New Forests **6**: 279—309, 1992.

Review paper

Allozyme markers in breeding zone designation

R. D. WESTFALL[1,2] and M. T. CONKLE[1]

[1] *Institute of Forest Genetics, USDA Forest Service, Pacific SW Forest & Range Experiment Station, Box 245, Berkeley, CA 94701, USA;* [2] *National Forest Genetic Electrophoresis Laboratory, USDA Forest Service, 2345 Fruitridge Road, Camino, CA 95709, USA*

Received 7 December 1990; accepted 29 August 1991

Key words: geographic variation, multivariate analysis, trend-surface analysis, transfer risk

Application. Multi-locus analyses of allozymes indicate significant geographic variation in widespread forest trees. Under these conditions, allozymes in tandem with other traits, are useful in the development of breeding zones.

Abstract. Early studies of allozyme variation in plant populations suggested that allelic frequencies in some loci vary by geography. Since then, the expectation that allozymes might be useful in describing geographic patterns has generally not been borne out by single locus analyses, except on the broadest scale. Multi-locus analyses reveal the converse: canonical correlation analysis of individual, uniformly-spaced genotypes describe statistically-significant, complex patterns with geography. Multi-locus scores in four major species, *Abies concolor, Pinus lambertiana, P. ponderosa*, and *Pseudotsuga menziesii*, of the mixed conifer forest in the Sierra Nevada correlate 0.40 or greater with the first canonical vector of a geographical trend surface equation. The different species follow similar patterns by latitude and elevation. In contrast with patterns in the Sierra Nevada, large-scale differentiation is weak ($R^2 < 0.20$) among populations of *Pseudotsuga menziesii* in the Coast Ranges and Siskyou Mountains of northern California and southern Oregon, where differentiation may be local. For the purpose of forming zones, we subdivided scores of the first two to four canonical vectors into groups and plotted them as multidimensional contour intervals. Reclassification by discriminant analysis serves as an approximate guide to transfer risks within and among these groups.

Introduction

Early studies of the variation in allozyme markers in plant populations suggested strong associations with geography (Allard and others 1972). Similar results in tree species (Bergmann 1975; Tigerstedt 1973) led some forest geneticists to speculate that allozymes might contribute towards the

development of breeding zones (Conkle 1974; Feret and Bergmann 1976). But subsequent results of studies have so diminished such expectations that the present conventional wisdom asserts that allozymes offer little potential usefulness, except in the broadest sense (Adams and Campbell 1981; Falkenhagen 1985; Muona 1990). While mean genetic diversity in many forest tree species is usually high, little of this variation is distributed among populations. Mean heterozygosity in polymorphic loci for both gymnosperms and perennial woody plants is 0.30, while the proportion of electrophoretic variation among populations (G_{st}) is about 0.07 (Hamrick and Godt 1990). Moreover, disequilibria between pairs of loci often do not deviate significantly from zero (Brown 1984; W. T. Adams, pers. comm.). Although patterns have been found by latitude (Bergmann 1975), elevation (Lundkvist 1979), aspect (Mitton and others 1977), and climatic measures (Guries and Ledig 1981), most studies fail to detect ordered geographic differentiation (Moran and Adams 1989; Neale and Adams 1985).

Smouse and others (1982) recognized that the cumulative effects of small differences among allelic frequencies, when summed across loci, can increase differentiation among populations so that individuals can be reliably allocated to biological groups on the basis of their multilocus genotypes. Later, Conkle and Westfall (1984), Guries (1984), and Yeh and others (1985) showed that geographic patterns in trees could be detected with a variety of multivariate statistical techniques. Moreover, individual genotypes in classes formed by multidimensional contour intervals could be reclassified by discriminant analysis into their original groups in higher frequencies than than those randomly expected (Conkle and Westfall 1984). Thus multivariate descriptions of allozyme patterns can contribute to developing and monitoring breeding zones.

The discussion that follows will be limited to allozyme variation. Allozymes, compared with biochemical markers, have found the greatest use in forest population genetics, because of their ease of analysis and their Mendelian basis. Monoterpenes, of the other markers, also have found extensive application to population studies in forest trees (Hanover 1992, this issue, pp. 159—178), but have not been widely applied to the development of breeding zones. Monoterpenes can contribute to zone design, but their greatest use might be in gene conservation (Millar and Westfall 1992, this issue, pp. 347—371).

In the sections that follow, we compare the relative utility of allozyme versus metric traits for describing geographic patterns. Next we will outline the procedural steps used in describing geographic variation and applying the results to the formation of breeding zones. Because the ability to detect pattern is fundamentally a matter of the analytical methods used, we focus on multivariate techniques. We include a summary of our results

in four conifer species in California, present a case example of the contribution of allozyme patterns to developing zones for one of these species in the Sierra Nevada Mountains, review unresolved issues, and point to further developments.

Our intention in this review is to demonstrate that significant, regular geographic patterns exist in allozyme loci in forest tree species. Depending on the objectives of the breeder, information on allozyme patterns can therefore augment that of other traits in constructing breeding zones.

The utility of allozymes in developing breeding zones

Zone designations for breeding programs are made on the basis of two perspectives: the source populations and the distribution of planting sites (Namkoong and Kang 1990). When local populations are confirmed or assumed to be optimal (Rehfeldt 1990a), then source populations, which define the seed zone, and the breeding zone, defined by the planting environments, occupy the same zone. But when non-local sources are optimal (Namkoong 1969), planting site zones and the commercially important traits determine the appropriate zones in the source populations (Namkoong and others 1988: 52—53). Once a breeding program advances beyond the establishment of a base population, zones lose their utility. Even in later generations, zones have relevance for defining variation in potential commercial traits and managing gene conservation programs.

Zone designation under the local-is-best assumption has been based on geoclimatic and ecological data (Buck and others 1970) or descriptions of metric traits from short-term tests (Campbell 1984; Rehfeldt 1986). But under the non-local assumption, zones are based on long-term tests (Dorman 1976). Performance of metric traits can be attributed to ecological factors and, if commercially important traits are measured, optimum germplasm source locations for these traits can be determined. But at least 3 years are needed to gather data after the dissemination of propagules, low heritabilities reduce the precision of estimating geographic patterns, and genotype-environment (G x E) interactions complicate the development of seed transfer guidelines. In contrast, allozyme data can be amassed within a few months after collection and processing plant material. Also, multilocus genotypes can be inferred directly and are not subject to G x E interactions.

The utility of geographic patterns in allozymes depends on the objectives of a breeding program and the underlying assumptions. If non-local sources are optimal, only differences among source populations in the economically important traits are relevant (Namkoong 1969). Although breeding for single, broadly-adapted varieties in an attractive goal for

breeding programs, multiple local breeding populations may be more stable and allow for greater flexibility over the long run (Namkoong and others 1988: 51—52, 112—115). Also, current forest practices are shifting to those that minimally alter the genetic structure in natural populations: practices that conserve temporal and spatial heterogeneity (Namkoong 1985; Namkoong 1990) and minimize gene flow from plantations to native stands (Ledig 1988). Allozymes, along with other genetic markers, are applicable under this conservative management philosophy by providing data on natural population genetic structures and measuring gene flow (Loveless; Epperson; Mitton; Adams; and Ellstrand 1992, this issue), although consistent associations between allozyme genotypes and ecological factors have not been established for forest tree species (Hamrick and Godt 1990; Hamrick and others 1992, this issue, pp. 95—124). Moreover, the precision of methods for detecting geographic trends in single loci is low and patterns in allozymes have not consistently correlated with those in morphological traits (Hamrick and Godt 1990). Yet, patterns in some morphological and physiological traits are not necessarily correlated among one another (Dickinson and others 1988; Rehfeldt 1986), and it is these independent patterns in sets of traits that contribute to zonation (Campbell 1986). Nevertheless, metric traits are more highly weighted than isozymes in defining source populations for breeding zones.

Methodological considerations

Single-locus methods

In forest genetics, the method of choice for inferring geographic patterns from distance measures computed from single loci is cluster analysis. Almost without exception, the average linkage (UPGMA) clustering algorithm (Dunn and Everitt 1982) has been applied to matrices of Nei's distance (Nei 1978). However, the little-used Cavalli-Sforza and Edwards arc and chord distances (Cavalli-Sforza and Edwards 1967), which weight differences in allelic frequencies by their sampling errors, are better suited for determining the existence of geographic pattern. Additionally, the average linkage algorithm assumes equal rates of divergence. However, the distance-Wagner procedure is more appropriate for visualizing clines (e.g., Millar and others 1988), as it is not constrained by the assumption of equal rates of divergence (Swofford 1981).

A serious drawback with clustering algorithms is that they are sensitive to error variation, diminishing their ability to detect population differences (Archie and others 1989; Dunn and Everitt 1982). When population

samples encompass relatively large areas and contain large numbers of individuals per population (greater than 50), means estimated from these populations will reduce sampling errors and clines will be more easily detected than with means from smaller sample sizes (Millar and others 1988). However, clines may be highly nonlinear and therefore not readily detectable in the cluster analysis.

Multi-locus methods

Multivariate analyses are preferable to univariate because single loci are not apt to reveal patterns well, even when patterns are significant (Thorpe 1985). Secondly, multivariate analyses are more efficient: only one analysis is needed for all loci simultaneously. If a pattern exists for a single locus that is independent of all others, it will be clearly shown in any multivariate analysis.

Below we discuss techniques for the multilocus analysis of allozyme data starting with a scoring procedure that transforms these data to a form that is suitable for multivariate analytical methods. Then we discuss four methods that, with the exception of principal components analysis, are special cases of the multivariate general linear model.

Transforming genotypic data. Because multivariate statistical methods require that linear combinations of traits must be normally distributed, haploid or diploid genotypes must be transformed. An appropriate and elegant transformation for diploid genotypes is outlined by Smouse and Williams (1982). For each allele at a locus minus one, assign a value of 0.5 when the allele is present and 0.0 when the allele is absent. Scores for an allele in a homozygote would be $0.5 + 0.5 = 1.0$, for a heterozygote would be $0.5 + 0.0 = 0.5$, and for homozygotes and heterozygotes of other alleles would be 0.0. A similar scoring method for haploid genotypes is given in Weir (1990: 143). Diploid scores for a locus with three alleles are as follows:

	Genotype					
	11	12	13	22	23	33
Allele 1	1.0	0.5	0.5	0.0	0.0	0.0
Allele 2	0.0	0.5	0.0	1.0	0.5	0.0

Other additive scoring systems, such as that in Spielman and Smouse (1976), are equivalent to the above, except that the Smouse and Williams'

(1982) score represents the frequency of an allele at a given level of sampling. For a diploid individual, the score is the frequency of that allele in the individual and for a population, the score is the frequency for the population. According to the expectations of quantitative genetic theory (Kempthorne 1969: 18—19), linear combinations of such allelic scores over ten or more loci are normally distributed (Smouse and others 1982). Such expectations are confirmed by our analyses: residuals in canonical models fit the normal distribution. Although scores for codominant alleles are additive, the infrequent presence of null alleles will generate some small proportion of dominance variance.

