Theor Appl Genet (1987) 73:799–808

Mating system and multilocus associations in a natural population of *Pseudotsuga menziesii* (Mirb.) Franco

F. C. Yeh [1] **and K. Morgan** [2, *]

[1] Department of Forest Science, University of Alberta, Edmonton, Alberta T6G 2H1, Canada
[2] Department of Genetics, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Received September 20, 1986; Accepted December 14, 1986
Communicated by P. M. A. Tigerstedt

Summary. Arrays of open-pollinated seeds were assayed for allozyme polymorphisms at ten loci (*Aat2, Est1, G6pd, Idh, Mdh2, Mdh3, Pgm, Sod, 6Pgd1, 6Pgd2*) to obtain estimates of the outcrossing rate and assess multilocus association in a natural population of coastal Douglas-fir, *Pseudotsuga menziesii* (Mirb.) Franco. The allele frequencies in the samples of adult trees and pollen-gamete pool were similar. Maximum-likelihood estimators of the outcrossing rate for individual loci and two multilocus models were derived using counting methods. The single-locus maximum likelihood estimates (MLEs) of the outcrossing rate were significantly heterogeneous; they varied over a more than two-fold range from 0.404 to 0.935, with an average MLE of 0.741. Both multilocus MLEs of the outcrossing rate were 0.887. The sample of trees was in random mating equilibrium when assessed on a pairwise-locus basis using Burrows' composite measure of gametic disequilibrium, with one exception (*Mdh2 Sod*) that was attributable to a rare "gametic" class. In the sample of pollen gametes, 5 of the 45 pairwise-locus associations were nominally significant at the 0.05 level: *Idh Est1, Mdh2 Sod, Aat2 Est1, Aat2 Mdh3,* and *Est1 Mdh3.* These apparent associations were attributable in most cases to the relative excess of uncommon or rare paternal gametes of discernibly outcrossed embryos. An additional two-locus association was identified for *Mdh2 Pgm* which was marginally significant for the major partition of the contingency table that excluded paternal gametes with the rare allele $Mdh2^2$.

Key words: Douglas-fir – Mixed mating model – Outcrossing rate – Gametic disequilibrium – Counting method

Introduction

Mating system studies in forest trees using allozyme polymorphisms have provided knowledge pertaining to the genetic structure of continuous populations, and to the relative contributions to population differentiation of selection, family structure and gene flow (for example, as reviewed by Brown et al. 1985). A common finding among studies of the mating system in forest trees is the heterogeneity of single-locus estimates of the outcrossing rate for the same set of embryos. Violations of the assumptions on which single-locus models are based have been suggested as the causes of the apparent heterogeneity (Shaw et al. 1981; Ellstrand and Foster 1983; Brown et al. 1985; Schoen and Clegg 1986).

An estimate of the outcrossing rate based on the information obtained by genotyping each embryo for several polymorphic isozyme loci, a multilocus estimator, will be more robust than individual-locus estimates (Shaw et al. 1981). In the first step of a two-step estimation procedure of mating system parameters, the embryo and maternal genotypes are compared in order to categorize each embryo as either discernibly outcrossed or ambiguous. Each additional polymorphic locus that is scored increases the likelihood that an outcrossed embryo will be discerned. One assumption in formulating the multilocus estimator is that no associations exist between alleles at different loci; that is, it is assumed that the pollen pool is in gametic phase equilibrium. (We use the term "gametic disequilibrium" instead of "linkage disequilibrium" in order to refer to associations between independent as well as linked loci.) Since evidence of gametic disequilibrium has most often been obtained in selfing plants, either in natural populations or in field populations construct-

* Present address: Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec H3A 1A2, Canada

ed from composite crosses (Brown and Allard 1970; Allard et al. 1972; Clegg et al. 1972; Kahler et al. 1975), the prediction for outcrossing plants such as forest trees has been that any initial gametic disequilibrium will rapidly decay in the absence of strong epistatic selection. This is a prediction, however, of long-term behavior. In the reestablishment of a forest tree population after a fire, for example, there could initially be gametic disequilibrium which will slowly decay for closely-linked genes and if there is nonrandom mating. This consideration is appropriate for coastal Douglas-fir, *Pseudotsuga menziesii* [Mirb.] Franco, in British Columbia because its distribution has been associated with the recent history of fire occurrence (Schmidt 1960).

In this paper, results are reported of analyses of ten polymorphic enzyme loci assayed by electrophoresis in a population of coastal Douglas-fir naturally regenerated after fire during the year 1927. The reproductive system in gymnosperms offers an unique opportunity to study segregation, linkage, gametic disequilibrium and outcrossing from the information provided by the various components of seed. The female gametophyte (megagametophyte) is haploid and all cells within the seed coat including the egg-cell (ovule) have been derived mitotically from a single megaspore. The maternal and pollen contributions to an embryo can be unambiguously determined, in principle, by genotyping the megagametophyte and the corresponding embryo from a single seed. At the same time, the most likely genotype of the tree can be assigned by inspection of the segregation patterns among its megagametophytes from a small number of seeds (Morris and Spieth 1978). This investigation includes single-locus and multilocus estimates of the outcrossing rate using the mixed mating model, tests for Hardy-Weinberg and gametic disequilibrium in the adult trees, and analyses of pairwise-locus associations in the pollen gametes.

