are approximately continuously varying. For example, allele frequencies, p_i, in a collection of n quadrat subsamples ($i = 1, \ldots, n$) are mapped so that each subsample is assigned to a point location. The second method is for data of nominal types, in this case genotypes, where a single nominal type can be assigned to a sample point location. Spatial autocorrelation statistics for both types of traits have better properties where the sample point locations are both fairly large in number and regularly spaced. Ideally, sample points are located on a regularly spaced sample grid or lattice (Epperson 1989).

Spatial statistical analyses of allele frequencies in quadrat subsamples proceeds through establishing distance measures between the pairs of quadrats. Usually this measure is simply the physical distance between the centerpoints of quadrats, but other distance measures may better reflect long term gene flow (Gabriel and Sokal 1969; Cliff and Ord 1981).

One important measure of correlation is based on the unweighted

Moran's I-statistic. First, the pairs of quadrats are classified into distance range classes. For each distance class k Moran's I statistic is calculated by: $I = n \sum\sum_{i \neq j}^{i,j \in k} Z_i Z_j / W_k \sum_{i=1}^{n} Z_i^2$, where $Z_i = p_i - \bar{p}$, and $\bar{p}$ is the mean allele frequency of all n quadrats and W_k equals twice the number of pairs of quadrats in the distance class k. Under the random hypothesis, I has expected value $u_1 = -1/(n-1)$. The variance, u_2, is given for example in Sokal and Oden (1978) and Cliff and Ord (1981). If the number of quadrats is fairly large and the number of genotypes per quadrat is moderate, then the statistic $(I - u_1)/\sqrt{u_2}$ has an approximate standard normal distribution under the random hypothesis (Cliff and Ord 1981).

A set of unweighted I statistics for mutually exclusive distance classes is known as an I-correlogram. Thus I-correlograms measure relative correlations in allele frequencies as a function of the distance measure. I-correlograms can be tested as a whole for significant deviation from the random hypothesis (Oden 1984), but exact tests are lacking for differences between correlograms from different data sets, for example, frequency distributions for different loci (Sokal and Wartenberg 1983).

An alternative method is to calculate separate weighted I statistics, in which the weights, w_{ij}, between locations i and j, can be specified based on any independent information on the relative strengths of correlations among pairs of subsample quadrats. In general, $I = n \sum\sum_{i \neq j} w_{ij} Z_i Z_j / W \sum_{i=1}^{n} Z_i^2$, where W is the sum of w_{ij} over all i, j such that $j \neq i$ ($w_{ii} = 0$). In addition, modified I statistics can be used to test the fit of a pattern to a theoretical (determined) map (Cliff and Ord 1981).

A more general method employs the calculation of a single statistic, the Mantel statistic (Cliff and Ord 1981), which tests the independence of two matrices. In the present context, one matrix contains values of a measure of differences in allele frequencies [for example Nei's genetic distance (Nei 1973)], and the other contains values of a physical distance measure.

Directionality in a spatial data set, such as may be caused by directional dispersal, maybe detected by classifying pairs of sample points by their respective locations in subareas defined by sectored concentric rings of an encircled total sample area (Oden and Sokal 1986). Alternatively, directionality along a predetermined spatial axis may be included in the distance measure in each of the statistical models (Upton and Fingleton 1985).

The second method involves using join counts to analyze a spatial map of n point sample genotypes. Two criteria are used to form subsets of the $n(n-1)/2$ total number of pairs of points, or joins. The first is based on some distance measure of the spatial relationship between pairs of points, and distance classes k are formed. The joins are further defined by the two genotypes for a pair of points. Thus, $n_{ij}(k)$ is the number of joins

between genotypes i and j for distance class k (i.e., the number of pairs of points which have genotype i at one point and j at the other and are separated by distances that fall within distance class k). For example, each distance class k may contain all pairs of sample points separated by d sample lattice units, where $k - 0.5 \leqslant d < k + 0.5$. Figure 1 shows the distribution of single locus heterozygotes for one of three rare alleles, 1a, 1b, 1c (denoted A, B, C, respectively in Fig. 1) and for one of two common alleles (alleles 2 and 3) of locus GOT I in a sample of trees from a stand of lodgepole pine (Epperson and Allard 1989). In this example the total sample size is 204, and the total number of joins is $204 \times 203/2 = 20{,}706$. For the first distance class, $0.5 \leqslant d < 1.5$ lattice units (note that this includes strict nearest neighbors and "diagonal" neighbors), there are 15, 5, 1, 3, 0, 0 joins of A × A, A × B, A × C, B × B, B × C, and C × C, respectively out of a total number of 731 joins for distance class one. Other joins would also be calculated which included pairs in combination with the genotypes with alleles 2 and 3.