Principal components analysis (PCA). PCA is a method used to describe patterns among populations in multidimensional space whereby principal axes in this space are aligned sequentially in the direction of greatest variance and covariance among traits (Morrison 1990). When all traits are uncorrelated, there will be as many principal components as there are traits. But if all traits are highly correlated with one another, their variation will collapse onto one or two principal components. PCA is sensitive to error variation resulting from sampling and genetic recombination (Dunn and Everitt 1982; Gittins 1985; Wartenberg 1985). Hence, analyses are based on population means, which will increase linkage disequilibrium among loci and reduce the number of significant principal components. Used for the analysis of geographic patterns among populations, PCA is hypothesis-free; no *a priori* hypotheses are imposed. Instead, the intent of PCA is to generate hypotheses about population structure (Guries 1984). Although PCA is poorly suited to describing complex clines, principal component scores of populations can be regressed against measures of their geographic location (cf Campbell 1986).

Canonical variate analysis (CVA). CVA, also known as canonical discriminant analysis, is the multivariate equivalent of a one-way analysis of variance (Gittins 1985). Allelic scores are weighted by their ratios of between-to-within group variation, thus reducing error variances. Consequently, CVA is usually better suited than PCA for describing stepped clines or ecotypic patterns (Gittins 1985). Because Mahalanobis distances from CVA (Morrison 1990) have properties similar to the Cavalli-Sforza and Edwards arc distances, they can also be applied to hierarchical cluster analyses (cf Merkle and others 1988). CVA is also hypothesis free when based on samples stratified by populations and individuals or subpopulations and when prior hypotheses about associations have not been imposed on the groups of populations (Kinloch and others 1986; Knowles 1985; Yeh and others 1985). CVA is also sensitive to sampling effects: the first few canonical vectors will focus on alleles that are uniquely represented in one or a few populations. As with PCA, CVA is limited in its

ability to detect nonlinear patterns, although canonical scores can be regressed against geographic measures (cf Merkle and others 1988).

Discriminant analysis. Discriminant analysis is related to CVA except that it functions in classification, whereby individuals are assigned into one of several groups (Morrison 1990: 269—287). The method was the first multilocus procedure to be applied to allozymes by Smouse and others (1982), who used the method to allocate Amarindians to populations. The utility of discriminant analysis in describing geographic differences is not so much in classification per se, but in the comparative classification errors among candidate groups. Relative percentages of classified and misclassified individuals indicate similarities among groups in comparison with their geographic distances. For reliable results, sample sizes must be very large or the classification must be validated by an independent dataset. The quadratic function (Morrison 1990: 275—278), used when the within-group variances and covariances are heterogeneous, has limited use in allozyme data because some loci are fixed in one or more groups and not in others, resulting in unequal numbers of variables among groups.

Canonical correlation analysis (CCA). CCA is the multivariate equivalent of multiple regression, but with more than one dependent variable, and the procedure is related to CVA in the same way that regression is related to ANOVA. CCA partitions the coefficient of multiple determination (R^2) of the multiple regression of single variables into separate, independent models, which are shared among dependent variables in the analysis. Although CCA has been applied to ecological studies (Gittins 1985), it's use in population genetics has been only recent (Wartenberg 1985). With the exception of sensitivity to outliers, CCA is less sensitive to errors than PCA and CVA, and errors that occur usually reside in the residual variation. CCA is also superior to PCA for detecting irregular geographic structure (Wartenberg 1985).

Multi-locus geographic patterns and the development of breeding zones

Zone delineation depends on the assumptions behind the choice of optimal source populations. These assumptions in turn determine procedures that are appropriate for sampling, data analysis, and subdivision of source populations and planting sites into zones. Because premises that underly using data from allozyme loci or other genetic markers are similar to those associated with short-term test data, methods that have been used with short-term tests also can be applied to those from allozyme allelic scores (Campbell 1986; Rehfeldt 1990b).

Sample sizes

Multivariate techniques require large sample sizes and complete sets of data. As to the latter issue, each observation with an unresolved locus is deleted from the analysis, thus effectively reducing sample size. With the exception of PCA, all analytical methods require that the sample size to be larger than the number of parameters. A general rule developed from simulation studies of CVA, is that the sample size per group, N_i, should be at least equal to the number of dependent variables, p, in the model (Williams and Titus 1988). In allozyme data, the number of dependent variables is equal to the number of allelic variables. Thus, the minimum sample size is p times the number of groups, k, in CVA, or the number of independent variables in CCA. This rule appears to be quite conservative: Williams and Titus (1988) showed that because errors stabilize more rapidly with increasing sample sizes as the number of dependent variables increase, N_i can be less than p for $p > 30$. Our experience with opportunistically jacknifing data having more than ten loci indicates that samples with three times the total number of variables $(p + k)$ are sufficient for stable estimates of the first vector. Thus El-Kassaby (1990) errs in stating that the sample size in any group cannot be smaller than the number of dependent variables. The parameters in canonical analyses or the discriminant function can be estimated as long as $N - p - k$ is greater than zero (preferably much greater), where N is the total sample size.

To meet the requirement for multivariate normality, our experience suggests that data from ten polymorphic loci are sufficient. When each locus averages three alleles, or two allelic variables, p equals 20. With nine independent variables in the model, the minimum sample would be 180.

Sampling methods

If irregular or stepped-cline patterns exist, then CVA is applicable and a hierarchical sampling structure can be used, as in Merkle and others (1988). But when patterns are clinal, sampling should be systematically made as in a factorial design, on a regular grid or set of transects (Box and others 1978; cf Campbell 1986; Campbell and others 1989). Pairs of trees should be sampled at about two thirds of the intersections to provide samples for estimating the variance within locations and to test lack-of-fit in the model (Box and Draper 1987).

In practice, rigid grids are not feasible because species' ranges are not arrayed on a regular grid. Alternatively, a uniform distribution of samples over the range of interest would be suitable. Or when precise estimates of genetic diversity statistics are desired, 20 to 50 trees could be sampled at

strategically-spaced locations, and single and paired samples collected at the remaining grid intersections within practical limits of the study.

Analytical procedures

Initially, an exploratory analysis might be desired and for this, PCA and CVA are appropriate. Even when geographic patterns are not obvious from the exploratory analysis, the existence of higher-order patterns and their statistical signifiance should be tested by regression, either by regressing measures of location on principal or canonical scores (Campbell 1984), or on allelic scores by CCA. Because principal and canonical vectors might not be maximally alligned with geography in multi-locus space, CCA is the preferred method. The simplest models should be tested first, beginning with the trend-surface analysis of principal or canonical scores (Gittins 1968) or on allelic variables with its multivariate extension, canonical trend-surface analysis (CTSA) (Lee 1969; Wartenberg 1985). The model used in trend-surface analysis is a second-order polymomial of the measures of geographic location and can be applied by such statistical procedures as SAS' PROC GLM in the univariate condition and PROC CANCORR in the multivariate (SAS Institute Inc. 1985). The importance of additional measures of geography or of higher order terms can be tested by their "extra sums of squares" (Box and Draper 1987), using the Type I sums of squares of canonical scores in PROC GLM, for example (SAS Institute Inc. 1985). To reduce the number of allelic variables in the model, we drop alleles that are near zero for both the standardized canonical weights (< 0.4 approx.) and structure-correlations between the canonical score and the trend surface equation (approx. < 0.05).

For describing pattern, the important statistical data from CCA are the R^2s, the unbiased R^2s, and the statistical significance of the canonical vectors (PROC CANCORR, SAS Institute Inc. 1985). Also useful are the structure correlations between the dependent variables and the scores of the canonical vectors formed by the independent variables. Such structure correlations are the multivariate equivalent of the correlation between the dependent variable and the model described by the independent variables in multiple regression. The square of this structure correlation is the proportion of variance in the dependent variable described by the canonical model, and when summed over k canonical vectors, is the total proportion of variance described by the vectors, also called the redundancies (Gittins 1985).

Final choice of a model depends in part on its adequacy. In addition, we rely on various descriptions of the fitted surface to subdivide that

surface into zones. We test lack-of-fit in the model and the relative importance of linear and quadratic terms of the trend-surface model, using a program such as PROC RSREG (SAS Institute Inc. 1985). Using this same SAS procedure, we also describe the shape of the surface of canonical scores from CTSA by canonical analysis (Box and Draper 1987). To aid in the visual description of the patterns, we compute predicted canonical scores and their residuals using the same model from CANCORR in GLM. Contour plots of the predicted scores show the pattern in geographic space (e.g., by latitude and longitude), whereas contour plots of the residuals will suggest discontinuities in the pattern. The latter plots, along with a plot of the residuals against the canonical scores can show biases in the model and suggest additional terms or transformations of the data. (Draper and Smith 1966). To evaluate the normality of the residuals we use both the direct statistical test and the normal probability plot in PROC UNIVARIATE (SAS Institute Inc. 1985). In our experience with allozyme allelic score data in various tree species, residuals do not significantly deviate from normality, or when they do, normal probability plots indicate no important deviations.

Subdivision into zones

Zone formation is a matter of balancing operational needs for a minimal number of units with the biological necessity to minimize risk of transfer within each unit. To apply allozyme data to this process, we determine the pattern and amount of multi-locus geographic variation and integrate this information with data from other traits and with the objectives of a breeding program.

Contour plots of predicted canonical scores indicate regions of similarity in multilocus frequencies in geographic space. Canonical analysis (Box and Draper 1987; PROC RSREG, SAS Institute Inc. 1985) also estimates rates of change in canonical scores in directions within geographic space and provides an indicator of number of subdivisions in these directions. To aid in visualizing patterns in more than one canonical vector, we partition the scores of each vector into contour intervals, assign a letter to each combined interval, and map the multi-vector groups by latitude and longitude. Mapping these interval groups over topographic maps helps to visualize associations with elevation, aspect, watersheds, and other physiographic features.

Differentiation between and within interval groups can be evaluated by discriminant analysis and transfer risk, respectively. Reallocation of individuals into interval groups by discriminant analysis not only indicates similarities by geographic distance, but also suggests pattern unrelated to

distance. Such patterns then can be used to develop alternative divisions of the data. Groups having a large proportion of individuals classified into neighboring groups suggest that these groups might be consolidated into a single one. Alternative groups can be evaluated by omitting individuals in the test alternatives from the analysis, recomputing the discriminant functions, and classifying the test individuals into the remaining groups (SAS Institute Inc. 1985: 321—322). For example, this procedure can be used to test subdivisions or consolidations of geographically distant individuals that are grouped into the same contour interval.

The latter method, transfer risk, estimates the proportion of genotypes at one location that does not match those at another location (Campbell and Sugano 1987; Westfall 1991). In metric traits, the number of subdivisions is based on a predetermined minimum risk within groups or zones (Campbell and Sugano 1987). An application of minimum transfer risk in multi-locus allozyme genotypes to the location of gene reserves is illustrated in Millar and Westfall (1992, this issue, pp. 347—371).

