Analysis of Mating System Parameters and Population Structure in Douglas-fir Using Single-locus and Multilocus Methods[1]

D. V. Shaw[2] and R. W. Allard[3]

Abstract: Two methods of estimating the proportion of self-fertilization as opposed to outcrossing in plant populations are described. The first method makes use of marker loci one at a time; the second method makes use of multiple marker loci simultaneously. Comparisons of the estimates of proportions of selfing and outcrossing obtained using the two methods are shown to yield additional information about inbreeding and the genetic structure of populations. Single-locus estimates of the proportion of selfing in Douglas-fir were heterogeneous over marker loci and lower than estimates obtained with the multilocus estimator. Family structure and/ or microhabitat selection were discussed as likely causes of the biased estimates of the proportion of selfing obtained from the single-locus estimation procedure.

An important goal of research in forest genetics is the development of methods that allow for the breeding of superior stock and also provide for the long-term maintenance of adequate germplasm resources. In realizing this dual goal, knowledge of the forces that are responsible for the distribution of genotypes in natural stands of forest trees, including mating system and gene flow, will be particularly useful. Several methods are available for the quantification of parameters which specify different aspects of mating systems, family structure, and gene flow. These methods have been used to estimate inbreeding coefficients (F) and the proportions of selfing (s) and outcrossing (t) (Sorensen 1973, Rudin 1977), as well as the effects of inbreeding depression on quantitative characters (Franklin 1970).

Electrophoretically detectable loci have been particularly useful in the study of mating systems and gene flow in plant populations (Brown and Allard 1970, Clegg and others 1978) for two main reasons. First, electrophoresis often provides an abundance of markers; second, enzyme loci are frequently codominant, allowing all genotypic classes to be identified directly and thus increasing the efficiency of the estimation procedures.

This paper presents estimates of the proportions of selfing and outcrossing in two stands of Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) based on two different methods of estimation. We show that comparisons of the estimates from the two methods yield additional information about the mating system and about gene flow within and between stands of this species.

MATERIALS AND METHODS

The data we use to illustrate the estimation procedures are from a study of two stands of Douglas-fir. The two stands occupy ridge-top sites which are located about 40 kilometers east of Springfield, Oregon, and about 20 kilometers from each other. The two sites, Springfield 3 (S3) and Springfield 5 (S5), which are both at about 500-m elevation, are occupied by young stands of naturally regenerated trees. Trees on S3 are about 30 years of age and S5 contains trees of 40 to 45 years of age. The stands are composed almost entirely of Douglas-fir, with an occasional western hemlock (*Tsuga heterophylla* [Raf.] Sarg.) interspersed. *Figure 1* gives the spatial distribution of the maternal trees sampled on the two sites. Sampling followed a modification of the Nelder Design (Nelder 1962).

Seeds were germinated and both diploid embryo tissue and haploid gametophytic tissue were analyzed by starch gel electrophoresis.[4] A minimum of 7 haploid-diploid pairs were analyzed per maternal tree sampled. The following 11 loci were scored for each haploid-diploid pair (numbers in parentheses represent the number of alleles scored for each locus): Got-I (3), Got-III (3), G6pd (2), Gdh (2), To (2), Est (4), Lap-I (3), Lap-II (3), Pgi (3), Pgm-I (3), Pgm-II (3).

STATISTICAL TECHNIQUES

Single-locus Estimation

Fyfe and Bailey (1951) gave the theory of estimation of the proportion of selfing (s) and outcrossing (t = 1 -s) for the mixed mating model, using recessive markers. They used

[1]Presented at the symposium on Isozymes of North American Forest Trees and Forest Insects, July 27, 1979, Berkeley, Calif.

[2]Graduate Student, Department of Genetics, University of California, Davis, Calif.

[3]Professor of Genetics and Agronomy, University of California, Davis, Calif.

[4]Starch gel electrophoretic techniques were developed by M. Thompson Conkle. (Personal commun. 1978.)

