MEASURING POLLEN CONTAMINATION IN CLONAL SEED ORCHARDS WITH THE AID OF GENETIC MARKERS

D. B. Smith and W. T. Adams

Abstract.--The availability of electrophoretic techniques has provided a large number of simply inherited genetic markers (allozymes) which can be used to estimate levels of pollen contamination in seed orchards. Using several allozyme loci, the multilocus genotypes of both the male and female gametes contributing to a conifer embryo can be determined. Because natural stands contain many more adult genotypes than do clonal orchards, they are expected to produce a much greater variety of multilocus pollen genotypes, even when there are no differences in single locus allele frequencies between orchards and background stands. The level of pollen contamination in an orchard can be estimated by analyzing a sample of orchard seed and determining the proportion of pollen gametes having multilocus genotypes which could not be produced in the orchard. This proportion is then divided by the probability that a background pollen grain has a multilocus marker (i.e., a genotype different from those that can be produced by orchard clones) to compute the level of contamination. Estimates of pollen contamination in several blocks of a Douglas-fir seed orchard in western Oregon are used to illustrate this multilocus technique. Preliminary recommendations for its use are also discussed.

Additional keywords: allozymes, migration models, pollen management, Pseudotsuga menziesii.

In order to avoid pollen contamination from non-orchard (background) sources, seed orchards should be located far from stands of the same species. In most circumstances, however, this is not practical and pollen dilution zones are often employed in an attempt to keep levels of pollen contamination low. Although few direct estimates of pollen contamination have been made, it appears that pollen dilution zones may not he as effective as intended and that pollen contamination can he a serious problem in seed orchards (Squillace 1967, Squillace and Long 1981, Friedman and Adams 1981). In addition to pollen dilution zones, other management techniques for minimizing pollen contamination have been suggested, but their effectiveness is largely unknown (Silen and Keane 1969, Denison and Franklin 1975, Bridgwater and Trew 1981).

1/ Research Assistant and Associate Professor of Forest Genetics, Department of Forest Science, Oregon State University, Corvallis, Oregon 97331. We thank Don Oliver of the Siuslaw National Forest (USDA Forest Service) and Jack Barringer and Howard Dew of Barringer and Associates, Inc. for their cooperation in this study. We also thank Bob Joly and Dr. Scott Merkle for reviewing the manuscript. Research was funded by USDA, Science and Education Administration Grant No. 59-2411-1-1-610-0.

Recommendations for seed orchard location and pollen management (e.g., size of dilution zones) have generally been based on measures of pollen dispersion and differences between orchards and background stands in floral phenology (Wang et al. 1960, Schmidt and Hamblett 1962, Silen 1962, 1963). Such measures give some indication of the relative proportions of orchard and non-orchard pollen present in an orchard at the time of pollination, but do not provide estimates of their relative effectiveness in fertilization. The availability of techniques for directly estimating pollen contamination would enable seed orchard managers to determine the need for and success of pollen management efforts.

Direct estimates of pollen contamination can be made with the aid of genetic markers and the following well known migration model:

$$Q_o = q_b m + q_o (1-m) \qquad (1)$$

where

Q_o = frequency of a marker in pollen gametes which have fertilized orchard ovules,

q_b = mean frequency of the same marker in pollen produced by background stands,

q_o = frequency of the marker in orchard pollen,

m = proportion of orchard ovules fertilized by pollen from background stands, and

1-m = proportion of orchard ovules fertilized by orchard pollen (Spiess 1977).

The rate of pollen contamination (m) is estimated by solving (1) for m and substituting estimates of marker frequencies so that:

$$\hat{m} = \frac{\hat{Q}_o - \hat{q}_o}{\hat{\bar{q}}_b - \hat{q}_o} \qquad (2)$$

Particularly useful are markers unique to background pollen sources such that (2) reduces to:

$$\hat{m} = \hat{Q}_o / \hat{\bar{q}}_b \qquad (3)$$

Estimates of pollen contamination using genetic markers have recently been reported for slash pine (Squillace and Long 1981) and loblolly pine (Friedman and Adams 1981).