Multi-locus patterns in four species of the mixed conifer zone in the Sierra Nevada

With support from many individuals in the Pacific Southwest Region Tree Improvement Program, we have analyzed geographic patterns in allozyme data from four species of the mixed conifer zone in California's Sierra Nevada: White fir (*Abies concolor* Lindl.) (Conkle and Westfall 1988), sugar pine (*Pinus lambertiana* Dougl.), ponderosa pine (*Pinus ponderosa* Laws.), and Douglas-fir (*Pseudotsuga menziesii* Franco). We will also summarize results from analyses of Douglas-fir from the northern Coast Range in California (Conkle and Westfall 1987). The seed parents genotyped to provide diploid data were superior trees that were well distributed throughout the commercial range of each species.

The latitudinal range for the white fir sample is from south of Mt. Lassen to south of Yosemite National Park (Fig. 1a). The ponderosa pine sample has the widest range, from Mt. Shasta in the north to the Greenhorn Mountains in the south (Fig. 1c). Northernmost samples for Douglas-fir and sugar pine are near Mt. Lassen (Figs. 1b and d). The southermost samples of Douglas-fir are near the southern limits of the species' distribution in the Sierra Nevada (Griffin and Critchfield 1972; Fig. 1d).

Individual trees were genotyped using megagametophytes. Diploid genotypes were transformed to allelic scores and analyzed by CTSA and discriminant analysis. Seed parent origins were identified by latitude and longitude (nearest minute), and elevation (nearest 30 m). These three

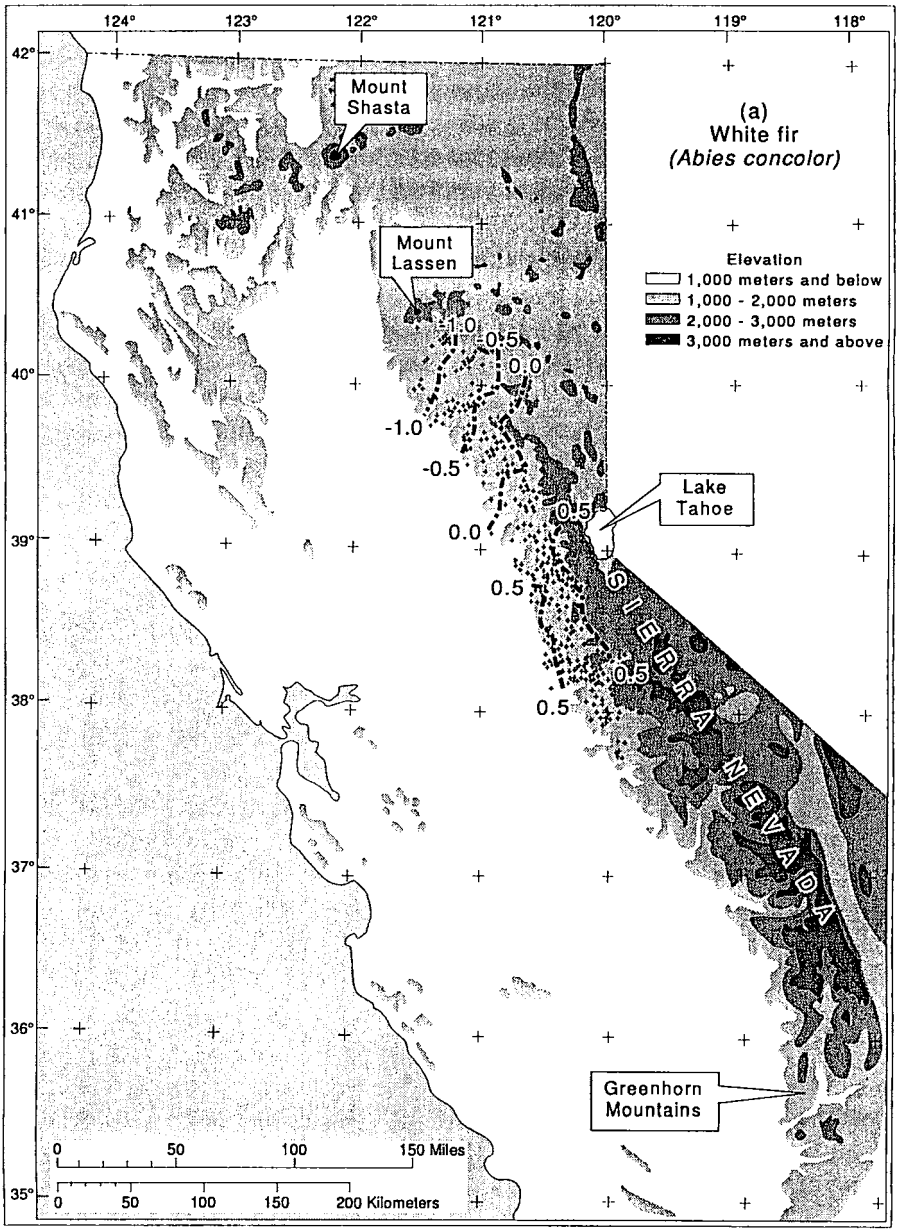
124°
123°
122°
121°
120°
119°
118°
42°
41°
40°
39°
38°
37°
36°
35°
Mount Shasta
Mount Lassen
(a)
White fir
(Abies concolor)
Elevation
1,000 meters and below
1,000 - 2,000 meters
2,000 - 3,000 meters
3,000 meters and above
-1.0
-0.5
0.0
-1.0
-0.5
0.0
0.5
0.5
0.5
Lake Tahoe
SIERRA NEVADA
Greenhorn Mountains
0
50
100
150 Miles
0
50
100
150
200 Kilometers

Fig. 1a

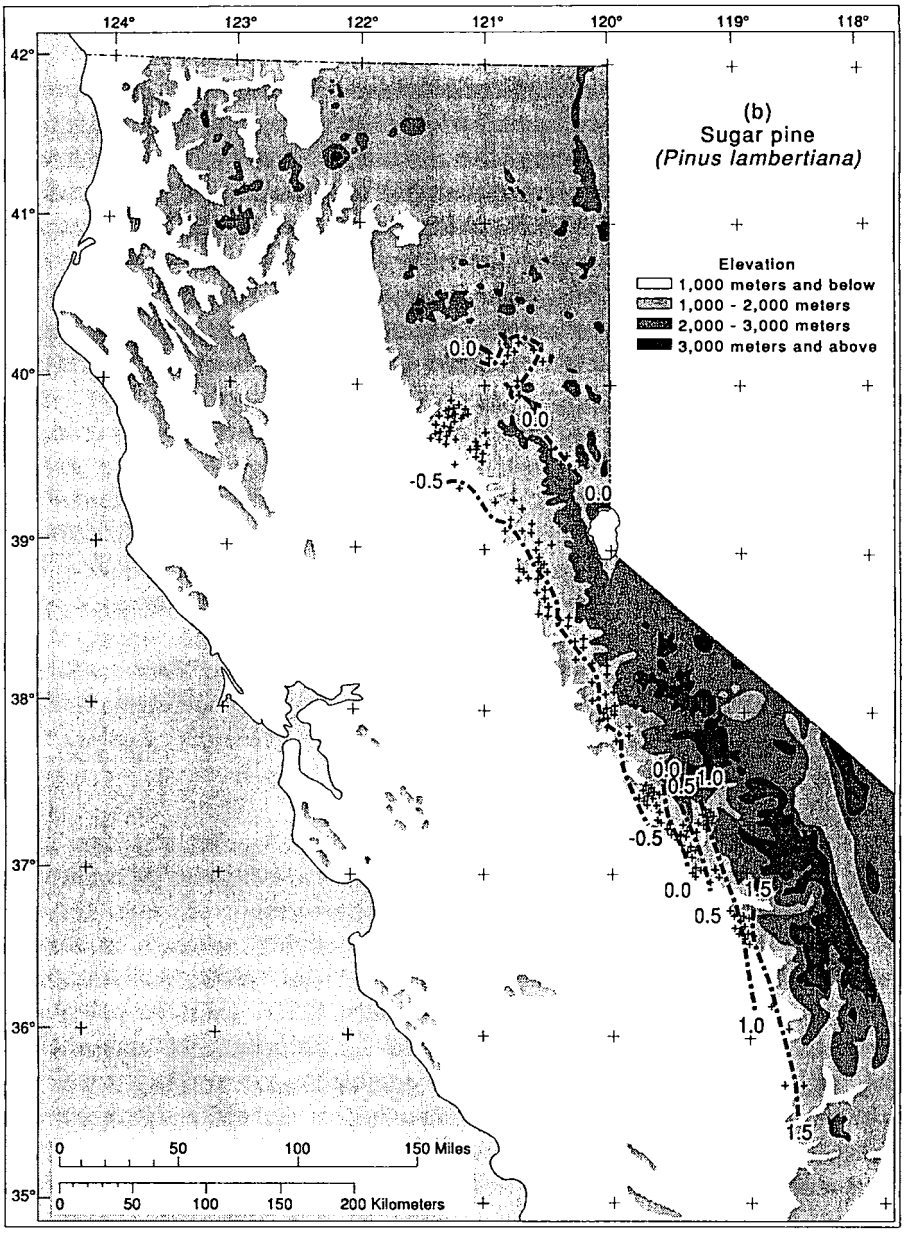
124°
123°
122°
121°
120°
119°
118°
42°
41°
40°
39°
38°
37°
36°
35°
(b)
Sugar pine
(Pinus lambertiana)
Elevation
1,000 meters and below
1,000 - 2,000 meters
2,000 - 3,000 meters
3,000 meters and above
0.0
0.0
-0.5
0.0
0.0
0.5
1.0
-0.5
0.0
0.5
1.0
1.5
0
50
100
150 Miles
0
50
100
150
200 Kilometers