Test statistics can be calculated for the null hypothesis, H_0, that the sampling distribution of the numbers of joins is "random", i.e., that produced by sampling pairs *without replacement* from the total sample of genotypes. Effectively, H_0 purports that the locations of sample genotypes are randomized. Under H_0 the expected number of joins between genotypes i and j for any distance class k is $u_{ii} = Wn_i(n_i - 1)/2n(n - 1)$ and $u_{ij} = Wn_in_j/n(n - 1)$ for $j \neq i$. Here n_i is the number of times that

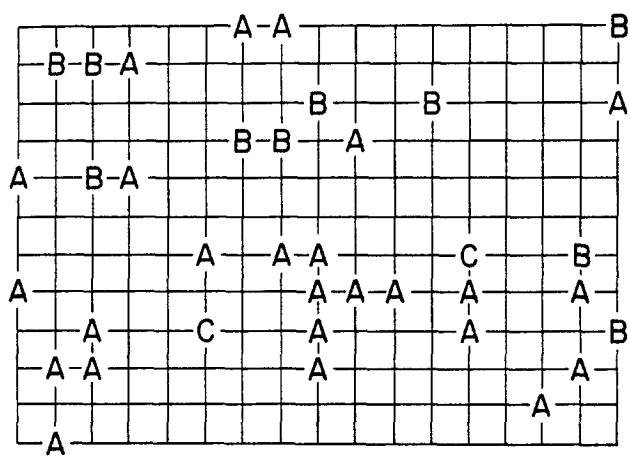

Fig. 1. Distribution of single locus heterozygotes for one of three rare alleles, 1a, 1b, and 1c (denoted A, B, and C respectively) and one of two common alleles (alleles 2 or 3) for locus GOT I in a sample of trees on a regularly spaced lattice from a population of lodgepole pine. Lattice points are separated by about 16 m. (50 feet).

genotype i occurs in the sample, and (suppressing k) W is twice the number of joins in total for class k. [Alternatively w_{ij} can be assigned to pairs of points as described above, and W will equal the sum of the w_{ij}]. The standard errors, $SE_{ij}(k)$, under H_0 can be found (Sokal and Oden 1978; Cliff and Ord 1981), and under H_0 the test statistic $SND_{ij} = (n_{ij} - u_{ij})/SE_{ij}$ (k suppressed) has an asymptotic standard normal distribution (SND) (Cliff and Ord 1981). These significance tests generally have high statistical power (Cliff and Ord 1981), and SND test statistics for short distances are particularly sensitive to most forms of structure (Epperson 1989). For selectively neutral genes, isolation by distance generally causes large positive SNDs for joins between like homozygotes at short distances, whereas the signs of SNDs between heterozygotes may be unpredictable and dependent on details of dispersal (Epperson 1989). SNDs are also useful as measures of spatial correlations. Spatial patterns of populations may be directly inferred from sample distributions (Ord 1980) to the resolution of the sample lattice. In addition, join counts can be used to test other null hypotheses (Epperson 1989).

Correlograms of SND statistics for mutually exclusive distance classes can be formed, analogous to I-correlograms. Like I-correlograms the distances at which SND correlograms intercept zero, known as the X-intercepts, provide measures of the spatial scales of sample autocorrelations (Sokal and Wartenberg 1983; Epperson 1990).

In some cases, it may be desirable to calculate other types of joins. One alternative is to simply combine genotypes into nominal classes. For example, more significant positive correlations were found between rare heterozygotes of the GOT I locus in the lodgepole pine data (see Fig. 1) after combining all rare heterozygotes into one nominal type, and then calculating the SND statistics for all joins between rare heterozygotes (Epperson and Allard 1989). In contrast, SND statistics for other sums of types of joins cannot be calculated by simply combining nominal types, as for example all joins between like homozygotes (Epperson 1989). Explicit formulae for the expected value and variance under H_0 have been given for one particularly informative summary measure, the total number of joins between unlike nominal types (Sokal and Oden 1978; Cliff and Ord 1981). This statistic is closely allied to measures of diversity as functions of distance. For diploid genotypes, it is closely related to measures of genotypic distance such as the probability of individuals having different genotypes as a function of distance. For haploid data it is inversely related to probabilities of gene identity (Epperson 1989). Multilocus data can be handled in a variety of ways (Epperson and Allard 1989), but the simplest is to place all multilocus genotypes into separate nominal classes.

Dispersal and spatial autocorrelations in populations and samples under isolation by distance for neutral genes

Limited seed dispersal creates kinship among adjacent individuals, and subsequent limits to pollen dispersal results in consanguinous matings and inbred progeny. The classical isolation by distance models of Wright (1943) and Malecot (1948) incorporate analogous measures of dispersal (Crawford 1984), e.g., marital distances or parent-offspring distances. *Spatial averaging* and calculations from dispersal parameters, produce a function of the average a priori coefficient of kinship on distance of separation, $\phi(d)$ [the probability of identity by descent for two genes, one from each of two individuals or alternatively two subpopulations, that are separated by a given distance, d (Malecot 1948, 1973)].