Materials and methods

Seed collection

Cones were collected in 1981 by rifle-fire from a pure coastal Douglas-fir population located on Mt. Prevost near Duncan, British Columbia, at 457 m elevation, latitude 48° 52′N and longitude 123° 45′W. The stand is situated on a gentle slope approximately 10 ha in area and composed of second growth trees under 60 years old, the result of natural regrowth after being burned in 1927. The site is reasonably uniform with respect to climate, landform, soil, and vegetation. A plot center was arbitrarily established and cone-bearing trees at a minimum distance of 45 m apart were sampled around the center, extending the plot to concentric circles until 60 trees were included. The individuality of the cone lots and, subsequently, seed lots were retained for the trees.

Electrophoretic procedures

The megagametophyte and embryo tissues were prepared for electrophoresis following the methods of Yeh and O'Malley (1980). Eight seeds per tree were picked at random from the seed lot and analyzed for variation by horizontal starch gel electrophoresis of the eight enzymes that could be reliably detected in the embryo tissue: aspartate aminotransferase (AAT, EC 2.6.1.1), esterase (EST, EC 3.1.1.1), glucose-6-phosphate dehydrogenase (G6PD, EC 1.1.1.49), isocitrate dehydrogenase (IDH, EC 1.1.1.42), malate dehydrogenase (MDH, EC 1.1.1.37), phosphoglucomutase (PGM, EC 2.7.5.1), sorbitol dehydrogenase (SOD, EC 1.15.1.1), and 6-phosphogluconic dehydrogenase (6PGD, EC 1.1.1.44). Buffer systems and staining techniques were described by Yeh and O'Malley (1980). Gels were prepared with electrostarch, lot 307 (Electrostarch Co., Madison WI).

When interpreting the electrophoretic banding patterns we followed the principles which have been outlined with special reference to lodgepole pine, *Pinus contorta* spp. *latifolia*, population data (Yeh and Layton 1979). The most common allozyme at each locus was designated *100*, with additional allozymes being given numerical values according to their migration relative to the *100* allozyme. When there was more than one locus for an enzyme, the isozymes were numbered in decreasing order of anodal mobility. The inheritance of the allozymes and linkage of the enzyme loci in coastal Douglas-fir have been reported by El-Kassaby et al. (1982a) and El-Kassaby et al. (1982b), respectively. The Mt. Prevost population is genetically different than the coastal Douglas-fir population studied by El-Kassaby et al. (1982a, b) with the result that the inventories of isozymes analyzed in the two samples do not entirely overlap.

Outcrossing rate

Maximum-likelihood estimates (MLEs) of the outcrossing rate of the mixed mating model were obtained by counting methods (Smith 1957) that use the EM algorithm (Dempster et al. 1977). A single-locus estimator of the outcrossing rate, an estimator of the average outcrossing rate of a set of independent loci, and two multilocus estimators were used. It is assumed that loci segregate independently in the formation of gametes for self-fertilization. An embryo that carries a non-maternal allele at any locus is categorized as discernibly outcrossed (DOC), otherwise as ambiguous (Shaw et al. 1981). An iterative procedure is used to estimate the posterior probability of being outcrossed for each embryo in the ambiguous class. This probability is a function of the outcrossing rate, the frequency of the paternal genotype in the outcross-pollen pool, and the expected segregation ratio of the maternal tree. One multilocus estimator restricts the set of possible multilocus genotypes of the pollen gametes to those observed in the sample of embryos. We call this "restricted-gametic" estimation. In the second multilocus estimator, the expected frequency of the multilocus genotype of the paternal gamete of an outcrossed embryo is the product of the appropriate allele frequency estimates of the outcross-pollen pool (e.g., Shaw et al. 1981). We call this "product-frequency" estimation. The restricted-gametic estimator is a straightforward extension of the single-locus estimator previously published (Cheliak et al. 1983).

For a test of the hypothesis that a single-locus estimate of the outcrossing rate is equal to 1 and a test for heterogeneity of outcrossing rates for independent loci, we used the Neyman-Pearson likelihood-ratio criterion Λ (Rao 1973). For the test of heterogeneity of outcrossing rates, we calculated the likelihood-ratio test statistic

$$-2\log_e \Lambda_c = 2\left[\sum l(\hat{t}_i) - l(\hat{\bar{t}})\right]$$

where $\hat{t}_i$ is the single-locus MLE of the outcrossing rate for the ith locus; $\sum l(\hat{t}_i)$ is the sum of the corresponding log-likelihoods of the independent loci; $\hat{\bar{t}}$ is the MLE of the average outcrossing rate, and $l(\hat{\bar{t}})$ is the corresponding log-likelihood. The likelihood-ratio test statistic is asymptotically distributed as χ^2 and for the heterogeneity test has degrees of freedom equal to the number of loci minus one. For the test of the hypothesis that the estimated outcrossing rate is $t = 1$, we calculated the likelihood-ratio test statistic

$$-2\log_e \Lambda_0 = 2[l(\hat{t}_i) - l(1)].$$

Assignment of diploid genotype, and multilocus associations

The genotype at each locus for a tree was assigned by inspection of the segregation pattern of the eight haploid megagametophytes. If the segregation ratio for a heterozygous locus is binomially distributed with an expected value of 1/2, the likelihood of not detecting a heterozygote by chance is $(1/2)^7$ or less than 0.01. Allele frequencies in the diploid, parental sample of 60 trees were estimated from the assigned genotypes by gene counting. A goodness-of-fit test of the composite hypothesis of Hardy-Weinberg and gametic equilibrium in the parental sample was carried out for each pair of loci; the test is based on Burrows' composite measure of gametic disequilibrium (Weir 1979). Alternatively, for a sample of diploid genotypes drawn from a population in random mating equilibrium, the multilocus gamete frequencies were estimated by a counting method, and the association between alleles at different loci was assessed by log-likelihood ratio test statistics (Hill 1975).