Fig. 1b

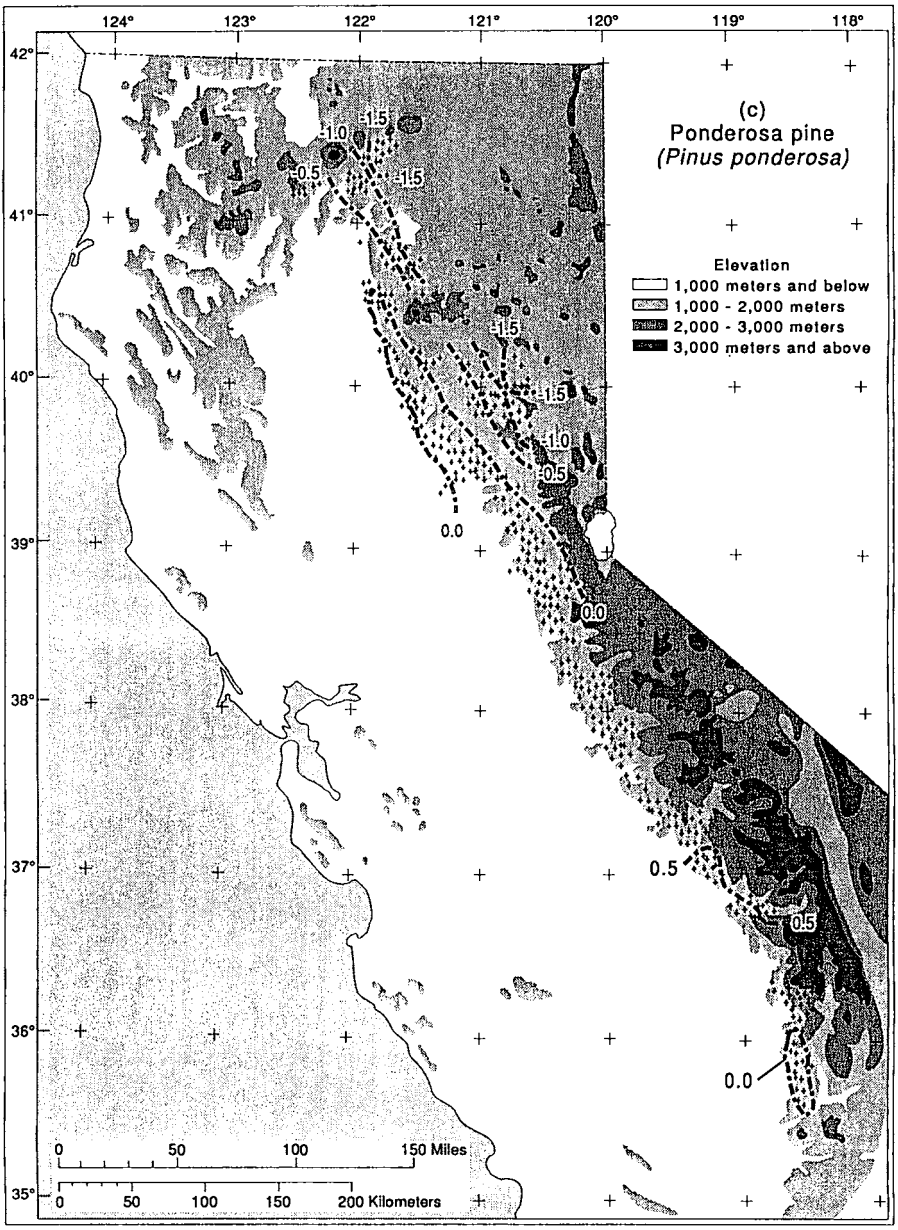
124°
123°
122°
121°
120°
119°
118°
42°
41°
40°
39°
38°
37°
36°
35°
(c)
Ponderosa pine
(Pinus ponderosa)
Elevation
1,000 meters and below
1,000 - 2,000 meters
2,000 - 3,000 meters
3,000 meters and above
1.5
1.0
0.5
0.0
0 50 100 150 Miles
0 50 100 150 200 Kilometers

Fig. 1c

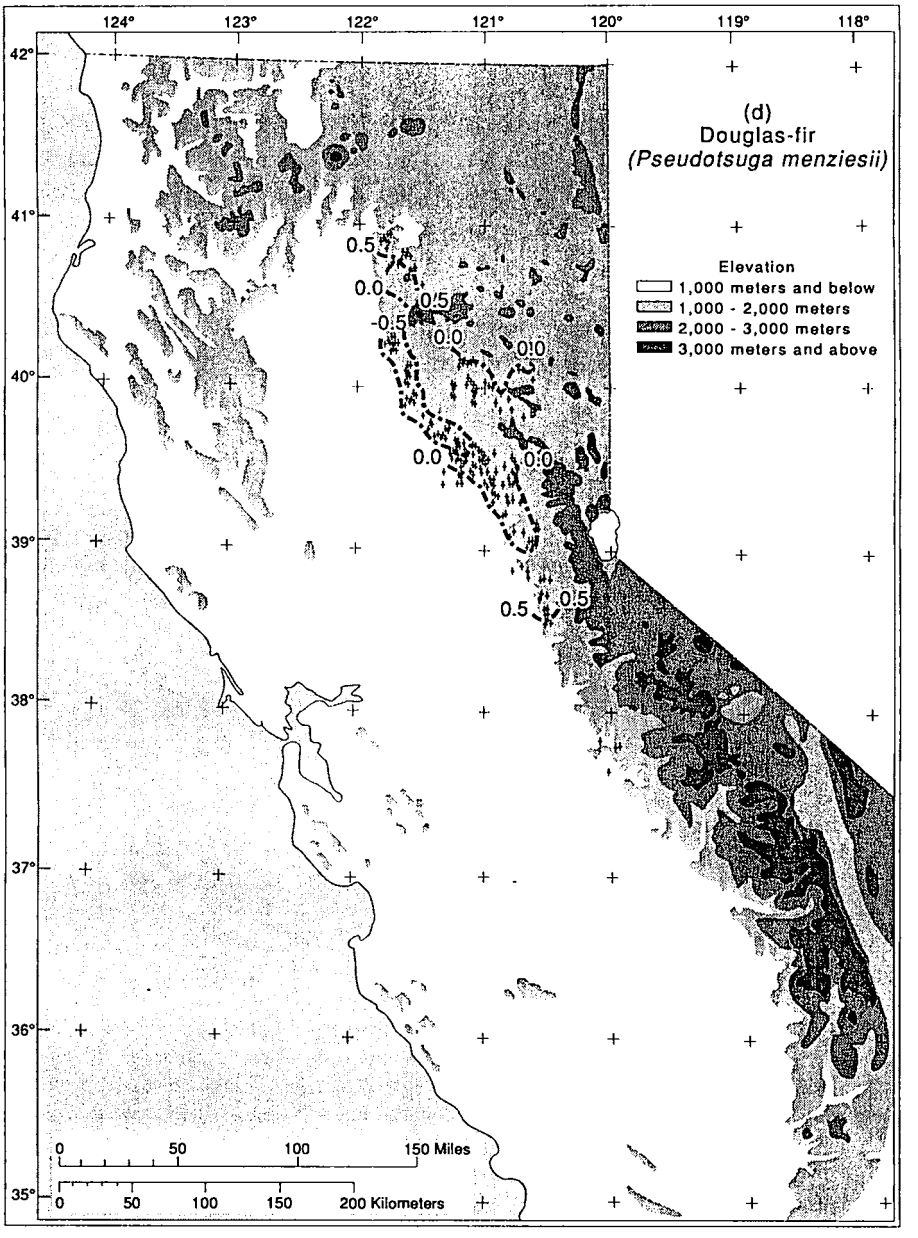

Fig. 1d

Fig. 1. Multi-locus contour plots of the first vector from canonical trend-surface models in four conifer species of the mixed conifer zone in the Sierra Nevada Mountains: *a*) white fir; *b*) sugar pine; *c*) ponderosa pine; and *d*) Douglas-fir. Contour intervals are in 0.50 standard deviation units.

measures of location formed the basis for the second-order trend-surface equations.

Geographic patterning was strongest for Douglas-fir and weakest for white-fir, both on the basis of the amount of variation described by the model in the first vector (43% vs 14%) and the amount of variation described by the first three vectors (63% vs 28%) (Table 1). In spite of the weak patterning in white fir, a multivariate test for departures from equilibrium (Morrison 1990: 292) was highly significant even though a multiple locus test of linkage disequilibrium (Brown 1984) was not. There is striking similarity in pattern for the first vectors among all four species (Fig. 1a—d). Canonical analyses of the trend-surface equations indicated that all are saddle-shaped surfaces (Box and Draper 1987: 346—350),

Table 1. Single- and multi-locus characteristics of variation in four conifer species in California.

Region/species	N	Nr loci	H_T	Diff. lat.	R_1^2	Signif. vectors	R_T^2	% Correct classif.
Sierra Nevada								
Abies concolor	373	10	0.20	2.8	0.14	2	0.28	0.34
Pinus lambertiana	167	12	0.42	4.8	0.31	2	0.44	0.53
Pinus ponderosa	516	32	0.19	6.1	0.25	4	0.40	0.50
Pseudotsuga menziesii	158	11	0.37	3.5	0.43	2	0.63	0.53
North coast								
Pseudotsuga menziesii[1]	315	27	0.21	3.8	0.17	0	0.35	0.52
Pseudotsuga menziesii[2]	374	11	0.34	2.0	0.14	2	0.24	0.37

Shown are the total sample size (N), number of loci in the analysis, expected total heterozygosities (H_T) of the loci in the analysis, latitudinal differences between the northern- and southern-most samples (in degrees), the proportion of variation in a linear combination of alleles described the the trend-surface model in the first canonical vector R_1^2, the total amount of variation described the first three canonical vectors R_T^2, the number of statistically significant vectors, and the percent of individuals correctly classified to multi-vector contour intervals by discriminant analysis.

R_1^2 = the proportion of variation described by the trend-surface model in the first canonical vector. These are the adjusted (unbiased) estimates (SAS Institute Inc. 1985) and are presented so that the data of varying sample sizes and number of parameters in the models can be compared directly.

$R_T^2 = \Sigma E_i/(1 + \Sigma E_i)$, where E_i is the i'th eigenvalue and $E_i = R_i^2/(1 - R_i^2)$ and R_i^2 = the unbiased R^2 for i'th vector. Eigenvalues are the multivariate equivalent of the sums of squares attributed to the model and the error sums of squares. These are summed over the first three vectors, assuming that the error terms are the same.

[1] Klamath National Forest samples (also includes samples from the Mendocino and Six Rivers NF).

[2] Mendocino, Trinity, and Shasta NF samples.

with the long axis of the saddle oriented in a northwest/southeast direction. In all species, the greatest rate of change in multi-locus frequencies in the first vector is along the northeast/southwest axis and the next largest on the northwest/southeast axis (cf Fig. 1a—d), whereas directions of greatest change in the second or third vectors usually include elevation. Such saddle-shaped forms of the allozyme patterns were also shared with those of growth traits from progeny tests in California (Kitzmiller 1990) whereby growth decreases along the Sierras to the northwest (increasing latitude and longitude) from the location of the growth maximum at a site, but increases to the southwest (decreasing latitude, increasing longitude, and decreasing elevation).

Multi-vector contour interval groups were constructed by dividing each the first three vectors into intervals above and below the mean score, which forms eight contour intervals in six-dimensional space (three geographic and three multi-locus dimensions). These groups, mapped over the sample ranges, show similarity in form among the four species (Fig. 2a—d). The proportions of correctly classified individuals into the eight groups by discriminant analysis are high relative to the random expectation of the percentage correctly classified, or 1/g, where g is the number of groups (Table 1). Because seven or eight groups are designated in each analysis, the random classification rate is thus 0.14 or less and the observed rates are substantially above this level. Although these proportions are biased upwards because the size of each group is small compared to the number of allelic variables, they do represent the degree of multi-locus distinctiveness in each group and do not necessarily follow trends in clinal differentiation among groups.

