Theor Appl Genet (1992) 83:500–508

THEORETICAL AND APPLIED GENETICS

© Springer-Verlag 1992

Pollen pool heterogeneity in jack pine (*Pinus banksiana* Lamb.): a problem for estimating outcrossing rates?

Y.B. Fu*,, P. Knowles and D.J. Perry*****

School of Forestry, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada

Received October 19, 1990; Accepted June 21, 1991
Communicated by A.L. Kahler

Summary. Pollen pool heterogeneity, which violates an assumption of the mixed-mating model, is a major potential problem in measuring plant mating systems. In this study, isozyme markers were used to examine pollen pool heterogeneity in two natural populations of jack pine, *Pinus banksiana* Lamb., in northwestern Ontario, Canada. Population multilocus estimates of outcrossing rate ranged from 0.83 to 0.95 and differed significantly between populations. Single-tree multilocus outcrossing rates were found to be homogeneous among trees in both populations. Computer simulation studies indicated that a consanguineous pollen pool (pollen gametes related to the mother tree) was capable of biasing population outcrossing estimates downward. Random pollen pool heterogeneity (uncorrelated with maternal genotypes) did not appear to affect population outcrossing estimates in the simulations. Heterogeneity *G*-tests and Spearman rank tests showed that pollen pool heterogeneity existed in the two natural populations examined; however, it did not have a major effect on population outcrossing estimates, since the consanguineous pollen pool detected was probably a relatively minor component of the outcross pollen pool in both populations. In addition, heterogeneity *G*-tests were found to be not sensitive in detecting pollen pool heterogeneity caused by consanguineous pollen pool.

Key words: Isozyme – Pollen pool heterogeneity – Mixed-mating model – Mating system – *Pinus banksiana* Lamb.

* To whom correspondence should be addressed
** Present address: Department of Botany, University of Toronto, Toronto, Ontario M5S 3B2, Canada
*** Present address: Department of Forest Resources, University of Minnesota, 1530 N. Cleveland Avenue, St. Paul/MN 55108, USA

Introduction

Reliable estimates of outcrossing are essential both for understanding the genetic structure of populations and for formulating breeding strategies and designing seed orchards (Brown 1990; Muona 1990). Plant mating system studies have greatly advanced since 1970 because of the use of isozyme markers and the development of appropriate statistical methods (Clegg 1980). The most commonly used procedure for estimating plant mating systems is the mixed-mating model (Brown and Allard 1970; Clegg 1980). This model assumes that mating takes place either by random outcrossing, at rate t, or by selfing, at rate $s=(1-t)$ (Jones 1916; Fyfe and Bailey 1951). The estimation procedure further assumes: (1) that outcross pollen allele frequencies, p(o), are constant for all maternal plants, (2) that the outcrossing rate is uniform for all maternal plants, and (3) that no selection affects the marker loci between the time of mating and the time when progeny genotypes are assayed (Clegg 1980). However, one or more of these assumptions are usually violated, especially in natural populations (Brown et al. 1985; Hamrick and Schnabel 1985). Schoen and Clegg (1984) indicated that violation of the assumptions may lead to inaccurate estimation of t. In fact, little is known about the bias of t from the violation of these assumptions.

The purpose of this study was to examine the effect of tree-to-tree variation in p(o) on the population estimates of t. Thus, it is necessary to determine whether heterogeneity of p(o) exists in populations. However, most previous related studies (e.g., Brown et al. 1975; King et al. 1984; Cheliak et al. 1985; Merzeau et al. 1989) provided only tentative evidence for the existence of heterogeneity of p(o). These studies usually used the conventional heterogeneity *G*-test (G_h; Sokal and Rohlf 1981) or χ^2 test to examine only the observed pollen pool (pollen gametes effective in fertilizing viable ovules, p), and p includes both self pollen pool, p(s), and outcross pollen pool, p(o). The observed heterogeneity detected by these tests may arise from the variation in t, in p(o), or in both (Brown et al. 1975). To distinguish

between these two sources, Perry and Knowles (1990) examined the correlations of the observed pollen allele frequencies in progeny of homozygous maternal trees with the estimated t and p(o) of the corresponding maternal trees, using Spearman's rank tests. However, they showed that the observed heterogeneity was probably due to the variation in t rather than in p(o).