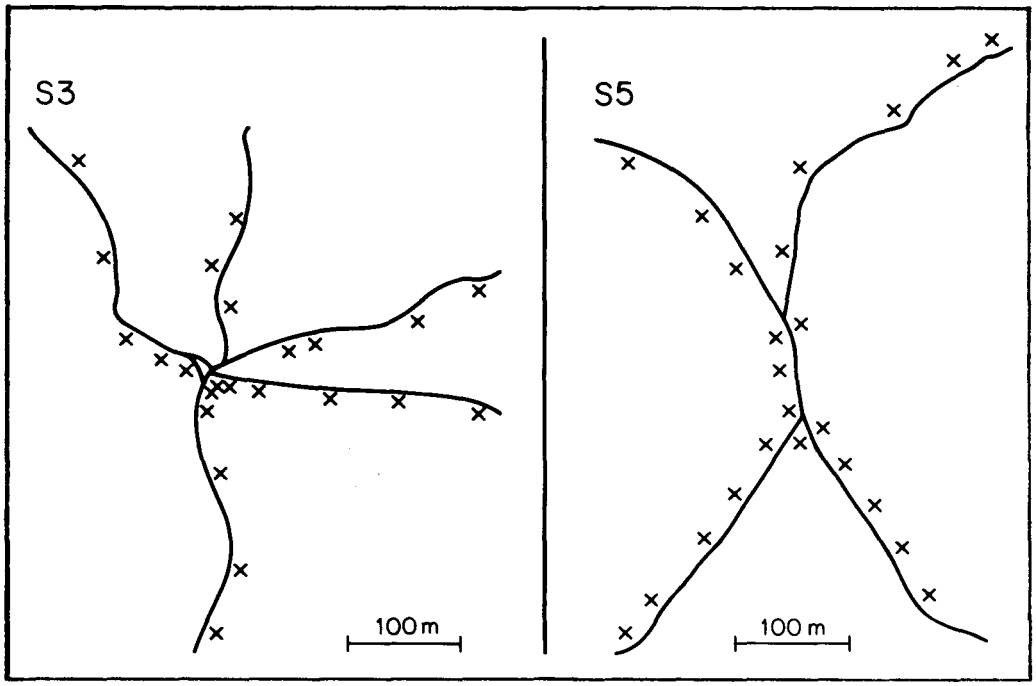

Figure 1—Spatial distribution of sample trees (*S3*) Springfield 3 site, (*S5*) Springfield 5 site.

the transformation f = (1-t)/(1+t), where f is the inbreeding coefficient, and formulated the joint maximum likelihood estimators and their variances for f and p (allelic frequencies). Brown and Allard (1970) and Clegg and others (1978) developed maximum likelihood methods for estimating outcrossing rates from the genotypic arrays of half-sib families in which the maternal parent is of unknown genotype. In these estimators the genotype of the maternal parent is inferred from the array of progeny produced by each maternal individual. This inference is affected by the maternal genotype, by allelic frequencies in the pollen pool, and by the mating system parameters themselves. When data from gametophytes are also available, as in the present case, the genotype of each maternal individual can be inferred from the gametic array of each maternal individual and an estimate of maternal genotypic frequencies can be obtained which is statistically independent of the other parameters. The estimator used in the present study was modified from that of Clegg and others (1978) to take advantage of the information provided by the gametophytic analysis. Fyfe and Bailey (1951) pointed out that an important assumption of the mixed mating model is that allelic frequencies are homogeneous in the pollen pool over the entire area sampled. A variety of causes (for example, family structure, microhabitat selection) can lead to heterogeneity of the pollen pool in natural stands and the resulting Wahlund effect leads to overestimation of the amount of selfing (Wahlund 1928). Consider the hypothetical sampling situation given by *figure 2*. Two contiguous stands, P_1 and P_2, are reproductively isolated from each other but mating occurs at random within each stand. For a diallelic locus the frequency of allele A_1 is 0.9 in P_1 and 0.1 in P_2. If an investigator were aware of the existence of the two stands, and sampled the two areas separately, the single-locus estimator would yield correct estimates of the amount of selfing ($\hat{s} = 0$) and outcrossing ($\hat{t} = 1$). If, however, the investigator were unaware that there were two stands, and sampled as if there were only one stand, the estimate of s would be 0.89 and t would be 0.11. This example demonstrates the potential of pollen heterogeneity for introduction of bias into single-locus estimates, and it points out the need for careful sampling technique. Two important additional assumptions in the mixed mating model are: (a) that the probability of an outcross will not be affected by the maternal genotype and, (b) that selection does not intervene between mating and the determination of progeny genotype distributions. The fact that significant between-locus differences in s and t are often obtained from a single sample of progeny indicates that these assumptions, like the assumption of homogeneity of the pollen pool, are also often not valid.