If an outside systematic pressure (of strength m), which may represent either mutation or immigration, is added to keep the population from becoming fixed for one gene, then for long distances (at least) an equilibrium kinship function on distance (d) is obtained of the general form: $\phi(d) \cong ad^{-c}e^{-bd}$. The constants a and b are positive and are controlled by the dispersal parameters and the value of m. There has been some controversy over how c may depend on the number of spatial dimensions, especially for the case generally of most interest, i.e., for short distances in a model with two spatial dimensions (Imaizumi et al. 1970; Morton 1973a, b). The utility of the a priori coefficient of kinship is limited because it is defined with respect to an ancestral population, and can be determined only by tracing probabilities of descent through genealogies back to the founding population.

A measure which is directly estimable from genetic survey data is the *conditional kinship,* r_{ij}, or kinship between i and j relative to the existing population (when i and j represent two subpopulations rather than two individuals, r_{ij} is essentially the covariance in gene frequencies; Malecot 1973). It has been suggested that the spatially averaged conditional kinship function on distance is, $r(d) \cong (1 - \mathrm{L})\phi(d) + \mathrm{L}$ (Morton 1973a, b). However, the validity of this equation and the genetic meaning of L (a negative constant) have been questioned (reviewed in Epperson 1989). Unweighted Moran's I statistics are closely related to conditional kinship. If both are defined using the same distance classes and the number of points (subpopulations) is large, then $\mathrm{I}(d) \approx r(d)\ F_{st}$, where F_{st} is Wright's measure of variation in gene frequencies among all subpopulations (Barbujani 1987). There is no corresponding direct theoretical relationship between individual kinship coefficients and genotypic join counts. As a complete set, join counts contain more information which

corresponds roughly to the information contained in additional descent measures (Epperson 1989).

Direct results for spatial autocorrelation of genetic variation within populations in two spatial dimensions were obtained in several Monte Carlo simulation studies (Sokal and Wartenberg 1983; Sokal, Jacquez and Wooten 1989; Epperson 1990). In each study, simulated populations consisted of 10,000 individual genotypes at a selectively neutral diallelic locus, where the individuals were continuously distributed over a regularly spaced lattice of 100 × 100 points. Distances and rates of dispersal of individual propagules ranged from very low to moderate (see Epperson 1989). Spatial correlations were characterized by first partitioning a simulated population into 400 nonoverlapping 5 by 5 quadrats and then calculating I-correlograms on quadrat allele frequencies. Generally, the values of I statistics for short distances increased rapidly as patch structures (large areas where several hundred individuals mostly have one homozygous genotype) built up from the initially random distributions during the first 30 to 50 generations. After 50 generations the I-correlograms changed little. Moreover, for the wide range of dispersal parameters, I-correlograms varied only slightly (Sokal and Wartenberg 1983; Sokal, Jacquez and Wooten 1989; Epperson 1990). A typical correlogram is shown in Fig. 2. The salient features are:

(1) a steep decrease in values of I as distance increases from 1 to 4 or 5 quadrat units (or from 5 to 20 or 25 "interplant" units);
(2) X-intercepts in the range of 4 to 6 quadrat units; and
(3) very little change in the values of I at greater distances.

Correlograms for join count statistics also reflect the patch structure (Epperson 1989). SNDs for joins between like homozygotes are in excess at distances up to about 20 to 25 interplant units (4 to 5 quadrat units). In addition, similar patch structures were observed in the simulated populations with highly limited dispersal in a study by Turner et al. (1982).

The amount of genetic correlation expected in samples of genotypes for a neutral locus collected from populations with low to moderate dispersal could be predicted by imposing a sampling grid over the typical patch structures, or more roughly by rescaling the typical I-correlogram. For example, suppose that a set of sample quadrats, with nearest neighbor quadrats having centers 15 interplant units (3 simulation quadrat units) apart, were collected from a frequency surface with a I-correlogram like that in Fig. 2. Then nearest neighbor pairs of these sample quadrats should have values of I near 0.10, and sample quadrats twice as far apart should have negligible correlation. In addition, the predicted covariance or

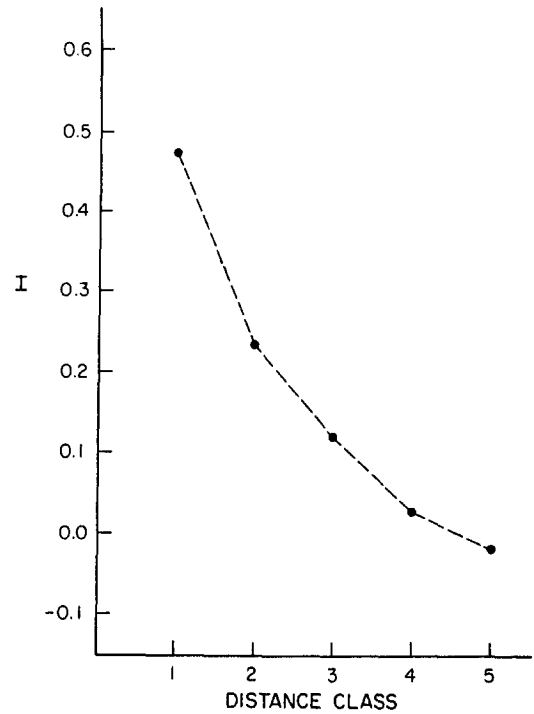

Fig. 2. Typical correlogram for simulated populations with low amounts of dispersal. Distance classes are in quadrat units, and each quadrat unit equals five times the distance between nearest neighbor individuals (i.e., five "interplant" units).