Patterns are also similar in all species near Mt. Lassen and southward to Lake Almanor (approx. 100 km NW of Lake Tahoe, Fig. 2a—d). The Sierra crest is more diffuse around 40 °N latitude, and the region is in a transition zone between the marine climatic influences of the Sierra Nevada west slope and the continental climate of the Great Basin; allozyme patterns tend to parallel that transition.

In contrast to the Sierra Nevada populations, multi-locus geographic variation in Coast Range Douglas-fir was weak, even though the sample spanned the Kalamath and Mendocino National Forests, ranging from the Oregon border to Clear Lake (which is approximately 130 km north of San Francisco) (Conkle and Westfall 1987; Kitzmiller 1990). The R^2 for the first vector was 0.43 in the Sierra Nevada populations and was 0.17 for populations in the Coast Range; R^2's for the first three vectors were 0.63 and 0.35, respectively (Table 1). Moreover, none of the vectors in the North Coast sample were statistically significant. These results parallel those of Merkle and others (1988) who described statistically significant,

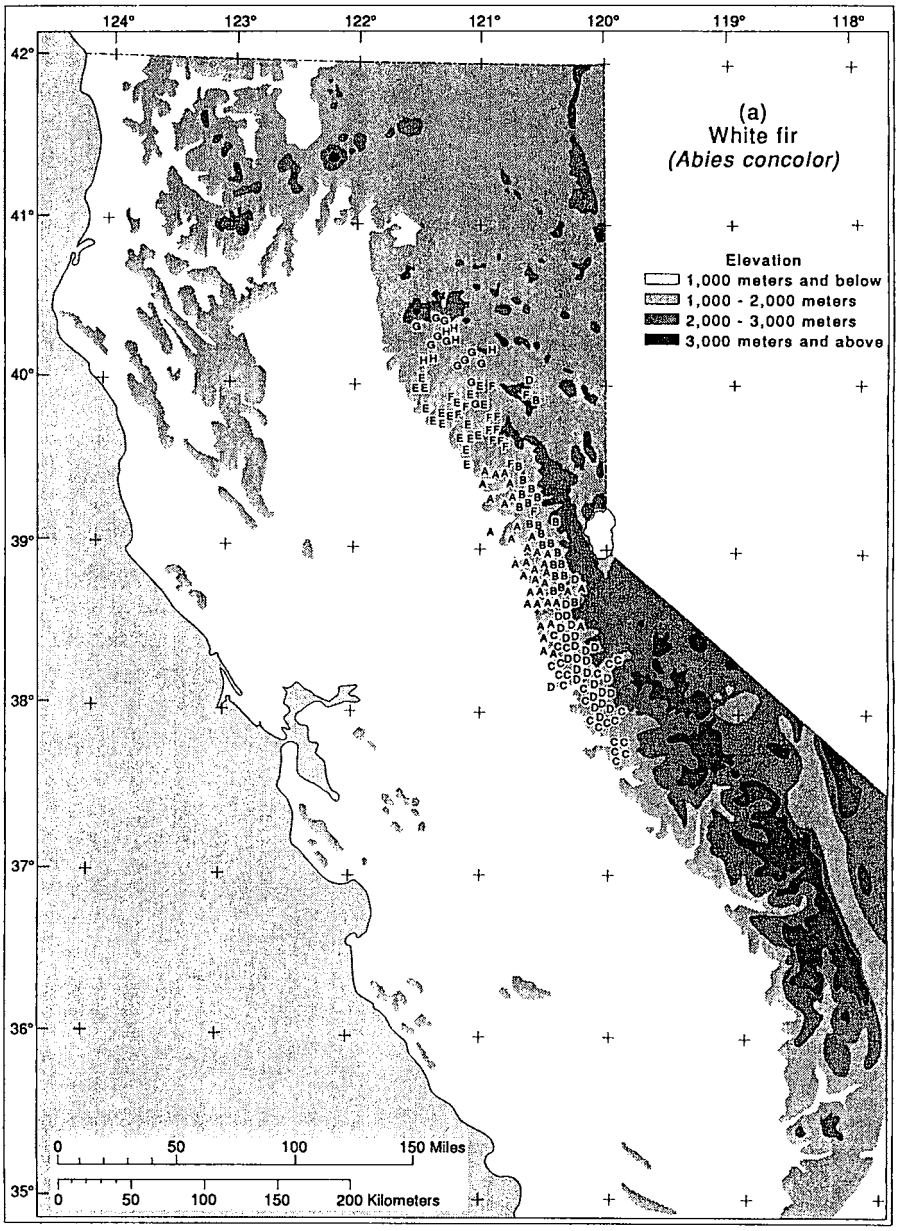
124°
123°
122°
121°
120°
119°
118°
42°
41°
40°
39°
38°
37°
36°
35°
(a)
White fir
(Abies concolor)
Elevation
1,000 meters and below
1,000 - 2,000 meters
2,000 - 3,000 meters
3,000 meters and above
0
50
100
150 Miles
0
50
100
150
200 Kilometers

Fig. 2a

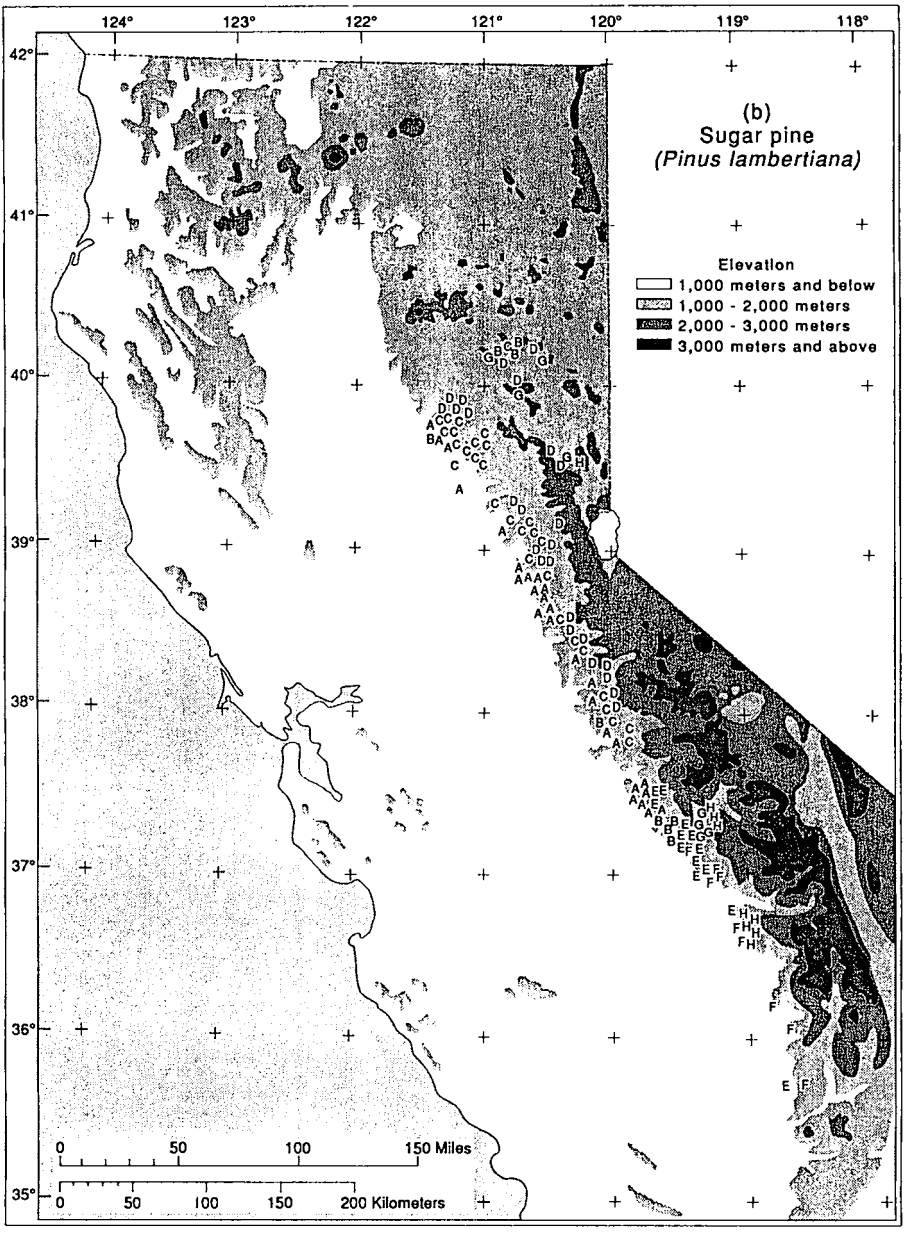
124°
123°
122°
121°
120°
119°
118°
42°
41°
40°
39°
38°
37°
36°
35°
(b)
Sugar pine
(Pinus lambertiana)
Elevation
1,000 meters and below
1,000 - 2,000 meters
2,000 - 3,000 meters
3,000 meters and above
0
50
100
150 Miles
0
50
100
150
200 Kilometers