Associations of allele frequencies at different loci in the sample of pollen gametes were assessed by contingency table analysis of pairs of loci. We used pairwise tests of association because of the sparseness of the tables of observed numbers of multilocus genotypes. Both the likelihood-ratio test statistic G^2 and the Pearson chi-square statistic X^2 were inspected for evidence of pairwise association. If the value of either statistic was greater than the value of χ^2 at $P = 0.05$, we considered the association nominally significant. In addition, the standardized cell residuals, $z^{(1)}$ and $z^{(3)}$ (Bishop et al. 1975), were computed for all two-locus tables. These have the following forms

Standardized cell deviate, component of X^2:

$$z^{(1)} = (x_i - m_i)/\sqrt{m_i}$$

Freeman-Tukey deviate:

$$z^{(3)} = \sqrt{x_i} + \sqrt{x_i + 1} - \sqrt{4m_i + 1}$$

where x_i and m_i are, respectively, the observed and expected numbers for the ith cell. An excess (or deficiency) of a gamete was judged to be significantly large if $|z^{(1)}| > 1.96$ *and* $|z^{(3)}| > 1.96\sqrt{\text{d.f./no. cells}}$ in a two-locus test (for the second criterion, see comments of Sokal and Rohlf 1981, p 762). The contribution of a particular genotype to the lack of fit can be assessed formally by comparing the reduction in G^2 after fitting the structurally incomplete contingency table with a fixed-zero cell (Fienberg 1980). Finally, sample randomization tests were carried out by Monte Carlo simulation to augment some of the large-sample tests of significance (e.g., Sokal and Rohlf 1981, pp 787–795).

Results

A total of 13 loci were assayed. Two of these, *Aat1* and *Mdh4*, were monomorphic. For *Mdh1*, 55 of 60 trees were homozygous for the *100* allele; only nine pollen gametes out of 475 were other than *Mdh1-100*. The remaining ten loci, *Aat2*, *Est1*, *G6pd*, *Idh*, *Mdh2*, *Mdh3*, *Pgm*, *Sod*, *6Pgd1*, and *6Pgd2*, were polymorphic in the sample and analyzed in detail. Occasionally the genotype of an embryo at a particular locus should not be unambiguously determined from the allozymes; the deduced genotype of the paternal gamete was scored as if "missing". We report the results of analyses of the pollen gametes using all available genetic information and often for a subset of 375 pollen gametes from embryos which could be genotyped for all of the ten polymorphic loci.

There were 56 distinct multilocus genotypes among the 60 trees. Among the 375 completely-scored pollen gametes, there are 149 distinct multilocus genotypes; the two most common genotypes, [1 1 1 1 1 1 1 1 1 1] and [2 1 1 1 1 1 1 1 1 1], were observed 34 and 37 times, respectively. (The left-to-right order of loci in a multilocus genotype follows the order of the loci in Table 1, and the allele designations are the ordinal numbers of the allozymes.) Estimates of allele frequencies in the tree sample and pollen pool are given in Table 1. For each locus the alleles are homogeneously distributed between the two samples, as determined by contingency table analysis ($P > 0.05$).

Single-locus and multilocus estimates of the outcrossing rate

The occurrence of a non-maternal allele identified an embryo as DOC. There were 261 DOC among the 375 completely-scored embryos: 132 embryos were DOC for only one locus, 103 for two loci, 19 for three loci, and 7 embryos were DOC for four loci. The proportion of DOC embryos, average 0.696, had a standard deviation of 0.236 among trees; the number of DOC embryos per tree ranged from 1 to 8. Of the 149 distinct pollen genotypes, 143 were present among the DOC embryos. The proportion of DOC embryos ranged from 0.026 for *6Pgd2* to 0.301 for *G6pd* (see Table 2). The single-locus estimates of the outcrossing rate varied over a 2.5-fold range from 0.404 for *Est1* to 0.935 for *Mdh2* (Table 2); the average single-locus estimate of the outcrossing rate was 0.741. The estimates of the outcrossing rate for all but two loci, *Mdh2* and *Sod*, were significantly less than $t = 1$ (Table 2). Re-estimation of the outcrossing rate for the 375 embryos with complete allozyme data resulted in small changes; the average of the estimates increased to 0.758, a proportional increase of about 2%. Both multilocus MLEs of the outcrossing rate were 0.887 and significantly less than $t = 1$ (likelihood-ratio test statistics: 39.41 and 56.86 for the restricted-gametic and product-frequency estimators, respectively). This is a

Table 1. Allele frequencies for ten polymorphic enzyme loci

Locus	Allozyme	Maternal alleles (N = 60 trees)		Pollen gametes (N = 480 embryos)	
		No.	Frequency	No.	Frequency
1. *G6pd*	*100*	51	0.425	208	0.447
	90	63	0.525	232	0.499
	80	6	0.050	25	0.054
2. *Idh*	*100*	92	0.767	374	0.803
	122	16	0.133	56	0.120
	90	2	0.017	17	0.036
	65	10	0.083	19	0.041
3. *Mdh2*	*100*	103	0.858	416	0.881
	110	3	0.025	2	0.004
	90	14	0.117	54	0.114
4. *Pgm*	*100*	103	0.858	394	0.828
	105	6	0.050	19	0.040
	94	11	0.092	63	0.132
5. *Sod*	*100*	114	0.950	451	0.949
	68	6	0.050	24	0.051
6. *6Pgd1*	*100*	108	0.900	434	0.935
	107	3	0.025	14	0.030
	85	9	0.075	16	0.034
7. *6Pgd2*	*100*	110	0.917	441	0.955
	122	10	0.083	21	0.045
8. *Aat2*	*100*	105	0.875	387	0.852
	112	8	0.067	39	0.086
	82	7	0.058	28	0.062
9. *Est1*	*100*	71	0.592	245	0.554
	110	29	0.242	88	0.199
	95	20	0.167	109	0.247
10. *Mdh3*[a]	*100*	109	0.908	427	0.910
	130	0	0.0	1	0.002
	84	10	0.083	39	0.083
	100r	1	0.008	2	0.004

[a] *100r* allozyme of the *Mdh3* locus has reduced staining activity

17% relative increase over the average outcrossing rate for the same set of 375 embryos.