conditional kinship could be calculated by multiplying values of I by estimates of F_{st} from the sample quadrats. Moreover, sample quadrats located within the same patch should have high genetic correlations and kinship. These theoretical results are supported by several studies of plant populations with low to moderate levels of dispersal (Epperson and Clegg 1986; Schoen and Latta 1989). Another example of sampling illustrates the effects of seed tree logging from such populations. Suppose that prior to logging there are 1,000 mature trees per hectare, and after logging 40 well-spread out seed trees per hectare. Then average genetic correlations among nearest neighbor seed trees would approximate that for distance class one in Fig. 2. For theoretical purposes suppose that there were only 10 seed trees per hectare, then correlations would be closer to that for distance class two.

Many conifer populations have greater distances of dispersal (relative to density) than those in the simulations described above, and the theoretical spatial correlations in such populations are not well studied. How-

ever, for populations with very high levels of dispersal (e.g., Wright's neighborhood size >1000), there is virtually random mating (among outcrosses), little isolation by distance and thus little spatial autocorrelation (Wright 1943).

Spatial patterns under (non-microenvironmental) natural selection and immigration

Within populations that have limited dispersal, spatial distributions of genotypes for a locus under natural selection differ markedly from those for loci that are selectively neutral. This is true even where there are no components of microenvironmental selection. I conducted a series of Monte Carlo simulations of a diallelic locus under additive directional selection with intensity s, in populations with highly limited dispersal (Epperson 1990). In these simulations, selectively removed individuals were replaced by neighbors. Thus selection operated like local competition. (The simulations also featured an outside systematic pressure which could represent either mutation or immigration from outside populations with constant allele frequencies). The I-correlograms for populations with $s = 0.1$ were greatly reduced in comparison to the neutral case, and selection eroded patches of deleterious genotypes down to small average sizes. Differences in patch structures and I-correlograms caused by selection were established within 30 to 50 generations, and were unaffected by either initial or equilibrium allele frequencies. In contrast, selection had very little effect where $s = 0.01$.

Contrasts of spatial patterns in populations of morning glory confirm these results. Dispersal in populations of *I. purpurea* is very limited. Genotypes for loci that are selectively neutral (as based on independent experiments; Epperson and Clegg 1987) are distributed in large patches (Epperson and Clegg 1986), in accordance with neutral theory. In contrast, large consistent reductions of patch sizes occur for deleterious white-flowered genotypes.

To my knowledge there are no theoretical results on the spatial distributions of genotypes within populations when fitnesses are nonadditive. In populations with low seed and/or pollen dispersal, individuals with deleterious genotypes (via either viability or fecundity deficits) are "replaced" or outcompeted by their neighbors. In such cases, it may be expected that patches of recessive deleterious homozygotes would have larger average sizes than for the additive case, for the same intensity of selection. This follows from the likelihood that heterozygotes near the boundary of a patch, although having full fitness, also contribute homozy-

gous progeny to the patch through local inbreeding. Thus such structures may be expected for cases where selection acts primarily through inbreeding depression, which is a common feature of many populations of forest trees (Sorensen and Miles 1982). (Conversely, patches of dominant deleterious homozygotes might be somewhat smaller). In contrast, heterozygote advantage might cause some reductions in patch sizes for both homozygotes but an increase in sizes of areas containing mostly heterozygotes. This is an important area for further theoretical studies. Contrasts of spatial correlations (especially join counts) could be useful in distinguishing between heterozygote advantage and inbreeding depression (caused by recessive deleterious genes).

Random immigration and mutation alone, with rates lower than 0.01, have little effect on spatial correlations (Epperson 1990). Slight changes in structure and I-correlograms can be caused by certain other types of immigration when rates are on the order of 0.01. In particular, when the immigrant source population has very different allele frequencies, when immigrants arrive at only one boundary of the recipient population, and when dispersal within the recipient population falls within a certain intermediate range, then immigration results in a weak cline of allele frequencies superimposed onto the patch structure (Sokal, Jacquez and Wooten 1989).

The independence of structure from most mutation and immigration processes, and independence from initial conditions and allele frequencies, strengthen our ability to make inferences based on contrasting patterns for loci that are selected with those for loci that are neutral. In addition, for pairs of unlinked neutral loci, there should be little correspondence or correlations over large areas (Sokal and Wartenberg 1983). Thus, contrasts and correlations among spatial patterns for different loci in multiple-locus studies provides a very powerful framework for studying natural selection.