Fig. 2b

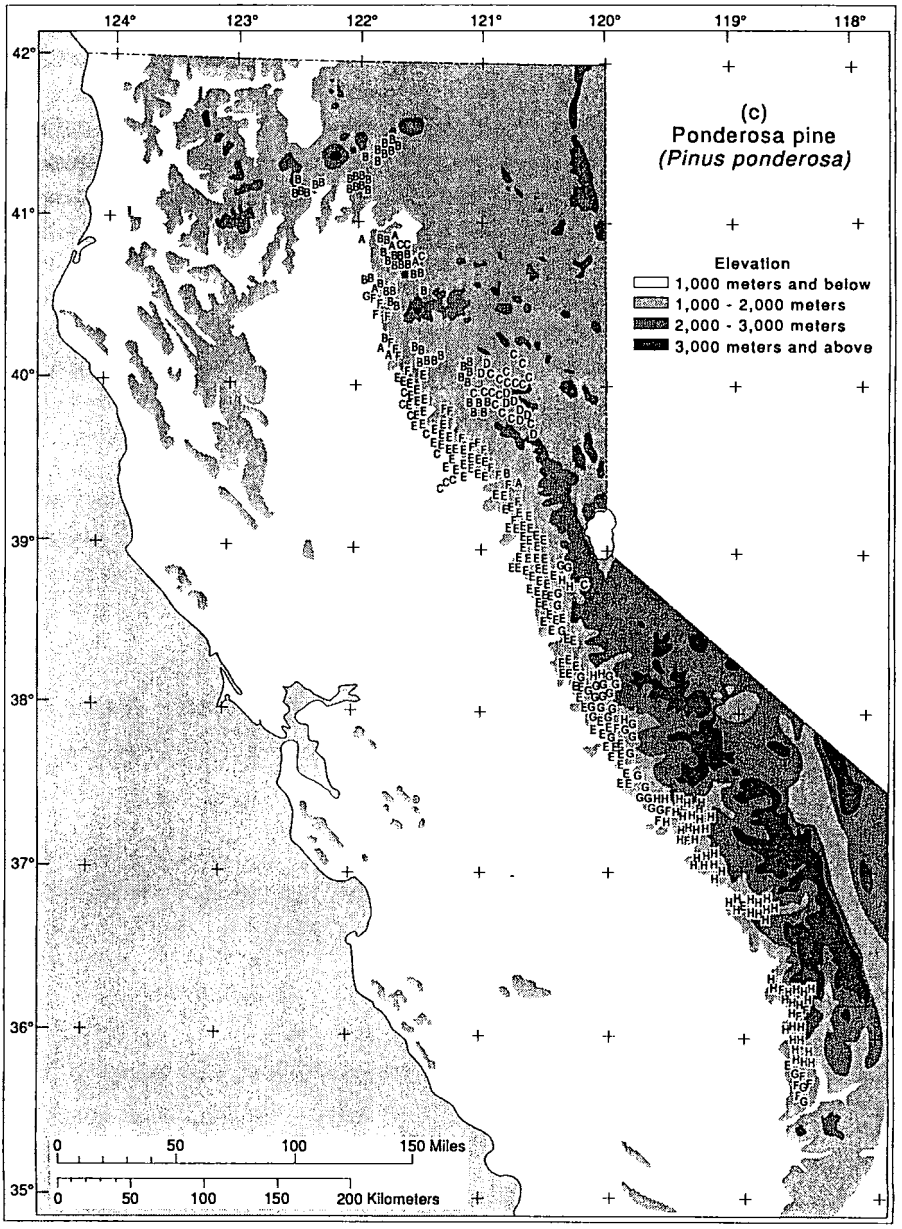

Fig. 2c

Fig. 2. Multi-locus contour class intervals of the first three vectors from canonical trend-surface models in four conifer species of the mixed conifer zone in the Sierra Nevada Mountains: *a*) white fir; *b*) sugar pine; *c*) ponderosa pine; and *d*) Douglas-fir. Contour intervals for each vector are above and below the mean. Class A contains those individuals

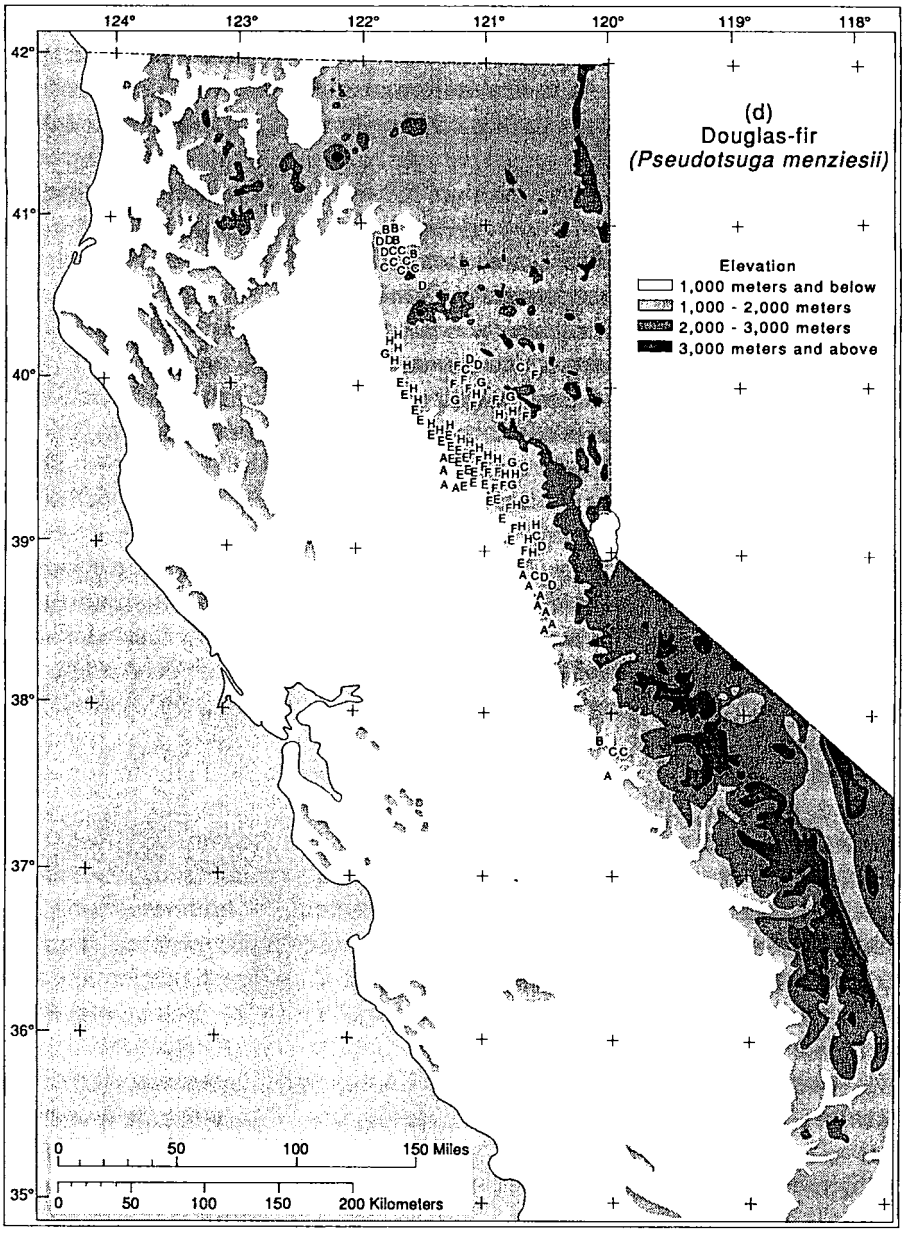

Fig. 2d

below the mean for all three vectors, class B individuals below the mean for the first two vectors and above the mean in the third, and so on to class H, which contains individuals above the mean in all three vectors.

but weak patterning in populations from Southern Oregon. The strongest pattern in multi-locus frequencies in the California North Coast data was east/west with some elevational differentiation, which follow existing breeding zone subdivisions for Douglas-fir in the region (Kitzmiller 1976; Kitzmiller 1990). We also analyzed patterns in Douglas-fir bud tissue allozymes of parent trees from the Trinity and Mendocino National Forests, plus those of the Shasta NF to the east. Levels of differentiation were similar to those in the Klamath data: the R^2 for the first vector was 0.14 and for the first three combined, 0.24. Moreover, geographic patterns in the area of overlap between the two datasets were nearly identical.

In these allozyme data, the correlation between an allele and the canonical model only occasionally exceeds 0.25. So, although very small proportions of the variability in any one allele is associated with geography, the aggregate pattern is much stronger (Table 1). Therefore, these loci and their alleles behave much in the way as that expected for quantitative trait loci.

Because large numbers of the resolved loci are not common to all of the analyses, we cannot easily test frequencies of associations and make generalizations regarding loci that pattern among all four species and regions. However, some of the enzymes associated with glycolysis and the Krebs cycle (ACO, IDH, MDH, 6PGDH, and PGM), and LAP2 and GOT pattern in most of the samples.

Geographic patterns in allozymes approximate those patterns for morphological traits, but matched comparisons are difficult to make. Even though progenies from the trees subjected to allozyme analyses are planted in tests, they are split into subsets and planting sites, and genotype-by-environment interactions and the allocation of families among sets make comparisons difficult. In spite of these encumbrances, current progeny test results in Klamath Mountains sites suggest that geographic differentiation in most family sets is on the same order as that suggested by the allozyme data (Kitzmiller 1990).

We have also compared allozyme and morphological patterns in ponderosa pine, where parent trees in the allozyme analysis described above were assessed in a three-year nursery test. Open-pollinated family means of 19 morphological and growth traits were analyzed by the same procedures used in the allozyme study. The trend-surface model in the first vector described 26% of the additive genetic variance in a linear combination of nursery traits, and the first three vectors described 36% of the variation, which are very similar in magnitude to proportions found in the allozyme data (Table 1). Three statistically-significant, saddle-shaped vectors in the nursery data described 44% of the variation in the four allozyme vectors, whereas four allozyme vectors accounted for 62% of the

variation in the nursery vectors. Contour intervals, classification percentages from discriminant analysis, and transfer risks were similar in both data sets and together suggest that California Seed Zones in the central Sierra (Buck and others 1970) are conservative for ponderosa pine and could be consolidated.

An example: breeding zone designation in Sierra Nevada Douglas-fir

Results from the analyses of the Sierra Nevada Douglas-fir sample described above contributed to the development of breeding zones, which embrace much of the commercial range of the species in this region of California (Kitzmiller 1976). Allozyme variation in these data is among the greatest of the four Sierran species: total heterozygosity is 0.37, and R^2's for the first and the first three canonical vectors are 0.37 and 0.63, respectively. The rate of change in multi-locus frequencies with geographic distance is largest in the northeast/southwest direction and, at right angles to this direction, is much lower (Fig. 1d). The direction of maximum change in the second vector is similar to that in the first, but the second largest rate of change is by elevation.

The initial step in applying these results to the formation of zones was to subdivide the first two statistically significant vectors each into theoretical thirds: 0.5 standard deviations (SD) below, ± 0.5 SD around, and 0.5 SD above the mean. This subdivision resulted in nine classes or joint contour intervals in five-dimensional space (Table 2 and Fig 3; group G contained no trees within its interval). Maximum differences in multi-locus genotypes were predicted to be between groups A and I.

Subdivisions among the groups are largely by latitude and elevation. Because of the saddle-shaped structure of the data, groups A and B are at both the northern and southern extremes of the sample (Figs. 1d and 3). However, these two groups are below 1470 m in the north and above that elevation in the south. Rates of change in pattern are most rapid between the mid elevations on the west slope of the northern Sierras (Group I) to the southern Cascades (A—C; Fig. 3). Groups D through F subdivided by elevation: D and E divides into contour intervals at 1070 m, and E and F at 1470 m. Groups H and I are located in the elevation band described by E. A heterogeneous assemblage of groups are found in seed zone 523, again suggesting a transition between the northern Sierra to the west and the Cascades to the northwest. Discriminant analysis shows similarity between groups D and E, E and F, and E and H (Table 2).

Breeding zones were developed from these allozyme patterns, ecological and geoclimatic information, and seed production needs. The major demand for seed is anticipated in California Seed Zones 524, 525, and

Table 2. Classification percentages for eight multi-locus allozyme groups in Sierra Nevada Douglas-fir

From class	Percent classified into class								Total number
	A	B	C	D	E	F	H	I	
A	43	14	29		14				7
B		62	12	12	12			3	8
C	8		62	8	15		8		13
D	4	4	7	56	11		15	4	27
E			3	14	52	9	21	2	58
F		8	8		17	33	8	25	14
H				11	30		52	7	27
I						17	17	67	6
Total									158
Percent correctly classified									

Table 3. Classification percentages for five candidate seed zones in Sierra Nevada Douglas-fir

From class	Percent classified into class					Total number
	522	523	524/526L	524/526M	524/526H	
522	53	13	3	13	17	30
523		63		26	11	19
524/526L	10	14	52	19	5	21
524/526M	3	15	10	69	3	71
524/526H	10	10		30	50	10
Total						151
Percent correctly classified						62

Classes 522 and 523 refer to the respective California Seed Zones. Classes 524/526L, M, and H refer to the low (< 1070 m), middle (> 1070 < 1470 m), and high (> 14870 m) elevation subdivisions of the pooled California Seed Zones 524, 525, and 526.