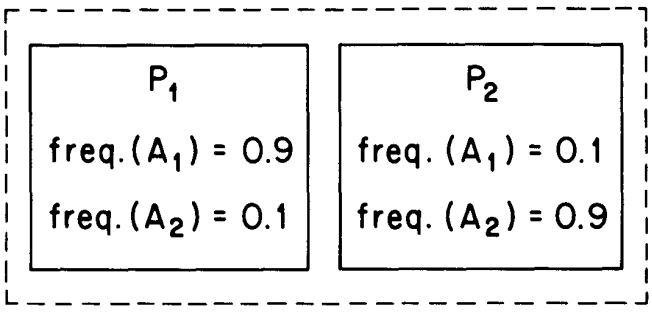

Figure 2—Hypothetical sampling situation showing actual stand boundaries (——) and sampling boundaries (- - -).

Multilocus Estimation

An alternate method for estimating proportions of outcrossing and selfing, which avoids many of the problems of single-locus estimation, has been developed.[5] The large number of marker loci provided by electrophoretic analysis makes application of this alternative method feasible in experimental situations. Briefly this method involves the identification of outcross progeny by the comparison of multilocus progeny genotypes with maternal genotypes and using the information gained from single-locus data for statistical compensation for outcrosses which are not identifiable. As more loci are considered the identification of true outcross progeny becomes nearly complete and estimates of s and t become increasingly dependent on observation and less dependent on statistical compensation. One advantage of this estimator is its insensitivity to failure of assumptions that can seriously affect single-locus estimation.

Another advantage is that comparisons of estimates derived from single-locus and multilocus techniques provide a means of separating the effects of selfing from other causes of nonrandom pollen dispersal, thus providing additional information concerning the breeding structure of populations.

RESULTS

Single-locus estimates of the proportions of outcrossing, t, for three loci in the Springfield 3 and Springfield 5 sites are presented in *table 1*. The estimates oft range of 0.66 to 0.97. This heterogeneity indicates that the estimates are affected by factors other than the mating system. Likelihood ratio tests indicate that 7 of the 9 single-locus estimates differ significantly from those expected under random mating (t = 1) and an eighth value approached significance (P = .06) (Rao 1973). Note also that the estimates obtained by combining the samples from the two sites are always lower than the mean of the two independent samples. This situation parallels the hypothetical case given in *figure 1* and demonstrates that bias is in fact introduced by combining heterogeneous samples.

Estimates of the proportion of outcrossing($\hat{t}_m$) obtained using the multilocus estimator are given below.

Stand:	$\hat{t}_m$ (S.E.)
Springfield 3	0.91 (0.03)
Springfield 5	.93 (.02)
Combined	.93 (.02)

These estimates are based on 11 loci. The estimates do not differ significantly between the two sites and they are not affected by the combination of the samples from the two sites. These estimates probably reflect the actual proportions of self and outcross progeny which germinated and survived to the time of electrophoretic assay.

[5]Shaw, D.V., A.L. Kahler, and R.W. Allard. A multilocus method for estimating mating system parameters in plant populations. (Manuscript in preparation.)

Table 1—*Single-locus estimates of outcrossing rates, $\hat{t}$ for two stands of Douglas-fir near Springfield, Oregon. Standard errors are in parentheses*

Stand	Locus PGM	LAP	EST
Springfield 3	0.91* (.05)	0.90† (.05)	0.83**(.05)
Springfield 5	.93* (.05)	.98 (.05)	.66**(.06)
Combined	.91**(.02)	.79**(.03)	.71**(.03)

†, *, **The likelihood ratio test H_0: $\hat{t}$ =1, is significant at probability levels, 0.06, 0.05 and 0.01, respectively.