In experimental studies, it is impossible to directly examine the effect of variation in p(o) on estimates of t because the true value of t is unobtainable. Thus, we conducted a computer simulation to examine the effect by determining which aspects of the outcross pollen pool could potentially affect population estimates of t. We expect that the consanguineous pollen pool [pollen gametes related to the mother tree; p(c)] is the major component of the p(o) in biasing population outcrossing estimates. There are several reasons for this expectation. First, since the multilocus mixed-mating model does not distinguish consanguineous mating (mating between relatives other than selfing), if any, from selfing (Ritland 1984; Hamrick and Schnabel 1985; Waller and Knight 1989), the p(o) of this model includes p(c) if p(c) exists. Thus, if p(c) exists, the estimate of p(o) will be biased and its joint estimate of t will probably be biased too. Second, evidence of consanguineous mating has been found in several species (Ritland and El-Kassaby 1985; Waller and Knight 1989; Barrett and Husband 1990). Third, several studies (Ennos and Clegg 1982; Ellstrand and Foster 1983; Shea 1987) have shown that family structure, which is one important source of a consanguineous pollen pool, has a great effect on t. And fourth, Brown et al. (1975) indicated that the net effect of pollen pool heterogeneity is a downward bias in t, but Fripp et al. (1987) and Sampson et al. (1990) reported that temporal heterogeneity of the outcross pollen pool did not have a major effect on population estimates of t. This suggests that some components of the spatial variation of p(o) are the possible biasing factors.

If this expectation is correct, the direct examination of p(c) will provide information about the bias of t. However, there are no efficient methods available to measure or detect p(c). Ritland and El-Kassaby (1985) used the regression of outcross pollen allele frequency on the additive value of the ovule genotype to measure consanguineous mating, but this method is based on single-tree estimates which usually have high variances. Shaw et al. (1981) indicated that the comparison between multilocus t and the average of single-locus t can detect consanguineous mating, but the estimates of t are easily biased by the number of loci used (Barrett and Husband 1990) and/or linkage disequilibrium, if this exists (Hedrick and Ritland 1990). Therefore, new efficient methods are needed for examining p(c). In this study, we present an indirect approach to detect p(c). This method is based on two outcrossing estimates for individual trees in which only estimates of p(o) differ, and it seems to be conservative.

In this study, isozyme markers were used to examine pollen pool heterogeneity in two natural populations of jack pine (*Pinus banksiana* Lamb.) in northwestern Ontario, Canada. The specific objectives were: (1) to determine whether heterogeneity of the outcross pollen pool existed in populations, and (2) to assess the effect of tree-to-tree variation in p(o) on the population estimates of t by combining experiments and computer simulations.

Materials and methods

Experimental studies

Study sites included two natural populations of jack pine, GLFP and CP11, near Thunder Bay, Ontario, Canada. GLFP, with a density of 5,939 stems/ha and an average tree height of 8 m, was a young stand (16 years old) originating via seeding from remnant slash following logging. CP11 was a mature stand of fire origin (104 years old), with a density of 3,750 stems/ha and an average tree height of 22 m. Twelve trees were randomly selected from each of the two populations. In both populations, most of the sampled trees were located at least 10 m apart, within an area of 60 × 60 m. The seeds from cones collected from the top crown of the sampled trees were extracted by hand, stored below 0 °C, and maintained by individual mother-tree indentity throughout storage and processing.

Seeds were germinated on moist filter paper until seedlings grew a 3- to 5-mm radicle (about 5 days). Each emerging embryo was excised from its surrounding megagametophyte and both tissues were homogenized separately, in 20 and 30 µl of extraction buffer (Yeh and O'Malley 1980), respectively. A total of 80 such embryo-megagametophyte pairs was assayed for each tree.