526 (Fig. 3; Buck and others 1970), which compose most of an existing breeding zone (Kitzmiller 1976). The allozyme results support that grouping, but also suggest three subdivisions by elevation: below 1070 m, between 1070 and 1470 m, and above 1470 m. A fourth zone was set to encompass Seed Zone 523, with no current subdivisions.

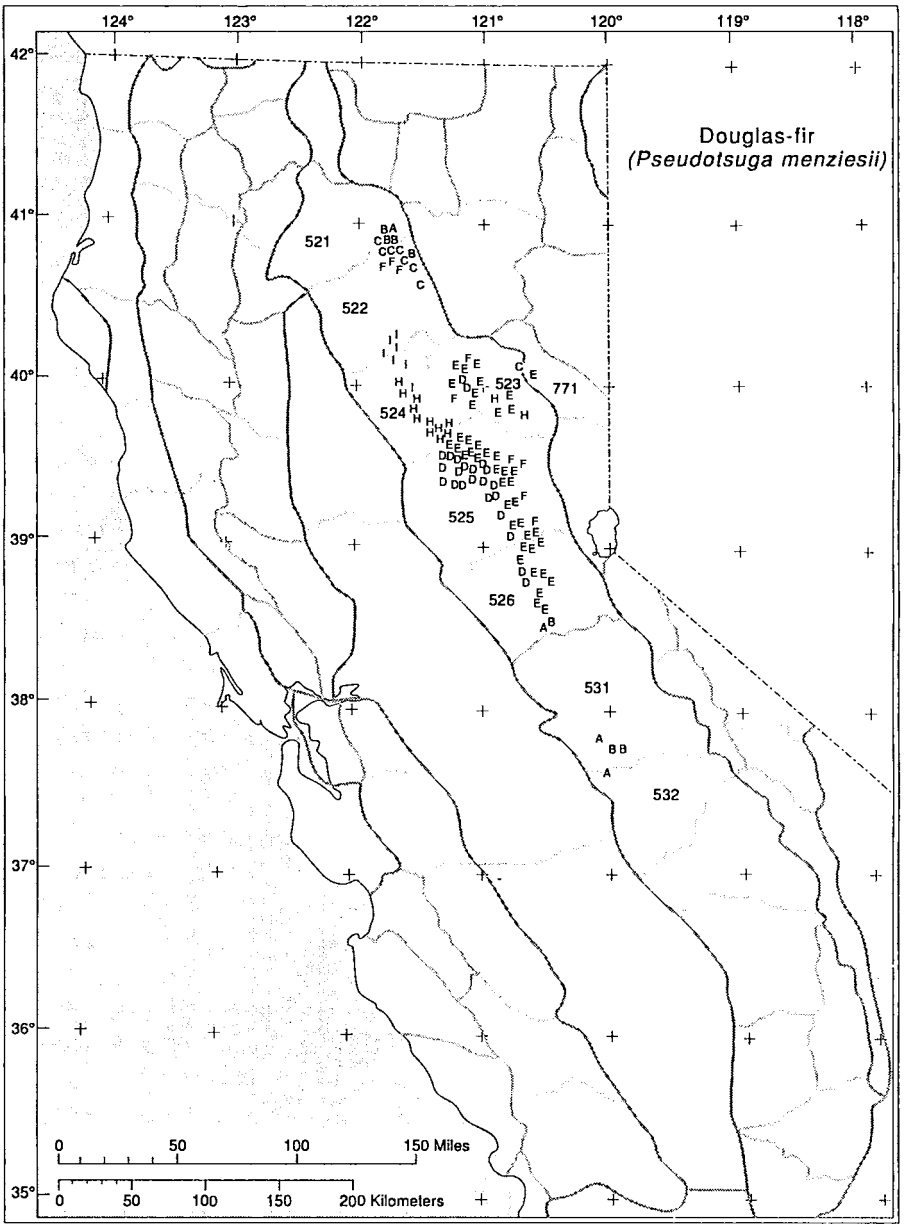

Fig. 3. Multi-locus classes in Sierra Nevada Douglas-fir, developed from intervals of the first two canonical vectors. Groups A-C, D-F, and G-I represent contour intervals in the first vector and the three subdivisions in each (e.g., A, B, and C) represent subdivisions in the second (see the text).

Discriminant allocation results for these zones plus that for Seed Zone 522 are shown in Table 3. The proposed zones formed moderately cohesive classes in comparison to the groups formed by contour intervals (Tables 2 and 3). For the proposed central Sierra zone (Zones 524—526), misclassifications were highest among adjacent elevational groups. Moreover, individuals from the northermost zone, 522, tended to classify into the middle to high elevation classes in the central Sierra group and in the eastern zone, 523.

Future prospects

Statistical stability of the analyses

An important issue to resolve is the minimum number of loci necessary to represent patterns. A procedure similar to that used by Thorpe (1985, 1987) could be applied to establishing minimum numbers. Thorpe determined optimum numbers of traits by bootstrap selection of varying numbers of traits (Efron 1979) and then correlating patterns at the differing trait numbers with the pattern generated from all the traits. In this work, using independent metric traits that were significantly correlated with geography, he found that about 10 randomly selected traits were sufficient for simple clines and 20 were needed to describe complex patterns.

An additional issue is the stability of the patterns, which again can be addressed by adapting procedures used in Thorpe (1987). A measure of stability can be assessed by repetitively resampling the data by bootstrapping, approximately 100 times. Then, for each sample, we compute predicted scores of the canonical trend-surface model and the correlation between predicted scores for the original model and those of the bootstrapped sample. From these bootstrapped samples, we compute a mean correlation with the pattern in the original dataset and 95% confidence interval. Limitations on the widespread use of this method are the daunting demands on time and computing resources.

Further analytical approaches

Spatial autocorrelation methods have received limited attention in forest genetics (Epperson 1990; Epperson 1992, this issue, pp. 257—278). A potentially useful approach is the Mantel test (Manly 1986) of the correlation between genetic and geographic distance matrices, using the residuals

from trend surfaces (Bocquet-Appel and Sokal 1989). Significant correlations would suggest localized population structure, perhaps related to factors not considered in the original trend surface models.

DNA polymorphisms

Recent surveys of restriction fragment length polymorphisms (RFLPs) of nuclear and plastid DNA suggest that a greater number of polymorphic loci can be obtained than with allozymes (Neale 1992, this issue, pp. 391—407). If so, then additional loci, some representing different parts of the genome than allozyme loci, would be available for studies of population structure. However, if such polymorphisms only result from mutations in the third codon position or from non-coded regions, their utility in describing geographic patterns will be limited.

Conclusions

In contrast to single locus results, multi-locus analyses reveal statistically significant geographic patterns among tree populations. These multi-locus surfaces tend to follow major geoclimatic patterns in California and also parallel patterns in metric traits. Consequently, these allozyme data can contribute useful information to the formation of breeding zones.

Perhaps the most important questions raised by our results and those from CVA pertain to the causes of the patterns. The geographic patterns may result from chance and historical events. But if they result from selection or selection and drift, can we make global statements about genetic variation in populations within a specific geographic region? If so, can mechanistic models be constructed to explain the patterns, rather than the empirically based trend-surface models?

Acknowledgments

J. H. Kitzmiller enthusiastically supported the work described here. G. Parks, P. Stover, and S. Samman supplied information about the parent trees. We thank P. Hodgskiss, C. Niebling, and S. T. Friedman who supplied the electrophoretic data. And we thank R. K. Scagel and J. Baldwin for helpful suggestions and F. T. Ledig, C. I. Millar, the associate editor, and two anonymous reviewers for constructive comments.

References

Adams, W. T. 1992. Gene dispersal within forest tree populations. This issue (pp. 217—240).

Adams, T. and Campbell, R. K. 1981. Genetic adaptation and seed source specificity, pp. 78—85. In: Hobbs, S. D. and Helgerson, O. T, (Eds) Reforestation of Skeletal Soils, Forest Res. Lab., Oregon State Univ., Corvallis, OR.

Allard, R. W., Babbel, G. R., Clegg, M. T. and Kahler, A. L. 1972. Evidence for coadaptation in *Avena barbata*. Proc. Nat. Acad. Sci. USA 69: 3043—3048.

Archie, J. W., Simon, C. and Martin, A. 1989. Small sample size does decrease the stability of dendrograms calculated from allozyme-frequency data. Evolution 43(3): 678—683.

Bergmann, F. 1975. Adaptive acid posphatase polymorpism in conifer seeds. Silvae Genet. 24(5—6): 175—177.

Bocquet-Appel, J.-P. and Sokal, R. R. 1989. Spatial autocorrelation analysis of trend residuals in biological data. Syst. Zool. 38(4): 333—341.

Box, G. E. P. and Draper, N. R. 1987. Empirical Model-building and Response Surfaces. J. Wiley & Sons, New York.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. 1978. Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building. J. Wiley & Sons, New York.

Brown, A. D. H. 1984. Multilocus organization of plant populations, pp. 159—169. In: Wöhrmann, K. and Loeschcke, V. (Eds) Population Biology and Evolution. Springer-Verlag, Berlin.

Buck, J. M., Adams, R. S., Cone, J., Conkle, M. T., Libby, W. J., Eden, C. J. and Knight, M. J. 1970. California Tree Seed Zones. USDA Forest Service.

Campbell, R. K. 1984. Procedures for determining the biological limits of breeding zones in the Pacific Northwest, pp. 24—33. In: Proc. Progeny testing, servicewide genetics workshop, Charleston, SC, December 5—9, 1983. USDA Forest Service, Timber Management.

Campbell, R. K. 1986. Mapped genetic variation of Douglas-fir to guide seed transfer in southwest Oregon. Silvae Genet. 35(2—3): 85—96.

Campbell, R. K. and Sugano, A. J. 1987. Seed zones and breeding zones for sugar pine in southwestern Oregon. USDA Forest Serv., Res. Pap. PNW-RP-383, 18 pp.

Campbell, R. K., Pawuk, W. A. and Harris, A. S. 1989. Microgeographic genetic variation of Sitka spruce in southeastern Alaska. Can. J. For. Res. 19: 1004—1013.

Cavalli-Sforza, L. L. and Edwards, A. W. F. 1967. Phylogenetic analysis: Models and estimation procedures. Evolution 21: 550—570.