Multilocus selection may be expected to interact complexly with the spatial structure of genetic variation. Even with nonepistatic selection, the average fitness of single locus genotypes may depend on linkage disequilibrium created from correlations between loci over space (Prout 1973). With nonadditive or epistatic selection, and linkage, the interplay between marginal fitnesses and multilocus spatial structure must be very complex.

In a detailed spatial autocorrelation analysis of two study populations of lodgepole pine, the effects of selection and spatial structure could be disentangled. Multiallelic genotypes at eleven of fourteen allozyme loci showed no evidence of spatial structure or convincing evidence of linkage disequilibrium (Epperson and Allard 1987, 1989). This accords with the high recombination rates among these eleven loci, combined with the

facts that the populations are nearly 100 percent outcrossing, and have Wright's neighborhood sizes exceeding 1,000. There was significant disequilibrium between some pairs of alleles of the three other, tightly linked, loci (GOT I, PER I, and PER II). Single locus spatial correlations were small but statistically significant for some genotypes of GOT I, PER I and PER II. Genotypes for different loci were not significantly correlated spatially. Thus structure could not be responsible for the observed disequilibrium (Epperson and Allard 1989), which was consistent only with epistatic selection. Both epistatic and nonepistatic (but microenvironmental) components of selection may be acting on these loci themselves or on adjacent chromosomal segments.

Microenvironmental selection—theory and statistical measures

As mentioned briefly above, most forest tree populations are prolific enough to tolerate tremendous genetic loads, and they contain high levels of genetic variation for both allozymes and morphological traits. Microenvironmental heterogeneity in numerous important abiotic and biotic factors can exist on spatial scales ranging down to a few meters. Thus there is considerable potential for microenvironmental selection. Differences in the genotypes that are favored between different microenvironments could cause a variety of changes in the spatial patterns of genetic variation within populations. Microenvironmental selection can also result in correlations between genetic variation and spatially distributed parameters of small scale microenvironmental heterogeneities. Both types of correlations will depend on several features of microenvironmental selection and the stochastic events inherent in local systems of mating system and seed and pollen dispersal patterns. Correlation analysis may be used to detect important factors of microenvironmental selection (Epperson 1989). Such information points to genetic and microenvironmental factors that may warrant further studies to measure fitness directly, possibly including the use of genealogical survey data to directly measure the reproductive success of individual mature trees. However, spurious correlations for the genotypes of a neutral locus can occur simply because of the patch structures created by limited dispersal (Epperson 1989). Spatial structure and correlations can be predicted for some cases in which the scales of both dispersal and population densities, and the grains of microenvironmental heterogeneities, take certain extreme values. For intermediate values of these factors, more complex spatial statistical methods may be used.

First, consider cases where microenvironmental heterogeneity is scaled much larger than the average spacing between (mature) individuals, where

microenvironmental selection strongly favors opposite genotypes is different microenvironmental zones, and where dispersal is extremely limited. This situation would create strong associations between genotypic frequencies and the type of zone. In each zone, gene frequencies should simply evolve to values near the equilibrium values predicted from the relative fitnesses within a zone. In addition, if instead dispersal is very great, strong microenvironmental selection could create predictable local deviations in gene frequencies as a stand matures; however, dispersal would erase the spatial correlations during reproduction. Gene frequencies in different zones in a mature stand could be near those predicted after one generation of selection from the fitness differentials between genotypes, for a zone. These patterns would generally exhibit less sharp features than where gene flow is highly limited. Spatial patterns of microenvironmental factors may take on any form, including very irregularly shaped zones, and thus may produce irregular patterns of gene frequencies. It is useful to consider theoretical results for autocorrelations calculated for artificially generated irregular patterns of values (e.g., gene frequencies). Sokal and Oden (1978) and Sokal (1979) found that:

(1) X-intercepts of I-correlograms are nearly equal to the average size of zones of spatial data where different zones vary in size;
(2) X-intercepts are closer to the smaller dimension of zones that are rectangular or irregular in shape; and
(3) helical and "ridge" clines of spatial data both result in large positive correlations at short distances and negative correlations at longer distances.

Naturally spatial associations of genotypic frequencies with microenvironmental zones would be essentially complete, and correlations between allele frequencies and microenvironmental values should reach values near 1.0 in the low dispersal case.

Where environmental heterogeneity is scaled much smaller than both the spacing of (adult) individuals and the dispersal distances, then the progeny from each tree genotype will fall into zones with probabilities that are nearly independent of the parents' location. Thus overall average fitnesses of genotypes may be essentially independent of spatial structure, even though the average fitnesses of genotypes may depend on the total amounts of each microenvironmental type. In mature stands the spatial correlations of genotypes with microenvironments would be near zero.