Electrophoretic procedures, staining recipes, and enzyme nomenclature followed those of Cheliak and Pitel (1984). The following enzyme systems were assayed: aconitase (ACO; E.C.4.2.1.3); esterase (fluorescent) (F-EST; E.C.3.1.1.1); glutamate dehydrogenase (GDH; E.C.1.4.1.3); malate dehydrogenase (MDH; E.C.1.1.1.37); malic enzyme (ME; E.C.1.1.1.40); 6-phosphogluconate dehydrogenase (6PGD; E.C.1.1.1.44); phosphoglucose isomerase (PGI; E.C.5.3.1.9); and phosphoglucomutase (PGM; E.C.2.7.5.1). These eight systems yielded nine polymorphic loci (*Aco*, *F-Est3*, *Gdh*, *Mdh1*, *Me*, *6Pgd1*, *6Pgd2*, *Pgi*, *Pgm*) that could be clearly scored for both haploid megagametophyte and diploid embryo tissues. Examination of the available data in this study and comparison with related references (Cheliak et al. 1984; Knowles 1985; Snyder et al. 1985; Tobolski 1979) showed that the nine employed isozymes exhibited distinct, codominant expression and simple Mendelian segregation in their mode of inheritance.

Population single-(t_s) and multilocus (t_m) estimates of t and p(o) were calculated for both populations, using the maximum likelihood procedures of Ritland and El-Kassaby (1985). Tests of the proportions (Zar 1974) of allele frequencies were conducted to examine the differences in allele frequencies between outcross pollen and maternal parents for each of the populations. Tests for heterogeneity for both t_s among loci and t_m among populations, and tests of the hypotheses $H_o: t_{(m \text{ or } s)} = 1$ were conducted using the likelihood ratio test (Rao 1973).

Single-tree multilocus estimates of outcrossing rate (t_f) (f denotes single tree) and $p(o)_f$ for both populations, also using the maximum likelihood procedures of Ritland and El-Kassaby (1985), were made to: (1) determine heterogeneity of t_f (t_{f2}, see definition below) among trees using a likelihood ratio test (Rao 1973), and (2) detect p(c). Two methods for t_f estimation were used as follows: one (t_{f1}) with p(o) jointly with the combined data of the population, and the other (t_{f2}) with joint estimation of t_f and $p(o)_f$ for each tree. Obviously, the t_{f1} estimation assumes homogeneity of outcross pollen allele frequencies, and the t_{f2} estimation does not. With the second method (t_{f2}), the outcross pollen allele frequencies, which include the consanguineous pollen pool, can be estimated for single trees (Ritland and El-Kassaby 1985). Therefore, any differences in estimates obtained using these two methods (t_{f1} and t_{f2}) may be due to the consanguineous pollen pool. The Wilcoxon signed-rank test (Zar 1974) was used to assess this difference. Any significant differences between t_{f1} and t_{f2} detected by this test may suggest the existence of a consanguineous pollen pool. The Wilcoxon signed-rank test was chosen because both the nature of the distribution of estimated values per tree and the correlation, if any, of two estimates (t_{f1}, t_{f2}) per tree are unknown.

The particular structure of coniferous seeds (haploid megagametophyte and diploid embryo) allows inference of pol-

len contributions to each embryo, because the ovule from which a given seed arises shares the same haploid genotype as its associated megagametophyte (Müller 1976). The observed pollen allele frequencies in progeny arrays of maternal trees with like genotypes were compared in contingency tables using heterogeneity G-tests. One G_h value per locus was obtained by summation over genotypes. This test is a modification of the method described by Brown et al. (1975) in which only homozygous maternal parents were included using angiosperm data. We were not restricted to homozygous maternal parents since, in conifers, the pollen contribution to viable embryos may be determined regardless of maternal genotype. As mentioned in the introduction, the observed heterogeneity detected by G-tests may arise from the variation in t_f or in $p(o)_f$.

To determine whether heterogeneity of outcross pollen pool existed, we calculated the following Spearman's rank correlations: (1) $r_{s(t)}$, the correlation between t_{f2} and the frequency of alternate alleles (alleles other than the A allele) observed in progeny arrays of AA homozygous maternal trees; and (2) $r_{s(p)}$, the correlation between the alternate allele frequencies of $p(o)_f$ estimated jointly with t_{f2} and the frequencies of these alternate alleles observed in progeny of AA homozygous maternal parents. The results obtained from these tests may be one of the four possible types: (1) that only $r_{s(p)}$ is significant, (2) that only $r_{s(t)}$ is significant, (3) that both $r_{s(p)}$ and $r_{s(t)}$ are significant, and (4) that none of them is significant. However, only type 1 can strongly suggest that heterogeneity of $p(o)$ existed. It is worth noting that this method is conservative because of small sample sizes (Perry and Knowles 1990).