In the past, application of this method has been limited by the lack of suitable genetic markers; but with the advent of electrophoresis in recent years, a large number of simply inherited genetic markers (allozymes) have become available. There are many advantages of using allozymes as genetic markers for tree improvement research problems (Adams 1982). Advantages of

particular relevance to pollen contamination studies include: (1) the ability (in conifers) to identify the genotypes of the male and female gametes contributing to an embryo, and (2) the high levels of variation at allozyme loci.

Pollen contamination estimates based on single locus genetic markers require at least moderate gene frequency differences between an orchard and the background stand; otherwise, the efficiency of estimates is low, and very large sample sizes are required to achieve a reasonable level of precision. Unfortunately, the majority of allozyme variation that has been observed in forest tree species seems to occur within populations. Gene frequency differences among populations, even those separated by large distances, are often quite small (Brown and Moran 1979, Adams 1982). The same lack of differentiation appears to be true when allozyme frequencies in seed orchards and background stands are compared. For example, gene frequencies among clones (25 to 27) in three of ten 5 acre (2.0 ha) blocks in the Beaver Creek Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] seed orchard (USDA Forest Service, Siuslaw National Forest, Oregon) and among adult trees in the surrounding natural stand are compared for three allozyme loci in Table 1. Ead orchard block represents a different breeding zone (geographical region), al of which are different from the zone within which the orchard is located. Although large differences among blocks, and between blocks and the background stand, might be expected, only limited gene frequency differences and the presence or absence of alleles in low frequency were found.

Table 1.--Allele frequencies at three allozyme loci in three blocks of the Beaver Creek Douglas-fir seed orchard and in the surrounding (background) natural stand.[a/]

Locus	Allele	Orchard block 7	Orchard block 5	Orchard block 10	Background
GOT3	1	.027	.142	.217	.078
	2	.970	.858	.749	.904
	3	.003	--	.034	.018
LAP1	2	.361	.614	.449	.482
	5	.236	.238	.296	.255
	7	.403	.148	.255	.263
IDH	1	.084	.113	.149	.168
	2	.872	.814	.820	.743
	3	.041	.026	--	.080
	4	.003	.047	.031	.009

[a/] Allele frequencies are based on the genotypes of all 25 to 27 clones in each of the three orchard blocks and on a sample of 183 (LAP1, IDH) or 57 (GOT3) trees in the background stand.

While limited gene frequency differences between orchards and background stands may preclude the use of single locus allozyme markers for estimating pollen contamination, a technique based on combining allozyme data over multiple loci shows a great deal of promise. In this paper, we introduce and illustrate this multilocus technique which can be used in clonal seed orchards even in the absence of gene frequency differences between orchards and background stands.

DESCRIPTION OF THE MULTILOCUS TECHNIQUE

The multilocus estimation technique takes advantage of the large amount of allozyme variation observed among trees within populations. It is based on the premise that a limited number of multilocus genotypes exist among the clones in any one orchard, relative to the number in background stands. Thus, even when there are no differences in allele frequencies between an orchard and a background stand, the background stand can be expected to produce a greater variety of multilocus pollen genotypes than the orchard. Furthermore, as the number of variable loci included in an analysis increases, the ability to distinguish pollen gametes produced in the background stand from those produced in the orchard will also increase. The expected proportion of orchard ovules fertilized by background pollen with multilocus markers (i.e., multilocus genotypes that are unique to the background stand) is:

$$b = md \tag{4}$$

where

m = probability that an orchard ovule is fertilized by a pollen grain from the background stand, and

d = probability that a background stand pollen grain has a multilocus marker.

The rate of pollen contamination is then estimated as:

$$m = b/d. \tag{5}$$

Application of this multilocus pollen contamination estimator (m) requires that the allozyme genotypes of all clones in an orchard be determined for each locus used in the analysis. In conifers, clone genotypes can be inferred from the allelic composition of megagametophytes in a sample of seeds, or determined directly by analyzing bud or needle tissues (Adams 1982). The multilocus genotypes of the clones are then used to generate the array of all multilocus pollen genotypes which can be produced by the orchard. For example, using 14 variable allozyme loci, the number of multilocus pollen genotypes that can be produced by the clones in each Beaver Creek block ranges from 154 to 416. In contrast, the number of multilocus gametic combinations that are possible with the alleles in each block is over a million. Therefore, it is not unreasonable to expect the natural stand surrounding this orchard (with much larger numbers of adult genotypes) to produce many pollen gametes with multilocus genotypes which could not be produced in the orchard blocks.