Where dispersal distances are intermediate, the spatial autocorrelations of genetic variation depend complexly on how the genotypic fitness differentials vary in strength and direction over space. Some theoretical results are available from Monte Carlo simulations. Sokal, Jacquez and Wooten

(1989) simulated populations with low to moderate dispersal and with several patterns of microenvironmental selection. A gradient of strengths of directional selection (i.e., always against the same allele of a diallelic locus) produced I-correlograms that differed little from the neutral case, except that I-statistics for long distances were much more negative. It appears that a cline is superimposed onto patch structure. Discrete microenvironmental zones that were similar in size to patches produced in the neutral case, accordingly produced I-correlograms that were very similar to those for the neutral case. More simulation studies need to be done, with different spatial scales and patterns, and different forms of within zone selection.

Correlations *between* genetic variation *and* microenvironmental factors also are confounded with a genetic patch structure overlay supported by limited dispersal. Patch structure alone could cause spurious correlations between genotypes for a neutral locus and microenvironmental factors in small samples (Epperson 1989). Conversely, gene flow also may blur spatial differences caused by microenvironmental selection, especially near the boundaries of microenvironmental zones.

A number of spatial statistical methods can be used for partitioning autocorrelations due to limits to gene flow from correlations caused by location (i.e., local selection) in known microenvironments. Important considerations of sampling design and statistical details are discussed in a previous paper (Epperson 1989). In cases where microenvironmental heterogeneities are represented as classification variables and the spatial point data are genotypic, then loglinear and similar statistical models, modified with spatial autocorrelation parameters, can be employed. If instead, genetic data are allele frequencies, then modified analysis of variance would be appropriate.

Using multiple regression methods we can model a spatially distributed set of n subsamples of allele frequencies, Y_i, $i = 1, n$ as the observed vector Y of the dependent variable, and X as an $n \times m$ matrix of values of m different microenvironmental factors (independent variables) for each of the n sample locations. The usual multiple regression model is then: $Y = Xb + e$, where b is a vector of coefficients of the strengths of the effects of the microenvironmental selection factors, and e is a vector of independent and identically distributed error terms. Two methods of incorporating spatial autocorrelations are: 1. include a spatial autoregressive component so that $Y = pWY + Xb + e$; or 2. incorporate interactions into the error term so that $Y = Xb + pWe + u$ (u is a vector of independent identically distributed error terms). Here W is a matrix of relative weights of interactions, and p is a scalar which measures the overall strength of interactions, caused by proximity and gene flow. It is

anticipated that even very simple forms of spatial weighting in W (i.e., with only low order spatial lags) may remove most of the spatial autocorrelation due to limits to dispersal. Statistical procedures exist for obtaining estimates and tests of significance for b and p, as well as tests for model fit. Details and examples are presented in Cliff and Ord (1981) and Upton and Fingleton (1985). A preliminary indication of whether these more complex models are required, or whether the usual multiple regression model is sufficient, can be gained by testing the residuals of the multiple regression model, $\hat{e} = Y - Y\hat{b}$ (where $\hat{b}$ is the vector of the estimates of the b_j, $j = 1, m$), for autocorrelations, using a modified I statistic (see Cliff and Ord 1981).

Summary

Spatial analyses of the distributions of genetic variation provide important information on the population genetics of plant species. High levels of variation for allozyme loci are particularly useful fore forest trees. In addition, spatial statistics can be used to relate spatial structure to consanguinity relationships such as parent-offspring data. Spatial autocorrelation statistics can be used to detect natural selection, because patterns produced under different types of natural selection are distinct from those for neutral loci, at least in populations with low to moderate dispersal distances. Such distinctions between patterns appear to be maintained under many conditions, and in part emerge from the fact that patterns "capture" the cumulative effects of a number of generations.

The amount of spatial structure of genetic variation within a stand is itself important to forest genetics. In general, many forestry practices contain a component of spatial sampling. Amounts of variation, levels of inbreeding, and local adaptedness of genotypes in spatial samples depend on the amounts of isolation by distance and natural selection operating within stands. This affects regeneration methods and collections for germplasm. High levels of genetic variation and great tolerance for genetic loads are present within stands. In addition, forest stands contain considerable microenvironmental heterogeneity. Thus, the opportunity for microenvironmental selection is great. The presence of strong genotype by microenvironment interactions would have important implications for the genetic quality of forest regeneration. Further theoretical and experimental studies are needed to more fully understand the amounts and causes of spatial structure of genetic variation within stands.

Acknowledgments

This investigation was supported in part by National Science Foundation grants BSR-8614608 and BSR-9006984 to M. T. Clegg. The author thanks Jim Hamraick for stimulating discussions, and Mike Clegg and two anonymous reviewers for helpful comments on earlier drafts of the manuscript.

References

Barbujani, G. 1987. Autocorrelation of gene frequencies under isolation by distance. Genetics 117: 777—782.

Beckman, J. S. and Mitton, J. B. 1984. Peroxidase allozyme differentiation among successional stands of ponderosa pine. Amer. Midl. Nat. 112: 43—49.

Bennett, J. H. and Binet, F. E. 1956. Association between Mendelian factors with mixed selfing and random mating. Heredity 10: 51—56.