The heterogeneity of single-locus estimates of the outcrossing rate is evident from the results in Table 2. Two groups of loci differed with respect to their average outcrossing rates (Table 3): loci {8, 9}, *Aat2* and *Est1*, had a low average rate of 0.425, and the remaining eight loci {1, 2, 3, 4, 5, 6, 7, 10} had a high average rate of 0.849.

Pairwise-locus associations

There is little evidence for genotypic disequilibrium within loci or between pairs of loci in the tree sample. All but one of the 45 pairwise-locus comparisons had adequate fit to the composite hypothesis of gametic equilibrium based on Burrows' measure. The one exception was the locus-pair *Mdh2 Sod* ($X^2 = 11.62$, $G^2 = 7.64$ with 2 d.f.). There was an apparent relative excess of the gametic class $Mdh2^2 Sod^2$ (2 observed, 0.3 expected) which was the contribution of two double-heterozygotes in the SW quadrant of the sampling plot. A rare $Mdh2^2 Sod^2$ pollen gamete was also ascertained (see below).

For the pollen gametes, 5 of the 45 pairwise-locus comparisons were nominally significant at the 0.05 level as judged by either of the goodness-of-fit statistics, X^2 or G^2. The gametes judged to be in relative excess (or in relative deficiency in the case of *Aat2 Est1*) on the basis of large values of the deviates, $z^{(1)}$ and $z^{(3)}$, are given in Table 4. An additional pair of loci, *Mdh2 Pgm*, is included by this criterion and had a

Table 2. Single-locus and multilocus estimates of the outcrossing rate and allele frequencies of the outcross-pollen pool

Locus	No. scorable embryos[a]			Allele frequencies of outcross-pollen pool				Outcrossing rate[b], $\hat{t}$
	Outcross status							
	DOC	Ambiguous	Total					
1. *G6pd*	140	325	465	0.451	0.496	0.054		0.905
2. *Idh*	76	390	466	0.809	0.123	0.031	0.037	0.874
3. *Mdh2*	38	434	472	0.883	0.005	0.113		0.935
4. *Pgm*	58	418	476	0.815	0.047	0.138		0.751
5. *Sod*	20	455	475	0.949	0.051			0.927
6. *6Pgd1*	21	443	464	0.942	0.035	0.022		0.850
7. *6Pgd2*	12	450	462	0.962	0.038			0.819
8. *Aat2*	36	418	454	0.825	0.079	0.096		0.511
9. *Est1*	76	366	442	0.501	0.129	0.370		0.404
10. *Mdh3*	24	445	469	0.912	0.003	0.085	0.0	0.685
Average single-locus outcrossing rate								0.741[c] 0.758[d]
Multilocus outcrossing rate:	261	114	375					0.887

[a] The total number of seeds analyzed was 480 DOC is discernibly outcrossed
[b] Convergence criterion for estimate of number of outcross pollen gametes was 10^{-5}
[c] Results obtained with embryos scored for each locus independently
[d] Results for the subset of 375 embryos with complete allozyme data

Table 3. Heterogeneity of single-locus estimates of outcrossing rates for independent loci

Locus	No. embryos	Outcrossing rate, $\hat{t}$	Log-likelihood minus constant		Likelihood-ratio test of hypothesis: $t = 1$	
			$t = \hat{t}$	$t = 1$	G^2 (1 d.f.)	P
1. *G6pd*	465	0.905	− 399.70	− 401.72	4.05	< 0.05
2. *Idh*	466	0.874	− 311.96	− 317.99	12.06	< 0.001
3. *Mdh2*	472	0.935	− 180.06	− 180.54	0.95	> 0.30
4. *Pgm*	476	0.751	− 254.57	− 263.09	17.04	< 0.0001
5. *Sod*	475	0.927	− 94.51	− 95.03	1.03	> 0.30
6. *6Pgd1*	464	0.850	− 126.82	− 131.90	10.15	< 0.01
7. *6Pgd2*	462	0.819	− 80.92	− 85.43	9.02	< 0.01
8. *Aat2*	454	0.511	− 208.94	− 235.53	53.17	$< 10^{-6}$
9. *Est1*	442	0.404	− 364.48	− 439.19	149.42	$< 10^{-6}$
10. *Mdh3*	469	0.685	− 138.85	− 154.12	30.55	$< 10^{-6}$
Sum	−	−	− 2160.81	−	−	

Components	Average outcrossing rates				
	Rate	Log-likelihood minus constant	Likelihood-ratio test of heterogeneity		
	$\hat{t}$		G^2	d.f.	P
Total, Loci {1, ..., 10}	0.741	− 2,204.42	87.22	9	$< 10^{-6}$
Loci {1, 2, 3, 4, 5, 6, 7, 10}	0.849	− 1,592.43	10.08	7	> 0.15
Loci {8, 9}	0.425	− 573.98	1.13	1	> 0.25
Between sets of loci	−	− 2,166.41	76.01	1	$< 10^{-6}$

Table 4. Nominally significant two-locus associations in pollen gametes [a]