Once allozyme genotypes of the orchard clones are known, embryos from a sample of orchard seed are analyzed and the genotypes of pollen gametes inferred. The multilocus genotype of each pollen gamete in the seed sample is compared to the array of pollen genotypes which can be produced by the orchard clones. The observed proportion of pollen gametes with genotypes not matching those from the orchard clones (i.e., observed contaminants) is b. The probability of observing contaminants (if they are present) will increase as the number of variable loci used in the analysis increases.

To estimate d, genotypes of a representative sample of adult trees in the background stand must be determined. From these adult genotypes it is possible to infer the frequencies of multilocus pollen gametes produced in the background stand. The frequency of pollen gametes with multilocus markers (d) can then be calculated as one minus the total frequency of indistinguishable pollen gametes. Indistinguishable pollen gametes are those gametes which can be produced in both the orchard and the background stand.

Estimated frequencies of the indistinguishable pollen gametes in the background stand can be inferred from the adult genotypes in one of two ways. In the first, all possible allelic combinations of loci used in the analysis are generated for each sample tree. Since information on linkage relationships (i.e., whether linked genes are in "coupling" or "repulsion") is generally not available, we make the simplifying assumption that all gametic types from each sample tree are produced in equal frequency. The estimated frequency of each indistinguishable pollen gamete in the background stand is then calculated as its unweighted mean frequency among gametes of all sample trees. The second method of estimating the frequencies of indistinguishable pollen gametes utilizes single locus allele frequencies which are computed from the sample of adult genotypes. The frequency of each indistinguishable pollen gamete in the background stand is assumed to be the product of its component allele frequencies. This method is somewhat simpler computationally than the first, since all possible multilocus gametic combinations do not have to be generated for each sample tree. In practice, we have found that both methods lead to very similar estimates of d, and for the examples in the next section the second method was employed.

Both of the above methods for estimating multilocus pollen frequencies assume that: (1) the genes used in the analysis are in linkage equilibrium in the background stand, and (2) pollen productivity is independent of multilocus gametic genotype. In large outcrossing populations of forest trees, the first assumption is probably not violated to any great degree, and if it were, it is not clear what effect, if any, this would have on estimates of m. The second assumption seems to be reasonable, especially when a large number of variable loci are utilized, because each multilocus gamete can be produced by a variety of adult genotypes. Certainly, genetic markers based on alleles at one or a few loci would be much more sensitive to violations of the second assumption.

The number and distribution of sample trees must be considered in choosing a representative sample from the background stand. Sample size is dependent on the size of the background stand and on practical constraints on sampling. Preliminary computer simulations to test the effects of sample size on estimates of d suggest that 50 to 100 trees is a minimum for adequately estimating d. The distribution of these trees should be such that

all background sources of pollen are well represented. Even though pollen can travel considerable distances, it is likely that dominant and codominant trees close to the orchard will contribute the majority of contaminating pollen an orchard receives. Thus, for practical purposes, sampling should be restricted to these trees.

A large sample approximation of the variance of $\hat{m}$ is:

$$\mathrm{var}(\hat{m}) = \frac{b(1-b)}{nd^2} \qquad (6)$$

where

n = number of seed orchard embryos sampled to estimate b.

This must be considered a minimum variance because it does not take into account the fact that d is estimated. Since var(m) is directly proportional to $1/d^2$, it is highly sensitive to the magnitude of d.

ILLUSTRATION OF THE MULTILOCUS TECHNIQUE

Use of the multilocus technique is illustrated with data from the Beaver Creek (BC) seed orchard. The first heavy cone crop in the BC orchard was in 1980 (the year of sampling) when the average age of clones in the orchard was 14 years. The orchard is surrounded by a large, essentially continuous stand of Douglas-fir. No pollen dilution zones, either between adjacent blocks or between the orchard blocks and the surrounding natural stand, are present.

Pollen contaminants in any one BC block could come from other orchard blocks or from the natural stand. Ideally, allele frequencies used in estimating d should be based on a representative sample of genotypes in both background sources. At BC, however, genotypes of all clones in all blocks were already known. For each block the allele frequencies in all remaining blocks were very similar to those among a sample of 183[2]/ trees from the natural stand. Therefore, instead of randomly sampling ramets in other blocks, background stand allele frequencies for each block were calculated as unweighted means of the frequencies in other blocks and the natural stand.