Bradshaw, A. D. 1984. Ecological significance of genetic variation between populations, pp. 213—228. In: Dirzo, R. and Sarukhan, J. (Eds) Perspectives on Plant Population Ecology. Sinauer, Sunderland, Mass.

Brown, A. H. D. and Albrecht, L. 1980. Variable outcrossing and the genetic structure of predominantly self-pollinated species. J. Theor. Biol. 82: 591—606.

Brown, A. H. D. and Moran, G. F. 1981. Isozymes and the genetic resources of forest trees, pp. 1—10. In: Conkle, M. T. (Ed) Isozymes in North American Forest Trees and Forest Insects. Pac. SW Forest and Range Expt. Station Tech. Rep. # 48.

Campbell, R. K. 1979. Genecology of Douglas-fir in a watershed in the Oregon cascades. Ecology 60: 1036—1050.

Cliff, A. D. and Ord, J. K. 1981. Spatial Processes. Pion, London.

Crawford, T. J. 1984. The estimation of neighborhood parameters for plant populations. Heredity 52: 273—283.

Dewey, S. E. and Heywood, J. S. 1988. Spatial genetic structure in a population of *Psychotria nervosa*. I. Distribution of genotypes. Evolution 42: 834—838.

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, pp. 229—253. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (Eds) Plant Population Genetics, Breeding, and Genetic Resources. Sinauer Associates, Sunderland, Mass.

Epperson, B. K. 1990. Spatial autocorrelation of genotypes under directional selection. Genetics 124: 757—771.

Epperson, B. K. and Allard, R. W. 1984. Allozyme analysis of the mating system in lodgepole pine populations. J. Hered. 75: 212—214.

Epperson, B. K. and Allard, R. W. 1987. Linkage disequilibrium between allozymes in natural populations of lodgepole pine. Genetics 115: 341—352.

Epperson, B. K. and Allard, R. W. 1989. Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. Genetics 121: 369—377.

Epperson, B. K. and Clegg, M. T. 1986. Spatial autocorrelation analysis of flower color

polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). Am. Nat. 128: 840—858.

Epperson, B. K. and Clegg, M. T. 1987. Frequency-dependent variation for outcrossing rate among flower color morphs of *Ipomoea purpurea*. Evolution 41: 1302—1311.

Farris, M. S. and Mitton, J. B. 1984. Population density, outcrossing rate and heterozygote superiority in ponderosa pine. Evolution 38: 1151—1154.

Gabriel, K. R. and Sokal, R. R. 1969. A new statistical approach to geographic variation analysis. Syst. Zool. 18: 259—270.

Hamrick, J. L. and Godt, M. J. 1989. Allozyme diversity in plant species, pp. 43—63. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (Eds) Plant Population Genetics, Breeding, and Genetic Resources. Sinauer Associates, Sunderland, Mass.

Imaizumi, Y., Morton, N. E. and Harris, D. E. 1970. Isolation by distance in artificial populations. Genetics 66: 569—582.

Malecot, G. 1948. Les Mathématiques de l'Hérédité. Masson, Paris.

Malecot, G. 1973. Isolation by distance, pp. 72—75. In: Morton, N. E. (Ed) Genetic Structure of Populations. University of Hawaii Press, Honolulu.

Merzeau, D., DiGiusto, F., Comps, B., Thiebaut, B., Letouzey, J. and Cuguen, J. 1989. Genetic control of isozyme systems and heterogeneity of pollen contribution in Beech (*Fagus sylvatica* L.). Silvae Genetica 38: 195—201.

Mitton, J. B. 1983. Conifers, pp. 443—472. In: Tanksley, S. D. and Orton, T. J. (Eds) Isozymes in Plant Genetics and Breeding. Elsevier, Amsterdam.

Mitton, J. B. and Grant, M. C. 1984. Associations among protein heterozygosity, growth rate and developmental homeostasis. Ann. Rev. Ecol. Syst. 15:479—499.

Mitton, J. B., Linhart, Y. B., Hamrick, J. L. and Beckman, J. S. 1977. Observations on the genetic structure and mating system of ponderosa pine in the Colorado Front Range. Theor. Appl. Genet. 51: 5—13.

Mitton, J. B., Stutz, H. P., Schuster, W. S. and Shea, K. L. 1989. Genotypic differentiation at PGM in Engelmann spruce from wet and dry sites. Silvae Genetica 38: 217—221.

Moran, G. F. and Brown, A. H. D. 1980. Temporal heterogeneity of outcrossing rates in alpine ash (*Eucalyptus delagatensis* R. T. Bak.). Theor. Appl. Gent. 57: 101—105.

Morton, N. E. 1973a. Kinship and population structure, pp. 66— 69. In: Morton, N. E. (Ed) Genetic Structure of Populations. University of Hawaii Press, Honolulu.

Morton, N. E. 1973b. Kinship bioassay, pp. 158—163. In: Morton, N. E. (Ed) Genetic Structure of Populations. University of Hawaii Press, Honolulu.