Simulation studies

To determine which aspects of the outcross pollen pool could potentially affect population estimates of t, we developed three models, among which only $p(o)$ differs for the following pollen pools: (1) homogeneous pollen pools (p_{ho}), (2) random pollen pool heterogeneity (p_{her}), and (3) pollen pool heterogeneity with consanguineous (half-sib) mating (p_{hec}). For all three models, the frequency of allele A in the pollen gametes of mother tree f is

$$p_f = s\,p(s)_f + t\,p(o)_f, \quad (1)$$

where s and t are the selfing and outcrossing rates, $p(s)_f$ and $p(o)_f$ are the frequencies of allele A in self and outcross pollen gametes of tree f, respectively. However, each model has its own $p(o)_f$ defined differently as follows:

1) For p_{ho} model, all $p(o)_f = p(o)$ (i.e., all tree have same outcross pollen allele frequencies).

2) For p_{her} model, $p(o)_f$ is altered by the frequency of allele A in a randomly sampled neighbor tree f.

3) For p_{hec} model, $p(o)_f$ is correlated with the frequency of allele A in the maternal tree f.

To make the variation among $p(o)_f$ in the simulated populations comparable to that in the experimental populations, we used the genotypes of the 12 parental trees for the CP11 site and the estimated population outcross pollen allele frequencies for the nine assayed loci in the simulations. The simulations of each model had outcross pollen pools constructed as follows:

1) p_{ho}: $p(o)_f$ was simply $p(o)$ estimated for CP11.

2) p_{her}: $p(o)_f$ was equal to 0.7 times its frequency in the $p(o)$ of CP 11 plus 0.3 times its frequency (0, 0.5, or 1.0) in a neighbor tree randomly drawn from a hypothetical, Hardy-Weinberg equilibrium (assumed) population, with allele frequencies identical to those estimated for CP11. A high deviation of 0.3 was arbitrarily chosen to make the bias of t easier to detect. For each allele, the additive values 0, 0.5, or 1.0 were assigned to genotypes *aa*, *Aa* or *AA*, where '*A*' was the allele in question and '*a*' was the class of the other alleles at that locus. A neighbor tree was randomly drawn for each sample tree.

3) p_{hec}: $p(o)_f$ was equal to 0.7 times its frequency in the $p(o)$ of CP11 plus 0.3 times its frequency in the maternal tree (0, 0.5, or 1.0).

In all simulations, $p(s)_f$ was the same allele frequencies as the sample tree (0, 0.5, or 1.0).

Simulations of each model consisted of three replications, each with a sample of 80 seeds per tree. Each seed was determined to be outcrossed or selfed by sampling from a uniform distribution, so that the probability of outcrossing for each tree in 90%. The outcrossing rate of 90% was arbitrarily chosen from the potential values based on empirical work. The genotype of an outcrossed seed for a locus was determined by a second sampling from a uniform distribution, so that the probabilities were $p(o)_f^2$, $2p(o)_f[1-p(o)_f]$, $[1-p(o)_f]^2$ for *AA*, *Aa*, *aa*, respectively, noting that $p(o)_f$ was adjusted for the different models and for nine loci. Selection of a pollen genotype at one locus was independent of its genotype determined at other loci. For the genotypes of selfed seeds, the same method as for outcrossed seeds was used but $p(s)_f$ applied. The simulated data were analyzed by the same procedures as given in the experimental studies. The bias of t against the sampled t due to the variation in $p(o)$ was calculated as: $(t_{estimated} - t_{sample})/t_{sample}$.

Results

No significant differences in allele frequencies were observed between the ovule and outcross pollen pools within the two populations, with the exception of *Aco* in CP11 (Table 1). This indicated that the observed pollen pool was a random sample of the adult trees, as represented by the maternal trees.