Comparisons of single-locus and multilocus estimates of t are presented in *table 2*. Likelihood ratio tests were performed against the null hypothesis; H_0: $\hat{t} = \hat{t}_m$ (Rao 1973). The three loci each demonstrate a unique situation and they will be discussed separately.

1) Pgm-I: None of the three single-locus estimates of outcrossing based on this locus deviated significantly from $\hat{t}_m$ Combining the samples, therefore, did not generate any additional heterogeneity above and beyond the departure from randomness generated by the approximately 9 percent of selfing which occurred in both stands.

2) Lap-I: The single-locus estimate based on the combined samples differed significantly from $\hat{t}_m$. Thus the combining of data for the two stands introduced significant bias for this locus, indicating that allelic frequencies were not the same in the pollen pools of the two stands.

3) Est: All estimates based on this locus differed significantly from $\hat{t}_m$. This indicates that allelic frequencies in the pollen pool were not the same in the two stands and it indicates further that single-locus assumptions, probably including the assumption that the pollen pool is homogeneous within stands, do not hold.

DISCUSSION

A general expectation is that the pollen which falls on a given maternal tree is more likely to come from near neighbors than from distant trees. Also, the genotypes of neighboring trees may be more similar than those of

Table 2—*Comparison of multilocus* with *single-locus estimates of outcrossing rates for two stands of Douglas-fir near Springfield, Oregon*

Stand	$\hat{t}_m$	PGM	$\hat{t}$ LAP	EST
Springfield 3	0.91	0.93	0.90	0.83*
Springfield 5	.93	.93	.97	.69*
Combined	.93	.91	.79*	.71*

*Significant differences between $\hat{t}_m$ and $\hat{t}$ by likelihood ratio tests, at the .05 probability level.

random trees in the population due to family structure or microhabitat selection. If such nonrandom distribution of pollen occurs over maternal trees, resulting estimates of selfing, s, made using the mixed mating model, will be larger than the true amount of selfing occurring in the population. In this study we compared estimates of selfing and outcrossing made with single-locus techniques, which are affected by heterogeneity of pollen allelic frequencies, with multilocus estimates made using a multilocus technique which is insensitive to such heterogeneity. With one locus (Pgm-I) we observed no differences in single- and multilocus estimates of selfing. However, with a second locus (Lap-I) the results indicated heterogeneity of pollen gene frequencies between distant sites. With a third locus (Est) heterogeneity was indicated both between distant sites and within small areas. Comparisons of single-locus estimates with multilocus estimates indicated within-stand heterogeneity of pollen gene frequency for one locus (and thus nonrandom pollen dispersal), whereas this effect was not detected at other loci. Because different loci can give different results, it is desirable to include several loci in studies of mating systems.

Microhabitat selection, which causes individuals of similar genotype to be clustered nonrandomly within populations, is one possible cause of the heterogeneity observed in Douglas-fir. Another possible cause is the tendency of seeds to fall to the ground and grow near their maternal parent. The result of this restricted seed dispersal is that relatives, whose genotypes are more similar than those of random members of the populations, become concentrated in family clusters. Such nonrandom distribution of genotypes might lead to heterogeneity of the pollen pool. It can also lead to inbreeding: near neighbors are more likely to be relatives and they are also more likely to mate than are random members of the population.

The demonstration that nearly 10 percent of assayable embryos in Douglas-fir resulted from self-fertilization, and the indications that additional types of assortative matings may also occur, provide evidence that substantial inbreeding occurs in this species. Douglas-fir, however, shows moderate to severe inbreeding depression which raises the question of the equilibrium level of inbreeding that develops in response to the effects of these opposing forces. *Table 3,* which gives Est genotypic frequencies of adult maternal individuals and also genotypic frequencies of the progeny of these same individuals, provides information on this point. In the progeny generation homozygotes are in excess and heterozygotes are in deficiency. This is expected when mating is known to be assortative and when seedlings are assayed at a sufficiently early stage such that little selection is likely to have occurred between mating and time of assay. However, genotypic frequencies in the adult parents give a good fit to expectations based on the trinomial square rule. This pattern was repeated for the Lap-I and Pgm-I loci, using combined samples from the two stands (combined samples were used to obtain adequate sample sizes). Thus, even though significant inbreeding occurs in Douglas-fir, leading to excesses of homozygotes over expectations based on the assumption of random mating, it appears that selection removes the excess of inbred individuals. The adult populations fit Hardy-Weinberg expectations but only because the effects of nonrandom mating and selection balance and cancel each other.