Conkle, M. T. 1974. Enzyme poymorphism in forest trees, pp. 95—105. In: Proc. third North Amer. for. biol. workshop, Colorado State University, 9—12 Sept., 1974. College of Forestry and Natural Resources, Colorado State University, Fort Collins, CO.

Conkle, M. T. and Westfall, R. D. 1984. Evaluating breeding zones for ponderosa pine in California, pp. 89—98. In: Progeny testing, proc. servicewide genetics workshop, Charleston, SC, December 5—9, 1983. USDA Forest Service, Timber Management.

Conkle, M. T. and Westfall, R. D. 1987. Douglas-fir isozyme test results. USFS, Pac. SW For. & Range Expt. Stn., Unpublished report to the USFS R-5 Regional Tree Improvement Program, 5 pp.

Conkle, M. T. and Westfall, R. D. 1988. White fir report. USFS, Pac. SW For. & Range Expt. Stn., Unpublished report to the USFS R-5 Regional Tree Improvement Program, 13 pp.

Dickinson, T., Knowles, P. and Parker, W. H. 1988. Data set congruence in northern Ontario tamarack (*Larix laricina*, Pinacea). Syst. Bot. 13(3): 442—455.

Dorman, K. W. 1976. The Genetics and Breeding of Southern Pines. USDA, Agric. Handbook 471, 407 pp.

Draper, N. R. and Smith, H. 1966. Applied Regression Analysis. John Wiley & Sons, Inc., New York, 407 pp.

Dunn, G. and Everitt, B. S. 1982. An Introduction to Mathematical Taxonomy. Cambridge University Press, Cambridge.

Efron, B. 1979. Bootstrap methods: another look at the jacknife. Ann. Stat. 7: 1—26.

El-Kassaby, Y. A. 1990. Genetic Variation within and among conifer populations: review and evaluation of methods, pp. 59—74. In: Hattemer, H. H., Fineschi, S., Cannata, F. and Malvoti, M. E. (Eds) Biochemical Markers in the Population Genetics of Forest Trees. SPB Academic Publishing, The Hague.

Ellstrand, N. C. 1992. Gene flow among seed plant populations. This issue (pp. 241—256).

Epperson, B. K. 1992. Spatial structure of genetic variation within populations of forest trees.This issue (pp. 257—278).

Epperson, B. K. 1990. 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 Inc., Sunderland, MA.

Falkenhagen, E. R. 1985. Isozyme studies in provenance research of forest trees. Theor. Appl. Genet. 69(4): 335—347.

Feret, P. P. and Bergmann, F. 1976. Gel electrophoresis of proteins and enzymes, pp. 49—77. In: Miksche, J. P. (Ed) Modern Methods in Forest Genetics. Springer-Verlag, New York.

Gittins, R. 1968. Trend-surface analysis of ecological data. J. Ecology 56(3): 845—869.

Gittins, R. 1985. Canonical Analysis. A Review with Applications in Ecology. Springer-Verlag, New York.

Griffin, J. R. and Critchfield, W. B. 1982. The distribution of forest trees in California. USDA For. Serv., Research Pap. PSW-82, 118 pp.

Guries, R. P. 1984. Genetic variation and population differentiation in forest trees, pp. 119—131. In: Proc. eighth N. Amer. for. biol. workshop, Utah State Univ., Logan, UT, Jul. 30—Aug. 1, 1984. Dept. For Res., Utah State Univ., Logan, UT.

Guries, R. P. and Ledig, F. T. 1981. Genetic structure of populations and differentiation in forest trees, pp. 42—47. In: Proc. symposium on isozymes of North American forest trees and forest insects, Berkeley, CA, July 27, 1979. USDA For. Serv. PSW For. & Range Expt. Stn., Gen. Tech. Rep. PSW-48.

Hamrick, J. L. and Godt, M. J. W. 1990. 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 Inc., Sunderland, MA.

Kempthorne, O. 1969. An Introduction to Genetic Statistics. Iowa State Univ. Press, Ames, IA.

Kinloch, B. B., Westfall, R. D. and Forrest, G. I. 1986. Caledonian Scots pine: origins and genetic structure. New Phytol. 104: 703—729.

Kitzmiller, J. H. 1976. Tree Improvement Master Plan for the California Region. USDA Forest Service, San Francisco, 123 pp.

Kitzmiller, J. H. 1990. Genetic variation and adaptability of Douglas-fir in northwestern California. In: Proc. Joint Mtg. W. For. Genet. Assoc. & IUFRO Working Parties S2.02—05. 06, 12, & 14, Douglas-fir, Contorta Pine, Sitka Spruce, & Abies Breeding & Genetic Resources, Olympia, WA, Aug. 20—21, 1990.

Knowles, P. 1985. Comparison of isozyme variation among natural stands and plantations: jack pine and black spruce. Can. J. For. Res. 15: 902—908.

Ledig, F. T. 1988. The conservation of diversity in forest trees. BioSci. 38(7): 471—479.

Lee, P. J. 1969. The theory and application of canonical trend surfaces. J. Geology 77(3): 303—318.

Loveless, M. D. 1992. Isozyme variation in tropical trees: patterns of genetic organization. This issue (pp. 67—94).

Lundkvist, K. 1979. Allozyme frequency distributions in four Swedish populations of Norway spruce (*Picea abies* K.). I. Estimations of genetic variation within and among populations, genetic linkage and a mating system parameter. Hereditas 90: 127—143.

Manly, B. F. J. 1986. Multivariate Statistical Methods. A Primer. Chapman and Hall, New York, 159 pp.

Merkle, S. A., Adams, W. T. and Campbell, R. K. 1988. Multivariate analysis of allozyme variation patterns in coastal Douglas-fir from southwest Oregon. Can. J. For. Res. 18: 181—187.

Millar, C. I., Strauss, S. H., Conkle, M. T. and Westfall, R. D. 1988. Allozyme differentiation and biosystematics in the California closed-cone pines (Subsection *Oocarpae* Little & Critchfield, Genus *Pinus*). Syst. Bot. 13(3): 351—370.

Millar, C. I. and Westfall, R. D. 1992. Allozyme markers in forest genetic conservation. This issue (pp. 347—371).

Mitton, J. B. 1992. The dynamic mating systems of conifers. This issue (pp. 197—216).

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.

Moran, G. F. and Adams, W. T. 1989. Microgeographical patterns of allozyme differentiation in Douglas-fir from southwest Oregon. For. Sci. 35(1): 3—15.

Morrison, D. F. 1990. Multivariate Statistical Methods. McGraw-Hill, New York.

Muona, O. 1990. 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 Inc., Sunderland, MA.

Namkoong, G. 1969. Nonoptimality of local races, pp. 149—153. In: Proc. tenth south. for. tree improv. conf. Texas For. Serv., Texas A & M Univ. Press, College Sta., TX.

Namkoong, G. 1985. Considerations of respect for future forest populations, pp. 117—121. In: Proc. N. Amer. for. commission tree improv, Study Group. The state of the art in forest genetics and breeding. Int. Meeting, Satellite to the IX World Forestry Congress, Mexico City, Jul. 5, 1985.

Namkoong, G. 1990. Forest genetics: 2050. The quick and the dead. In: Proc. Joint Mtg. W. For Genet. Assoc. & IUFRO Working Parties S2.02-05. 06, 12, & 14, Douglas-fir, Contorta Pine, Sitka Spruce, & Abies Breeding & Genetic Resources, Olympia, WA, Aug. 20—21, 1990.

Namkoong, G. and Kang, H. 1990. Quantitative genetics of forest trees, pp. 139—188. In: Janick, J. (Ed) Plant Breeding Reviews, Vol 4. Timber Press, Portland, OR.

Namkoong, G., Kang, H. C. and Brouard, J. S. 1988. Tree Breeding: Principles and Strategies. Springer-Verlag, New York.

Neale, D. B. and Adams, W. T. 1985. Allozyme and mating-system variation in balsam fir (*Abies balsamea*) across a continuous elevational transect. Can. J. Bot. 63(2): 2448—2453.

Neale, D. B., Devey, M. E., Jermstad, K. D., Ahuja, M. R., Alosi, M.C. and Marshall, K. A. 1992. Use of DNA markers in forest tree improvement research. This issue (pp. 391—407).

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583—590.

Rehfeldt, G. E. 1986. Adaptive variation in *Pinus ponderosa* from Intermountain regions. I. Snake and Salmon River Basins. For. Sci. 32(1): 79—92.

Rehfeldt, G. E. 1990a. Gene resource management: Using models of genetic variation in

silviculture, pp. 31—44. In: Proc. genetics/silviculture workshop, Wenatchee, WA. Aug 27—31, 1990, USDA Forest Service.

Rehfeldt, G. E. 1990b. Genetic differentiation among populations of *Pinus ponderosa* from the upper Colorado River Basin. Bot. Gaz. 151(1): 125—137.

SAS Institute Inc. 1985. SAS® User's Guide: Statistics. SAS Institute Inc., Cary, NC, 956 pp.

Smouse, P. E. and Williams, R. C. 1982. Multivariate analysis of HLA-disease associations, Biometrics 38(3): 757—768.

Smouse, P. E., Spielman, R. S. and Park, M. H. 1982. Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. Am. Nat. 119(4): 445—463.

Spielman, R. S. and Smouse, P. E. 1976. Multivariate classification of human populations. I. Allocation of Yanomama Indians to villages. Am. J. Hum. Genet. 28(4): 317—331.

Swofford, D. L. 1981. On the utility of the distance Wagner procedure, pp. 25—43. In: Funk, V. A. and Brooks, D. R. (Eds) Advances in Cladistics. Allen Press, Lawrence, KS.

Thorpe, R. S. 1985. Clines: Character number and the multivariate analysis of simple patterns of geographic variation. Biol. J. Linn. Soc. 26(3): 201—214.

Thorpe, R. S. 1987. Complex clines: the predictivity of complicated patterns of geographic variation portrayed by multivariate analysis. Biol. J. Linn. Soc. 31: 75—88.

Tigerstedt, P. M. A. 1973. Studies on isozyme variation in marginal and central populations of *Picea abies*. Hereditas 75: 47—60.

Wartenberg, D. 1985. Canonical trend surface analysis: A method for describing geographic patterns. Syst. Zool. 34(3): 259—279.

Weir, B. S. 1990. Genetic Data Analysis. Sinauer Associates, Inc., Sunderland, MA.

Westfall, R. D. 1991. Developing seed transfer zones. In: Fins, L. and Friedman, S. T. (Eds) Manual of Quantitative Forest Genetics (in press).

Williams, B. K. and Titus, K. 1988. Assessment of sampling stability in ecological applications of discriminant analysis. Ecology 69(4): 1275—1285.

Yeh, F. C., Cheliak, W. M., Dancik, B. P., Illingworth, K. Trust, D. C. and Pryhitka, B. A. 1985. Population differentiation in lodgepole pine, *Pinus contorta* spp. *latifolia*: a discriminant analysis of allozyme variation. Can. J. Genet. Cytol. 27: 210—218.