Population estimates of t_m ranged from 0.83 to 0.95 with a mean of 0.89 and were heterogeneous ($P<0.01$) between populations (Table 2). The estimates of t_m for CP11 were significantly lower than for GLFP (Table 2). All t_m estimates were significantly less than 1. Population estimates of t_s ranged from 0.45 to 1.00, and were heterogeneous ($P<0.05$) over loci within each population (Table 2). Estimates of t_s showed a significant departure from complete outcrossing ($t=1.0$) at *Aco* in GLFP and at *Aco* and *6Pgd1* in CP11 (Table 2). Since the arithmetic mean over loci was an unrealistic estimate due to the low estimate of t_s for the *Aco* locus, minimum variance estimates of t were calculated (El-Kassaby et al. 1985) and used in the analysis. The differences between t_m and mean t_s were relatively large in GLFP but not in CP11 (Table 2). When *Aco* is excluded from the analysis, estimates of t_m and the mean of t_s are very high in both populations (0.98:0.98 in GLFP and 0.94:0.94 in CP11, respectively), which is contradictory to those ($t=0.88$) of previous studies in jack pine.

The mean estimates of t_{f1} in the two populations were close to the population estimates of t_m (Table 3). Estimates of t_{f2} ranged from 0.54 to 1.58 (Table 3) and were homogeneous ($P>0.05$) among trees in both populations (see Table 4), indicating that the assumption of uniformi-

Table 1. Allele frequencies for ovule and outcross pollen pools for GLFP and CP11 populations of *Pinus banksiana* (SE in parentheses)

Locus/ Allele		GLFP		CP11	
		Ovule	Outcross pollen	Ovule	Outcross pollen
Aco	1	0.42 (0.10)	0.35 (0.02)	0.38 (0.10)	0.32 (0.02)
	2	0.58 (0.10)	0.65 (0.02)	0.54 (0.10)	0.67 (0.02)
	3	0.00 (–)	0.01 (0.00)	0.08 (0.06)	0.01 (0.00)*
F-Est3	1	0.00 (–)	<0.01 (0.00)	0.00 (–)	0.01 (0.00)
	2	1.00 (0.00)	>0.99 (0.00)	1.00 (0.00)	0.99 (0.00)
	3	0.00 (–)	<0.01 (0.00)	0.00 (–)	0.00 (–)
Gdh	1	0.00 (–)	0.02 (0.00)	0.00 (–)	0.05 (0.01)
	2	1.00 (0.00)	0.98 (0.00)	1.00 (0.00)	0.95 (0.01)
Mdh1	1	0.88 (0.07)	0.92 (0.01)	1.00 (0.00)	0.95 (0.01)
	2	0.13 (0.07)	0.08 (0.01)	0.00 (–)	0.05 (0.01)
Me	1	0.00 (–)	<0.01 (0.00)	0.00 (–)	<0.01 (0.00)
	2	1.00 (0.00)	>0.99 (0.00)	1.00 (0.00)	>0.99 (0.00)
6Pgd1	1	0.38 (0.10)	0.50 (0.02)	0.58 (0.10)	0.47 (0.02)
	2	0.63 (0.10)	0.50 (0.02)	0.42 (0.10)	0.53 (0.02)
6Pgd2	1	0.17 (0.08)	0.11 (0.01)	0.04 (0.04)	0.13 (0.01)
	2	0.04 (0.04)	0.01 (0.00)	0.00 (–)	0.02 (0.00)
	3	0.79 (0.08)	0.84 (0.01)	0.92 (0.06)	0.81 (0.02)
	4	0.00 (–)	0.04 (0.01)	0.04 (0.04)	0.04 (0.01)
Pgi	1	0.92 (0.06)	0.94 (0.04)	0.96 (0.04)	0.92 (0.04)
	2	0.08 (0.06)	0.06 (0.01)	0.04 (0.04)	0.08 (0.01)
Pgm	1	0.00 (–)	<0.01 (0.00)	0.00 (–)	<0.01 (0.00)
	2	1.00 (0.00)	>0.99 (0.02)	1.00 (0.00)	>0.99 (0.00)
	3	0.00 (–)	0.00 (–)	0.00 (–)	<0.01 (0.00)
Sample size		12	960	12	960

* Significant at 0.05 level

Table 2. Single-locus and multilocus population estimates of outcrossing rate for GLFP and CP11 populations of *Pinus banksiana* (SE in parentheses)