Multilocus estimates (based on 14 allozyme loci) of pollen contamination from all background sources for each of three blocks and for all ten blocks combined are given in Table 2. For example, of 259 seeds sampled in block 7, 47 were found to have pollen gametes with multilocus genotypes which could not have come from any clone in the block (i.e., b = 47/259 = 0.181). Furthermore, d is estimated to be 0.423 indicating that approximately 42.3 percent of the pollen gametes produced by all background sources have multilocus markers. Thus, total contamination in block 7 is estimated to be 0.181/0.423 = 0.43. The average level of pollen contamination over all ten blocks was estimated to be 0.52. Although this estimate seems high, it is not unreasonable, given the low level of pollen production in the orchard. An estimate of pollen contamination in an older Dougas-fir orchard (20

2/ Genotypes of trees were inferred from a sample of seeds or determined directly from bud tissues. Of the 14 loci scored in seeds, only 8 could also be scored in buds. Thus, numbers of trees genotyped per locus ranged from 57 (6 loci) to 183 (8 loci).

Table 2.--Multilocus estimates of pollen contamination from all background sources ($\hat{m}$, other blocks and surrounding natural stand) and from the surrounding natural stand only ($\hat{m}_s$) in three blocks, and all ten blocks combined, at the Beaver Creek Douglas-fir seed orchard.[a/]

Orchard block	Clones	All sources[c/]					Natural stand only[c/]			
		N[b/]	$\hat{b}$	$\hat{d}$	$\hat{m}$	(SE)	$\hat{b}_s$	$\hat{d}_s$	$\hat{m}_s$	(SE)
BC-7	25	259	.181	.423	.43	(.057)	.031	.112	.28	(.096)
BC-5	25	582	.175	.366	.48	(.043)	.048	.112	.43	(.079)
BC-10	27	542	.207	.323	.64	(.054)	.044	.112	.39	(.079)
All blocks combined	251	3023	--	--	.52[d/]	(.061)	.044	.112	.39	(.033)

[a/] Based on 14 allozyme loci.

[b/] N = number of pollen gametes (seeds) sampled.

[c/] $\hat{b}$ and $\hat{b}_s$ are the estimated proportions of orchard ovules fertilized by pollen gametes with multilocus markers from all background sources and from the natural stand, respectively. $\hat{d}$ and $\hat{d}_s$ are the estimated proportions of multilocus markers in the pollen of all background sources and in the pollen of the surrounding stand, respectively.

[d/] Unweighted mean over all blocks.

years) which has been producing heavy pollen and cone crops periodically since 1971, was much lower ($\hat{m}$ = 0.29).

The multilocus technique was also used to determine the amount of pollen contamination in each block from the natural stand only. First, the proportion of pollen gametes in the seed sample of each block which could not have come from any of the orchard blocks was determined (b_s in Table 2). Then, allele frequencies in the natural stand were used to calculate d_s, and m_s was estimated as b_s/d_s (Table 2). Pollen from the surrounding natural stand accounted for a large proportion of the total contamination in BC blocks 5, 7, and 10. Over all the blocks, pollen from the natural stand (m_s = 0.39) made up 75 percent of the total contamination (m = 0.52). The high proportion of contamination from the natural stand again seems reasonable given the low level of pollen production in the orchard. As the orchard ages and pollen production within each block increases, it is likely that both total contamination and the proportion of contamination due to pollen from the natural stand will decrease.

As the number of clones in an orchard increases, the ability to detect pollen contaminants will decrease because fewer multilocus pollen genotypes in background sources will differ from those produced by the orchard clones. Nevertheless, when all 251 clones at BC are considered together, a moderate

proportion (a_s = 0.112) of the pollen in the surrounding stands is expected to have marker genotypes. Thus, even in an orchard with large numbers of clones or with small values of d, it is possible to use the multilocus technique to estimate m. However, larger sample sizes would be required to maintain low standard errors.