	Locus pair	N	v	ab	Gamete	Obs.	Exp.	Standardized cell residuals [b]	
								$z^{(3)}$	$z^{(1)}$
Case 1.	*Idh Est1*	430	6	12	[4; 2]	8	3.8	1.80	2.15
2.	*Mhd2 Pgm*	468	4	9	[3; 3]	13	7.3	1.86	2.13
3.	*Mdh2 Sod*	468	2	6	[2; 2]	1	0.1	1.23	2.80
4.	*Aat2 Est1*	419	4	9	[2; 1]	9	19.9	−2.81	−2.44
					[2; 3]	17	8.5	2.44	2.91
5.	*Aat2 Mdh3*	444	6	12	[2; 4]	2	0.2	1.85	4.42
6.	*Est1 Mdh3*	432	6	12	[2; 4]	2	0.4	1.54	2.56

[a] N is sample size, v degrees of freedom, and ab the number of cells in the two-way table
[b] Criteria for large deviates: the standardized cell residuals were $|z^{(1)}| > 1.96$ *and* $|z^{(3)}| > 1.96\sqrt{v/ab}$ in a two-locus test

marginally significant X^2 for the major partition of the contingency table that excludes the gametes with the rare allele $Mdh2^2$ (two $Mdh2^2 Pgm^1$ were observed).

It is informative to investigate the contribution of particular genotypes to the two-locus associations. For cases 3, 5 and 6 in Table 4, in presence of one or two pollen gametes that were expected to be rare was the reason for the lack of fit. Cases 5 and 6 involve the same pair of gametes, [1 1 1 1 1 1 1 2 2 4] and [2 1 1 1 1 1 1 2 2 4], of DOC embryos from one tree in the NW quadrant of the sample plot. For case 1, genotype $Idh^4 Est1^2$ was in excess. It was not, however, a rare genotype and did not account for the majority of the lack of fit. Case 4, *Aat2 Est1*, was perhaps the most statistically reliable test in that the table was not sparse. Primarily $Aat2^2 Est1^1$, which was in relative deficiency, contributed significantly to the lack of fit. With this cell fixed in the table, the reduction in the likelihood-ratio test statistic was 15.19. Fixing both this gamete and $Aat2^2 Est1^3$, which was in relative excess, resulted in a reduction of G^2 of 16.21. The MLEs of the gamete frequencies in the pollen pool were ($N = 419$)

0.516 0.150 0.191, 0.021 0.021 0.041, 0.031 0.017 0.012.

There was no significant association between these two loci in the sample of adult trees. The MLEs of the *Aat2 Est1* gametes in the sample of adult trees were

0.520 0.212 0.143, 0.016 0.028 0.023, 0.056 0.002 0.0.

Discussion

The single-locus and multilocus estimates of the outcrossing rate in the present study are similar to those obtained by Shaw and Allard (1982) for coastal Douglas-fir. They analyzed eleven isozyme loci in megagametophytes and embryos of open-pollinated seeds from eight natural populations from Oregon and Washington, and a seed orchard in Oregon. Although we do not know if the *Est1* locus analyzed in the present study is homologous to the one analyzed by Shaw and Allard, and the estimation procedures are somewhat different, in both studies the single-locus estimates of the outcrossing rate for *Est1* are among the lowest, as low as 0.40 in the Mt. Prevost population and 0.33 in Shaw and Allard (1982). The cause of high selfing at the *Est1* locus in coastal Douglas-fir is unknown at this time. Roberds and Conkle (1984), however, hypothesized that selection might have been the contributing factor in loblolly pine, *Pinus taeda* L., by acting either directly on the *Est* locus or at correlated loci. Excluding *Est*, the minimum outcrossing rate was 0.67 in the Oregon and Washington stands of coastal Douglas-fir, and the means of the single-locus estimates ranged from 0.88 to 0.93 with a grand mean of 0.91 (Shaw and Allard 1982). Single-locus estimates are clearly quite variable in both studies. Heterogeneity of single-locus estimates of the outcrossing rate is common among forest trees and has been reported for radiata pine, *Pinus radiata* Don. (Moran et al. 1980), ponderosa pine, *Pinus ponderosa* Laws. (Mitton et al. 1981), lodgepole pine, *Pinus contorta* var. 'latifolia' (Epperson and Allard 1984), and several *Eucalyptus* species (Brown et al. 1975; Phillips and Brown 1977; Moran and Brown 1980; Yeh et al. 1983).

The estimates of mating system parameters for each of the ten loci were made from a common set of embryos. Therefore, the actual proportions of outcrossed and selfed embryos are expected to be the same for every locus. The heterogeneity of single-locus estimates must be due to variation in information among loci to detect outcrossing events and estimate the rate, and/or violations in the assumptions of the population genetic model of the estimation procedure. Such violations include spatial heterogeneity of the pollen gamete pool, segregation distortion, assortative mating, and differential survival of inbred progeny due to recessive deleterious genes which are closely-linked to the enzyme loci, among other factors.