	GLFP	CP11
Single-locus		
Aco	0.45 (0.05)[a]	0.55 (0.03)[a]
F-Est3	–	–
Gdh	–	–
Mdh1	0.98 (0.03)	–
Me	–	–
6Pgd1	0.95 (0.05)	0.89 (0.05)[a]
6Pgd2	1.00 (0.04)	0.99 (0.04)
Pgi	0.99 (0.04)	0.88 (0.08)
Pgm	–	–
Mean		
t_s[b]	0.90 (0.02)[c]	0.80 (0.02)[c]
Multilocus		
t_m[d]	0.94 (0.02)[a]	0.82 (0.02)[a]
t_m[e]	0.95 (0.02)[a]	0.83 (0.02)[a]
Sample size	960	960

[a] Significant ($P<0.05$) departure from H_o: $t=1$
[b] Minimum variance mean
[c] Significant ($P<0.05$) heterogeneity of t_s among loci
[d] Multilocus estimate excluding the maternal genotype loci that have insufficient variation for single-locus estimates
[e] Multilocus estimate for all nine loci
–: Insufficient maternal genotype classes for estimation of t_s

ty of t_f for all maternal trees was valid for the population estimates of t. Estimates of t_{f2} were significantly less than 1 for two trees in GLFP and one tree in CP11 (Table 3). Wilcoxon signed-rank tests showed no significant differences between t_{f1} and t_{f2} in either population (see Table 4).

G-tests indicated significant heterogeneity of pollen allele frequencies in viable progeny among trees in three of nine loci for GLFP and in four of nine loci for CP11 (Table 5). Spearman rank correlations, $r_{s(t)}$, showed two significant negative values at *Gdh* and *Mdh1* in GLFP and one significant negative value at *Gdh* in CP11 (Table 5). Significant positive values for $r_{s(p)}$ were obtained at all loci in CP11 and in six of eight loci in GLFP (Table 5).

The results of the simulation studies are given in Tables 4 and 6, and are summarized as follows. (1) In all simulations, estimates of t_{f2} were homogeneous over maternal trees ($P>0.05$; Table 4). (2) The *G*-test was substantially less sensitive in detecting pollen pool heterogeneity when a consanguineous pollen pool existed (Table 4). (3) For population estimations, no apparent biases for either t_m or t_s were observed in p_{ho} or p_{her} simulations (Tables 4 and 6), i.e., random pollen pool heterogeneity did not appear to affect the population estimates of t. However, in p_{hec} simulations, single-locus estimates were biased down by 30.5% from the true sample value, nearly twice as much as the multilocus estimates (16.5%) (Tables 4 and 6). (4) The Wilcoxon signed-rank test for the existence of a consanguineous pollen pool (Table 4) showed its high reliability in the simulation studies. (5) For the t_{f2}, there were no apparent differences between mean sample t and mean estimated t_{f2} in all simulations (Table 6).

Discussion

The experimental studies reported above indicated that spatial heterogeneity of outcross pollen pool existed in two natural populations of jack pine. The *G*-tests showed significant heterogeneity of observed pollen pools at about 40% of the loci in both populations. Spearman rank correlations indicated that the observed pollen pool heterogeneity was probably due to actual heterogeneity of outcross pollen allele frequencies rather than actual variation of t. Moreover, this conclusion is strongly supported by the finding that no significant tree-to-tree vari-

Table 3. Multilocus single-tree estimates of outcrossing rate for two populations of *Pinus banksiana* using two methods: (1) $p(o)_f$ set constant and equal to estimated population frequencies, t_{f1}; and (2) $p(o)_f$ allowed to vary among trees [joint estimation of t_f with $p(o)_f$], t_{f2} (SE in parentheses)

Single-tree	GLFP		CP11	
	t_{f1}	t_{f2}	t_{f1}	t_{f2}
1	1.03	1.12 (0.18)	0.89	–
2	0.78	1.58 (0.82)	0.95	1.02 (0.09)
3	1.02	0.94 (0.09)	0.50	1.05 (0.91)
4	1.15	–	0.87	0.92 (0.20)
5	0.81	0.54 (0.17)	0.84	0.84 (0.10)
6	0.95	0.89 (0.07)[a]	1.13	1.02 (0.06)
7	0.84	1.48 (0.42)[a]	1.06	0.93 (0.09)
8	0.91	1.58 (1.27)	0.65	0.86 (0.15)
9	0.94	–	0.81	0.76 (0.24)[a]
10	–	1.01 (0.03)	0.85	0.92 (0.07)
11	0.91	0.95 (0.11)	0.72	1.03 (0.26)
12	0.97	1.00 (0.06)	0.50	0.83 (0.56)
Mean				
t_f[b]	0.94	1.11 (0.16)	0.81	0.92 (0.11)
t_f[c]		0.98 (0.03)		0.95 (0.03)
Population				
t_m[d]		0.94 (0.02)[a]		0.82 (0.02)[a]