In this paper we have illustrated methods of studying the mating system and breeding structure of populations of forest trees. Analysis of preliminary data from an experiment with Douglas-fir indicates that significant numbers of self-fertilized embryos survive through embryonic stages in this species, and that other phenomena, such as family structure and habitat selection, may also contribute to nonrandomness in family arrays derived from open pollination.

Knowledge of the breeding structure in natural stands of forest trees may influence the interpretation of results from open-pollinated genetic tests, and may be useful in planning plus-tree selections and germplasm conservation programs. The methods described appear to provide information useful to the planning and development of tree breeding programs.

Table 3—*Test for goodness of fit for Esterase genotypes to expectations assuming random mating for the parents and the progeny of the combined Springfield samples*

Class	Genotype					
	11	22	33	12	13	23
Progeny						
Observed	45	150	68	104	50	90
Expected	29.2	121.7	37.0	117.2	65.7	134
Chi-square$_{(3)}$ = 61.24***						
Parents						
Observed	1	14	3	9	8	14
Expected	1.76	13.2	4.1	9.6	5.8	14
Chi-square$_{(3)}$ = 1.88 n.s.						

***The chi-square goodness of fit test, with 3 degrees of freedom is significant at the .001 probability level; n.s. indicates nonsignificant deviation.

Acknowledgments

We are grateful to Dr. Rex McCullough and the Weyerhauser Company for providing financial support and for advice concerning the field aspects of the problem. We also thank Dr. M. Thompson Conkle for advice concerning electrophoretic techniques and Michael Carlson for assistance in the collection of seeds.

LITERATURE CITED

Brown, A. H. D., and R. W. Allard.
1970. **Estimation of mating systems in open-pollinated maize populations using isozyme polymorphisms.** Genetics 66:133-145.

Clegg, M. T., A. L. Kahler, and R. W. Allard.
1978. **Estimation of life cycle components of selection in an expelmental plant population.** Genetics 89:765-792.

Franklin, E. C.
1970. **Survey of mutant forms and inbreeding depression in species of the family Pinaceae.** USDA Forest Service Res. Paper SE-61, 21 p. Southeastern Forest Exp. Stn., Asheville, N.C.
Fyfe, J. L., and N. T. J. Bailey.
1951. **Plant breeding studies in leguminous forage crops. I. Natural cross-breeding in winter beans.** J. Agric. Sci. 41:371-378.
Nelder, J. A.
1962. **New kinds of systematic designs for spacing experiments.** Biometrics 18(3):283-307.
Rao, C. R.
1973. **Linear Statistical Inference and Its Applications.** (2d ed.) p. 417-418. John Wiley and Sons, New York.
Rudin, D., G. Eriksson, and M. Rasmunson.
1977. **Inbreeding in a seed tree stand of *Pinus sylvestris* L., in Northern Sweden. A study by the aid of the isozyme technique.** Research notes, Dep. Forest Genetics, Royal College of Forestry, Sweden.
Sorensen, F.
1969. **Embryonic genetic load in coastal Douglas fir, *Pseudotsuga menziesii* var. *menziesii.*** Amer. Nat. 103(932):389-398.
Sorensen, F.
1974. **Self-pollination effects on Douglas-fir and ponderosa pine seed and seedlings.** Silvae Genetica 23:135-165.
Wahlund, S.
1928. **Zasammensetzung von Populationen und Korrelationserscheinungen von Standpunkt der Verenbuugslehre aus befrachtet.** Hereditas 11:65-106.