CONCLUSIONS AND RECOMMENDATIONS

The multilocus method of estimating pollen contamination in clonal seed orchards has several advantages over single-locus methods described previously (Friedman and Adams 1981). Because a large proportion of contaminants are observed directly, multilocus estimates are more efficient (i.e., have lower standard errors) than single-locus estimates. Furthermore, unlike single-locus methods, the multilocus method does not require differences in allele frequency between the orchard and background stand. Finally, both methods require estimates of allele frequencies in the background stand. Frequencies of a number of multilocus markers are combined to estimate d in the multilocus method. Thus, d is not likely to be very sensitive to factors such as unequal pollen production, gametic selection or genetic subdivisions in the natural stand, all of which might cause single-locus marker frequencies in pollen to differ considerably from those in adult trees.

Although we are still testing the multilocus technique, some preliminary recommendations for its use can be made:

1. Multilocus genotypes of all clones in the orchard and a sample of trees from the background stand are necessary to estimate d. Unless radical changes in the composition of the orchard or background stand are made, d need only be calculated once. Levels of pollen contamination in subsequent years or after application of different pollen management techniques can then be determined by estimating b from a sample of the seed crop and dividing it by the previously estimated d.

2. Based on preliminary computer simulations, a sample of at least 50 to 100 dominant or codominant trees in the background stand is a minimum for adequately estimating d.

3. The availability of variable loci and the degree of precision required dictate the number of loci to use in estimating m. The magnitude of d decreases as the number of loci decreases. In BC block 5, estimates of m were similar using 14 or 9 loci but the standard error of the estimate based on 9 loci was approximately 23 percent larger. Thus, to maintain a certain level of precision with fewer loci, the sample of seeds in the orchard crop must be increased.

4. The number of seeds to sample in the orchard crop depends on the size of d and the degree of precision required. In the absence of previous knowledge of the magnitude of m, we suggest that sequential sampling be employed, with sample size progressively increased until the desired level of precision is reached.

LITERATURE CITED

Adams, W. T. 1982. Application of isozymes in tree breeding. In S. D. Tarksley and T. J. Orton (eds.), Isozymes in plant genetics and breeding. Elsevier Scientific Publ. Co. (in press).

Bridgwater, F. E., and Trew, I. F. 1981. Supplemental mass pollination. In E. C. Franklin (ed.), Pollen management handbook, pp. 52-57. USDA Agric. Handb. 587. Washington, D.C.

Brown, A. H. D., and Moran, G. F. 1981. Isozymes and genetic resources of forest trees. In M. T. Conkle (technical coordinator), Proc. Symp. Isozymes of North Am. Forest Trees and Forest Insects, pp. 1-10. Berkeley, CA. USDA Forest Serv. Gen. Tech. Rep. PSW-48.

Denison, N. P., and Franklin, E. C. 1975. Pollen management. In R. Faulkner (ed.), Seed orchards, pp. 92-100. Forest. Comm. Bull. No. 54. Her Majesty's Stationery Office, London.

Friedman, S. T., and Adams, W. T. 1981. Genetic efficiency in loblolly pine seed orchards. Proc. 16th South. Forest Tree Improv. Conf., Blacksburg, VA. pp. 213-224.

Schmidt, R. L., and Hamblett, K. C. 1962. Directional sources of extraneous pollen at three seed orchard sites on Vancouver Island. For. Chron. 38(2): 203-207.

Silen, R. 1962. Pollen dispersal considerations for Douglas-fir. J. For. 60:790-795.

Silen, R., and Keane, G. 1969. Cooling a Douglas-fir seed orchard to avoid pollen contamination. USDA Forest Serv. Res. Note PNW-101.

Spiess, E. R. 1977. Genes in populations. 780 p. John Wiley and Sons, Inc.

Squillace, A. E. 1967. Effectiveness of 400-foot isolation around a slash pine seed orchard. J. For. 65:823-824.

Squillace, A. E., and Long, E. M. 1981. Proportion of pollen from non-orchard sources. In E. C. Franklin (ed.), Pollen management handbook, pp. 15-19. USDA Agric. Handb. 587. Washington, D.C.

Wang, C. W., Perry, T. 0., and Johnson, A. G. 1960. Pollen dispersion of slash pine (Pinus elliottii Engelm.) with special reference to seed orchard management. Silvae Genet. 9:78-86.