[a] Significant ($P<0.05$) departure from H_o: $t=1$
[b] Arithmetic mean was calculated because the variance for the single-tree is unobtainable using the first method
[c] Minimum variance mean
[d] Population multilocus estimate of outcrossing rate excluding nonconvergent single-trees (see Ritland and El-Kassaby 1985)
–: Single-tree was not convergent (see Ritland and El-Kassaby 1985)

ation in t was observed in this study. The experimental studies aimed at assessing levels of pollen pool heterogeneity were not designed to address causal factors, and the sources of the heterogeneity detected in this study remain obscure. Theoretically, stochastic factors of pollen flow, microhabitat selection, and family structure can result in spatial heterogeneity of the outcross pollen pool (Shaw et al. 1981; Bijlsma et al. 1986; Fripp et al. 1987; Merzeau et al. 1989).

Direct examination of the effect of variation in $p(o)$ on estimates of t is impossible in an experimental study because the true value of t is unobtainable. Thus, we developed an indirect approach for examining its effect in which a computer simulation was used. The results obtained using this approach indicated that heterogeneity of outcross pollen pools detected in the experimental portion probably did not have a major effect on population outcrossing estimates.

First, the simulation studies showed that pollen pool heterogeneity with consanguineous mating biased estimates of t downward, but random pollen pool heterogeneity did not. As indicated, these simulated pollen pools differed only in $p(o)_f$. To understand these results, let us assume that $p(o)_f=(1-D)p(o)_f^*+Dp(x)_f$, where D is the deviation of $p(o)_f$ from $p(x)_f$ (the frequency of allele A in the unknown proportion of outcross pollen gametes of tree f) and $p(o)_f^*$ is the biased $p(o)_f$. If $p(x)_f$ is known as $p(c)_f$ (the frequency of allele A in the consanguineous pollen gametes of tree f), then $p(o)_f=(1-D)p(o)_f^*+Dp(c)_f$, noting that D here is t_c/t

Table 4. The variation in p and its effect on the population estimates of outcrossing rate in *Pinus banksiana* (degrees of freedom in parentheses)

Population or simulation model	Replication	*G*-test for heterogeneity of p[a]	Likelihood ratio test for heterogeneity of t_{f2}	Wilcoxon signed-rank test for $t_{f1}-t_{f2}$	% Bias	
					t_m	t_s
GLFP		3	14.4 (10) NS	13 (9) NS		
CP11		4	12.2 (11) NS	16 (11) NS		
p_{ho}	S11	0	6.7 (12) NS	20 (12) NS	−3.0	−6.9
	S12	0	12.5 (11) NS	29 (11) NS	0.4	0.4
	S13	0	5.5 (11) NS	29 (11) NS	0.8	0.0
	mean	0			−0.6	−2.1
p_{her}	S21	7	13.7 (12) NS	18 (12) NS	−0.3	2.2
	S22	8	12.3 (12) NS	27 (12) NS	−0.4	−4.1
	S23	8	13.2 (12) NS	24 (12) NS	−0.4	1.9
	mean	7.7			−0.3	0.0
p_{hec}	S31	3	8.3 (11) NS	1 (11)***	−16.4	−30.9
	S32	1	10.2 (12) NS	0 (12)***	−18.9	−32.1
	S33	2	6.3 (12) NS	3 (12)***	−14.3	−28.8
	Mean	2			−16.5	−30.5

[a] The number out of nine loci that were significant ($P<0.05$)
NS Not significant at 0.05 level
*** Significant at 0.01 level