Muona, O. 1989. Population genetics in forest tree improvement, pp. 282—298. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (Eds) Plant Population Genetics, Breeding, and Genetic Resources. Sinauer Associates, Sunderland, Mass.

Muona, O., Yazdani, R. and Rudin, D. 1987. Genetic change between life stages in *Pinus sylvestris*: allozyme variation in seeds and planted seedlings. Silv. Genet. 36: 39—42.

Neale, D. B. and Adams W. T. 1985. The mating system in natural and shelterwood stands of Douglas-fir. Theor. Appl. Genet. 71: 201—207.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. USA 70: 3321—3323.

Oden, N. L. 1984. Assessing the significance of a spatial correlogram. Geographical Analysis 16: 1—16.

Oden, N. L. and Sokal, R. R. 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. Syst. Zool. 35: 608—617.

Ord, J. K. 1980. Tests of significance using non-normal data. Geographical Analysis 12: 387—392.

Phillips, M. A. and Brown, A. H. D. 1977. Mating system and hybridity in *Eucalyptus pauciflora*. Aust. J. Biol. Sci. 30: 337—344.

Plessas, M. E. and Strauss, S. H. 1986. Allozyme differentiation among populations, stands and cohorts in Monterey pine. Can. J. For. Res. 16: 1155—1164.

Prout, T. 1973. Appendix to Mitton, J. B. and Koehn, R. K. Population genetics of marine pelecypods. III. Epistasis between functionally related isozymes in *Mytilas edulis*. Genetics 73: 487—496.

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

Sakai, K. 1985. Studies on breeding structure in two tropical tree species, pp. 212—225. In: Gregorius, H. R. (Ed) Population Genetics in Forestry. Lecture Notes In Biomathematics 60. Springer-Verlag, Berlin.

Schoen, D. J. and Latta, R. G. 1989. Spatial autocorrelation of genotypes in populations of *Impatiens pallida* and *Impatiens capensis*. Heredity 63: 181—189.

Shaw, D. V., Kahler, A. L. and Allard, R. W. 1981. A multilocus estimator of mating system parameters in plant populations. Proc. Natl. Acad. Sci. USA 78: 1298—1302.

Sokal, R. R. 1979. Ecological parameters inferred from spatial correlograms, pp. 167—196. In: Patil, G. P. and Rosenweig, M. L. (Eds) Contemporary Quantitative Ecology and Related Econometrics. International Cooperative Publishing House, Fairland, Md.

Sokal, R. R., Jacquez, G. M. and Wooten, M. C. 1989. Spatial autocorrelation analysis of migration and selection. Genetics 121: 845—855.

Sokal, R. R. and Oden, N. L. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. 10: 199—228.

Sokal, R. R. and Wartenberg, D. E. 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics 105: 219—237.

Sorensen, F. C. and Miles, R. S. 1982. Inbreeding depression in height, height growth, and survival of Douglas-fir, ponderosa pine and noble fir to 10 years of age. For. Sci. 28: 283—292.

Strauss, S. H. 1986. Heterosis at allozyme loci under inbreeding and cross-breeding in *Pinus attenuata*. Genetics 113: 115—134.

Strauss, S. H. and Libby, W. J. 1987. Allozyme heterosis in radiata pine is poorly explained by overdominance. Am. Nat. 130: 879—890.

Turner, M. E., Stephens, J. C. and Anderson, W. W. 1982. Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proc. Natl. Acad. Sci. USA 79: 203—207.

Upton, G. J. G. and Fingleton, B. 1985. Spatial Data Analysis by Example. Vol. 1. Point Pattern and Quantitative Data. Wiley, New York.

Wagner, D. B., Sun, Z.-X., Govindaraju, D. R. and Dancik, B. P. 1989. The population genetic structure of chloroplast DNA polymorphism in a *Pinus banksiana-P. contorta* sympatric region: spatial autocorrelation, pp. 19—32. In: Hallgren, J. E. (Ed) Molecular Genetics of Forest Trees. Proc. Frans Kempe Symp. Swedish Univ., Agric. Sci., Umea, Rep. 8.

Wheeler, N. C. and Guries, R. P. 1982. Population structure, genic diversity and morphological variation in *Pinus contorta* Dougl. Can. J. For. Res. 12: 595—606.

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.

Yazdani, R., Lingren, D. and Rudin, D. 1985. Gene dispersion and selfing frequency in a seed tree stand of *Pinus sylvestris* (L.), pp. 1139—1154. In: Gregorius, H. R. (Ed)

Population Genetics in Forestry. Lecture Notes In Biomathematics 60, Springer-Verlag, Berlin.

Yazdani, R., Muona, O., Rudin, D. and Szmidt, A. E. 1985. Genetic structure of a *Pinus sylvestris* L. seed-tree stand and naturally regenerated understory. For. Sci. 31: 430—436.

Yeh, F. C. and Layton, C. 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.