The average of single-locus estimates of the outcrossing rate (0.741) is expected to be lower than the

Table 5. Sample randomization tests of genotypic heterogeneity of progeny sets among homozygous maternal trees

Locus and tree genotype		No. embryos (8 per progeny set)				Sample randomization tests [a]		
						Approximate significance levels [b]		No. random samples per test
		No. trees	Not scored	Hetero-zygous	Homo-zygous	All embryos	Excluding unscored	
1. *G6pd*	*100/100*	13	4	52	48	0.298	0.249	1,000
	90/90	17	5	57	74	0.153	0.186	1,000
2. *Idh*	*100/100*	34	6	52	214	0.113	0.089	1,000
3. *Mdh2*	*100/100*	43	5	35	304	0.068	0.057	1,000
4. *Pgm*	*100/100*	44	3	46	303	0.008	0.016	1,000
5. *Sod*	*100/100*	54	5	20	407	0.004	0.0079	10,000
6. *6Pgd1*	*100/100*	48	13	21	350	0.031	0.151	1,000
7. *6Pgd2*	*100/100*	50	12	12	376	< 0.0001	< 0.0001	10,000
8. *Aat2*	*100/100*	45	13	32	315	< 0.0001	< 0.0001	10,000
9. *Est1*	*100/100*	22	14	35	127	0.0009	0.0092	10,000
	110/110	4	0	15	17	0.186 [c]	–	2,000
	95/95	2	0	4	12	0.214 [d]	–	2,000
10. *Mdh3*	*100/100*	49	9	24	359	0.379	0.254	1,000

[a] Sample randomization tests of the homogeneity of proportions of progeny types among trees were carried out by Monte Carlo simulation

[b] The results tabulated are the proportion of random samples with log-likelihood less than or equal to the log-likelihood of the observed sample assuming completely independent loci

Large-sample likelihood-ratio tests of heterogeneity:

[c] $G^2 = 4.57$ ($P = 0.21$), d.f. = 3

[d] $G^2 = 1.38$ ($P = 0.24$), d.f. = 1

Table 6. Segregation analysis of megagametophyte sets of size 8 for a specified locus [a]

Locus	1 : 7 or 7 : 1	2 : 6 or 6 : 2	3 : 5 or 5 : 3	4 : 4	No. progeny sets [b]
1. *G6pd*	2	6	10	6	24
2. *Idh*	0	6	9	3	18
3. *Mdh2*	0	2	5	7	14
4. *Pgm*	0	3	7	4	14
5. *Sod*	1	1	1	3	6
6. *6Pgd1*	0	3	4	2	9
7. *6Pgd2*	2	1	2	1	6
8. *Aat2*	1	1	5	2	9
9. *Est1*	0	4	8	4	16
10. *Mdh3*	0	4	3	2	9
Totals [c]	6	31	54	34	125

[a] The heterozygous genotype of a maternal tree was assigned by inspection of the segregation of the 8 megagametophytes (see text)

Chi-squared test of heterogeneity of segregation classes among loci using Smith's (1986) test when expected values are small: $X^2 = 5.34$, $m = 5.17$, $v = 0.46$, and the standardized normal variate $z_2 = 0.25$

[b] Data pooled over trees for each locus

[c] Data pooled over loci and over trees

Test of goodness-of-fit to the expected ratio of 8:28:56:35, $X^2 = 0.904$ with 3 d.f.

multilocus estimate (0.887) when cross-pollination occurs among family members (Ritland and Jain 1981; Shaw et al. 1981). It is likely that the genotypic identity among uncommon and, especially, rare pollen gametes is identity by descent; thus genetic differentiation of the population, that is, population structure, and inbreeding in addition to self-fertilization contribute to the apparent selfing rates. We assessed the combined effects of variability in individual selfing rates and heterogeneity of the pollen pool by analyzing the distribution of heterozygous progeny among homozygous adult trees (Brown et al. 1975). There is evidence for significant effects of population structure in the results of the sample randomization tests (Table 5).

There is in addition a potential bias in the estimation of apparent outcrossing rates that will tend to increase the multilocus estimate more so than the average single-locus estimate in the absence of gametic disequilibrium. This results when there is an appreciable likelihood of *not* detecting a true heterozygote because of segregation distortion, a relatively common phenomenon in forest tree species (Lundkvist 1974; Rudin 1977; O'Malley et al. 1979; Adams and Joly 1980; Eckert et al. 1981; El-Kassaby et al. 1982a; Cheliak et al. 1984; Strauss and Conkle 1986). We analyzed the

segregation of allozymes among the megagametophytes of heterozygous trees (Table 6). The analysis was restricted to those progeny sets which were complete for a given locus, that is, all 8 megagametophytes were scored for allozymes at the particular locus. Clearly, because of the method of genotype assignment (see "Materials and methods") there are no heterozygous trees identified with 8:0 (or 0:8) segregation. The data were pooled over trees for the test of heterogeneity of segregation classes among loci, and there was no significant heterogeneity. Similarly, there was very good fit to the expected ratio of segregation classes for the data pooled over loci and trees (Table 6). Therefore, the likelihood of not detecting a heterozygote is small, i.e., as expected for unbiased segregation in the megagametophytes. It is not increased by segregation distortion in these data.

The single-locus estimates of the outcrossing rate are not independent (Table 4). In the presence of gametic disequilibrium, the multilocus estimate of the outcrossing rate can depart downward in direction from the maximum-likelihood estimate of the average, which is computed on the assumption of independent loci. One way to minimize the effect of gametic disequilibrium in the outcrossing analysis would be to use the data of only one locus from any correlated pair (Brown et al. 1985).

The results on gametic disequilibrium presented in this report are similar to the results of a study on yellow polar, *Liriodendron tulipifera* L. (Roberds and Brotchol 1985). Gametic disequilibria were found mainly in the seedling populations, whereas in the adult trees only weak evidence for disequilibria was detected. The two-locus associations in the present study (Table 4) are not likely to be due to linkage disequilibrium between closely-linked allozyme loci. The ten polymorphic loci are apparently unliked. For a genetically different population of coastal Douglas-fir, El-Kassaby et al. (1982b) provided some information for 8 of the 10 loci of the present study. Unfortunately, *Mdh2* was monomorphic in the sample of trees that was previously studied, and the only close linkage detected was between *Aat2* and *Pgi2* at 2% recombination. The latter locus could not be studied in the sample of embryos from the Mt. Prevost population. We analyzed the joint segregation of *Mdh2* and *Pgm* in megagametophytes from 6 doubly heterozygous trees (data not shown). There was no significant departure from a 1:1:1:1 ratio for either the group of 4 progeny sets of size 8 or the group of 2 progeny sets of size 7 (results not shown). While the data are too few to detect loose linkage, we can conclude that these two loci are not closely linked.

Even without close linkage, there may be an association of alleles at different loci in the gametophytes because of biased representation of parental genotypes in the sampled pollen pool. The male and female gametophyte pools that contributed to the seeds that were sampled may have been genetically different. An example of this is the gametic disequilibrium created initially in a hybrid population as a result of the admixture of genetically different populations (Cavalli-Sforza and Bodmer 1971; Nei and Li 1973). Microdifferentiation in forest trees (Mitton et al. 1977; Linhart et al. 1981), probably due to selection and family structuring, would retard the decay of the initial disequilibrium. In addition, sampling embryos that are more closely related than maternal half-sibs could contribute to the apparent associations of gene frequencies at several loci. This results when the effective sample size is smaller than that given in Table 4 because successive pollination events are not independent (a tendency for fewer pollen parents than expected under open-pollination).

The large difference in outcrossing between the two groups of loci (Table 3) suggest other processes such as zygotic selection against selfed genotypes might have contributed to the disequilibria observed. Natural populations of coastal Douglas-fir have a large embryonic-lethal genetic load, the mean number of lethal equivalents is nine to ten per zygote (Sorensen 1969). The majority of this load is expressed in early embryonic development, particularly upon self-fertilization. Thus, there could have been many more zygotes initially produced from the fertilization events than survived as components of viable seeds. An explanation for the apparent gametic disequilibria, especially that of *Aat2 Est1*, would be either linkage disequilibrium of the isozyme markers with a recessive embryonic lethal or a deleterious effect of the isozymes themselves. Since all possible homozygous progeny at loci in gametic disequilibria (Table 4) were not observed in embryos and adult trees, the linkage disequilibrium could be complete or the deleterious effect would be that of a recessive lethal.

A marked reduction in viable seeds because of limitation of fertilization or loss of zygotes could give rise to the heterogeneity of single-locus estimates of the outcrossing rate in two ways. First, an allozyme locus that is in linkage disequilibrium with self-incompatibility alleles will have a high, estimated outcrossing rate. Second, there might be cross-incompatibility genes. Where an allozyme locus is in linkage disequilibrium with self-compatibility alleles then there would result a lower, estimated outcrossing rate.

We hypothesize that differential selection is primarily responsible for the variability of outcrossing rates among loci, especially between the two groups of loci (Table 3). The group of eight loci imply complete outcrossing, while the remaining pair of loci indicate a selfing rate of approximately 0.50. In the study of other natural populations of coastal Douglas-fir by Shaw and Allard (1982) and seed orchards by Ritland and El-Kassaby (1985), similar results were reported. Thus, there may be variation among regions of the genome of this conifer for closely-linked recessive deleterious genes. On this hypothesis and the evidence of significant viability depression due to selfing in coastal Douglas-fir (Sorensen 1969, 1982), the higher estimate of selfing in this study may be closer to the "true" rate in this effectively, outcrossing species. Controlled crosses will have to be used to evaluate the hypothesis that variability in selection among loci in linkage dis-

equilibrium with the isozyme markers is a major source of variability in effective outcrossing.

Acknowledgements. We thank A. H. D. Brown, W. M. Cheliak, M. T. Conkle, R. P. Guries, G. Namkoong, K. Ritland, and D. V. Shaw for helpful comments on an earlier draft of this paper; V. Farewell, J. Felsenstein, T. M. Holmes, C. Strobeck and E. A. Thompson for helpful discussions during this research; and D. C. Trust and C. Layton for laboratory assistance. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, A2282 (F.C.Y.) and A2350 (K.M.).

References

Adams WT, Joly RJ (1980) Genetics of allozyme variants in loblolly pine. J Hered 71:33–40

Allard RW, Babbel GR, Clegg MT, Kahler AL (1972) Evidence for coadaptation in *Avena barbata*. Proc Natl Acad Sci USA 69:3043–3048

Bishop YMM, Fienberg SE, Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge

Brown AHD, Allard RW (1970) Estimation of the mating system in open-pollinated maize population using isozyme polymorphisms. Genetics 66:133–145

Brown AHD, Matheson AC, Eldridge KG (1975) Estimation of the mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. Aust J Bot 23:931–949

Brown AHD, Barret SCH, Moran GF (1985) Mating system estimation in forest trees: models, methods and meaning. In: Gregorius H-R (ed) Population genetics in forestry. Springer, Berlin

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. Freeman, San Francisco, Calif

Cheliak WM, Morgan K, Strobeck C, Yeh FCH, Dancik BP (1983) Estimation of mating system parameters in plant populations using the *EM* algorithm. Theor Appl Genet 65:157–161

Cheliak WM, Morgan K, Dancik BP, Strobeck C, Yeh FCH (1984) Segregation of allozymes in megagametophytes of viable seed from a natural population of jack pine, *Pinus banksiana* Lamb. Theor Appl Genet 69:145–151

Clegg MT, Allard RW, Kahler AL (1972) is the gene the unit of selection? Evidence from two experimental plant populations. Proc Natl Acad Sci USA 69:2474–2478

Dempster AP, Laird NM, Rubin DB (1977) Maximum-likelihood from imcomplete data via the *EM* algorithm. J R Stat Soc, Ser B 39:1–38

Eckert RT, Joly RJ, Neale DB (1981) Genetics of isozyme variants and linkage relationships among allozyme loci in 35 eastern white pine clones. Can J For Res 11:573–579

El-Kassaby YA, Yeh FCH, Sziklai O (1982a) Inheritance of allozyme variants in coastal Douglas-fir (*Pseudotsuga menziesii* var. 'menziesii' (Mirb.) Franco). Can J Genet Cytol 24:325–335

El-Kassaby YA, Sziklai O, Yeh FCH (1982b) Linkage relationships among 19 polymorphic loci in coastal Douglas-fir (*Pseudotsuga menziesii* var. 'menziesii' (Mirb.) Franco). Can J Genet Cytol 24:101–108

Ellstrand NC, Foster KW (1983) Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor Appl Genet 66:323–327

Epperson BK, Allard RW (1984) Allozyme analysis of the mating system in lodgepole pine populations. J Hered 75: 212–214

Fienberg SE (1980) The analysis of cross-classified categorical data. MIT Press, Cambridge

Hill WG (1975) Tests for association of gene frequencies at several loci in random mating diploid populations. Biometrics 31:881–888

Kahler AL, Clegg MT, Allard RW (1975) Evolutionary changes in the mating system of an experimental population of barley (*Hordeum vulgare* L.). Proc Natl Acad Sci USA 72:943–946

Levene H (1949) On a matching problem arising in genetics. Ann Math Stat 20:91–94

Linhart YB, Mitton JB, Sturgeon KB, Davis ML (1981) Genetic variation in space and time in a population of ponderosa pine. Heredity 46:407–426

Lundkvist K (1974) Inheritance of leucine aminopeptidase isozymes in *Picea abies* K. Hereditas 76:91–96

Mitton JB, Linhart YB, Hamrick JL, Beckman JS (1977) Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range. Theor Appl Genet 51:5–13

Mitton JB, Linhart YB, Davis ML, Sturgeon KB (1981) Estimation of outcrossing in ponderosa pine, *Pinus ponderosa* Laws from patterns of segregation of protein polymorphisms and from frequencies of albino seedlings. Silvae Genet 30: 117–121

Moran GF, Bell JC, Matheson AC (1980) The genetic structure and levels of inbreeding in a *Pinus radiata* D. Don seed orchard. Silvae Genet 29:190–193

Moran GF, Brown AHD (1980) Temporal heterogeneity of outcrossing rates in Alpine ash (*Eucalyptus delegatensis* R.T. Bak.). Theor Appl Genet 57:101–105

Morris RW, Spieth PT (1978) Sampling strategies for using female gametophytes to estimate heterozygosity in conifers. Theor Appl Genet 51:217–222

Nei M, Li WH (1973) Linkage disequilibrium in subdivided populations. Genetics 75:213–219

O'Malley DM, Allendorf FW, Blake GM (1979) Inheritance isozyme variation and heterozygosity in *Pinus ponderosa*. Biochem Genet 17:233–250

Phillips MA, Brown AHD (1977) Mating system and hybridity in *Eucalyptus pauciflora*. Aust J Biol Sci 30:337–344

Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity 47:35–52

Ritland K, El-Kassaby YA (1985) The nature of inbreeding in a seed orchard of Douglas-fir as shown by an efficient multilocus model. Theor Appl Genet 71:375–384

Rao CR (1973) Linear statistical inference and its applications. Wiley, New York

Roberds JH, Brotschol JV (1985) Linkage disequilibria among allozyme loci in natural populations of *Liriodendron tulipifera* L. Silvae Genet 34:137–141

Roberds JH, Conkle MT (1984) Genetic structure in loblolly pine stands: allozyme variation in parents and progeny. For Sci 30:319–329

Rudin D (1977) Leucine-amino-peptidase (LAP) from needles and macrogametophytes of *Pinus sylvestris* L. Inheritance of allozymes. Hereditas 85:219–226

Schmidt RL (1960) Factors controlling the distribution of Douglas-fir in coastal British Columbia. Q J For 54: 156–160

Schoen DJ, Clegg MT (1986) Monte Carlo studies of plant mating system estimation models: the one-pollen parent and mixed mating models. Genetics 112:927–945

Shaw DV, Kahler AL, Allard RW (1981) A multilocus estimator of mating system parameters in plant populations. Proc Natl Acad Sci USA 78:1298–1302

Shaw DV, Allard RW (1982) Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor Appl Genet 62:113–120

Smith CAB (1957) Counting methods in genetical statistics. Ann Hum Genet 21:254–276

Smith CAB (1986) Chi-squared tests with small numbers. Ann Hum Genet 50:163–167

Sokal RS, Rohlf FJ (1981) Biometry, 2nd edn. Freeman, San Francisco, Calif

Sorensen FC (1969) Embryonic genetic load in coastal Douglas-fir, *Pseudotsuga menziesii* var. 'menziesii'. Am Nat 103: 389–398

Sorensen FC (1982) The roles of polyembryony and embryo viability in genetic system of conifers. Evolution 36: 725–733

Strauss SH, Conkle MT (1986) Segregation, linkage, and diversity of allozymes in knobcone pine. Theor Appl Genet 72:483–493

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35:235–254

Yeh FC, Layton C (1979) The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia*. Can J Genet Cytol 21: 487–503

Yeh FC, O'Malley DM (1980) Enzyme variation in natural populations of Douglas-fir, *Pseudotsuga menziesii* (Mirb.) Franco from British Columbia. 1. Genetic variation patterns in coastal populations. Silvae Genet 29:83–92

Yeh FC, Brune A, Cheliak WM, Chipman DC (1983) Mating system of *Eucalyptus citriodora* in a seed-production area. Can J For Res 13:1051–1055