(t_c is the level of consanguineous mating). In this case, we can extend Eq. 1 with $p(o)_f$ and $D=t_c/t$ as $p_f=s\,p(s)_f+(t-t_c)\,p(o)_f^*+t_c\,p(c)_f$. Thus, the outcrossing estimate is expected to be $(t-t_c)$ and will be biased downward by t_c. If $p(x)_f$ is not $p(c)_f$, $p(x)_f$ can still be assumed as the frequency of allele *A* in *some* proportion of outcross pollen gametes of tree f. Thus, the estimate of $p(o)_f$ remains unbiased and the outcrossing estimate is expected to be unbiased when $p(o)_f$ varies randomly. To examine our expectation about the degree of bias of t by t_c, we conducted further simulations of p_{hec} model with seven levels of D% (5, 7, 10, 15, 20, 30, 50), each with three replications. The results (Fig. 1) indicate some trends: (1) that the degree of bias increases as the level of D increases, and (2) that t_m is less biased than t_s. These results are identical to those of related studies (Ritland and Jain 1981; Ritland 1984; Schoen and Clegg 1984). While the replications in the simulations of this study are too small to make any generalizations, the results obtained indicate a pattern that the consanguineous pollen pool was capable of biasing population estimates of t.

Second, according to Wilcoxon signed-rank tests, the consanguineous pollen pool was a relatively minor component of the outcross pollen pools in the experimental populations. However, the power of this test is probably very poor, because the single-tree estimates on which it is based have high variances, and also because the difference in single-tree estimates obtained using two different methods could be due to other factors, such as chance and temporal variation of p(o). To draw a reliable conclusion about the existence of p(c), we performed a direct measure of consanguineous mating level following the regression method of Ritland and El-Kassaby (1985), in which the regression coefficient is the indicator of consanguineous mating. The regressions obtained have nonsignificant pooled slopes for both populations ($b_{glfp}=0.05$ and $b_{cp11}=0.02$). We also employed the likelihood ratio test (Rao 1973) to examine the difference between t_m and $\bar{t}_s$. As Shaw et al. (1981) indicated, the

Table 5. The results of *G*-test and Spearman rank test for pollen pool heterogeneity in *Pinus banksiana* (degrees of freedom in parentheses)

Locus	*G*-test	Spearman rank test	
		$r_{s(t)}$	$r_{s(p)}$
	GLFP		
Aco	81.75 (18)***	–[a]	–[a]
F-Est3	15.99 (22)	0.14 (8)	0.96 (8)***
Gdh	19.48 (11)	−0.65 (8)*	0.97 (8)***
Mdh1	19.08 (9)*	−0.74 (6)*	0.97 (6)***
Me	8.35 (11)	−0.34 (8)	0.99 (8)***
6Pgd1	17.36 (9)*	0.80 (2)	0.20 (2)
6Pgd2	28.33 (25)	0.23 (6)	0.26 (6)
Pgi	8.70 (10)	0.29 (7)	0.74 (7)*
Pgm	8.35 (11)	0.19 (8)	0.98 (8)***
	GP11		
Aco	157.08 (13)***	0.60 (3)	1.00 (3)***
F-Est3	10.38 (11)	−0.07 (9)	0.94 (9)***
Gdh	84.00 (11)***	−0.91 (9)***	0.98 (9)***
Mdh1	18.83 (11)	−0.35 (9)	0.98 (9)***
Me	7.19 (11)	0.40 (9)	1.00 (9)***
6Pgd1	36.38 (9)***	−0.40 (2)	1.00 (2)***
6Pgd2	38.83 (30)	0.36 (8)	0.89 (8)***
Pgi	19.80 (10)*	−0.34 (8)	1.00 (8)***
Pgm	12.12 (22)	0.13 (9)	0.99 (9)***

* $P<0.05$; *** $P<0.001$

[a] No homozygous maternal genotypes included in sampled trees

Table 6. Population and single-tree estimates of outcrossing rate in the simulation studies (SE in parentheses)

Simulation model	Replication	Sample	Population[a]		Single-tree	
		t	t_m	$\bar{t}_s$[b]	$\bar{t}_{f1}$[c]	$\bar{t}_{f2}$[b]
p_{ho}	S11	0.88	0.85 (0.02)	0.82 (0.02) ns	0.85	0.88 (0.02)
	S12	0.90	0.90 (0.02)	0.90 (0.02) ns	0.90	0.90 (0.02)
	S13	0.90	0.91 (0.02)	0.90 (0.02) ns	0.91	0.90 (0.02)
	mean	0.89	0.89 NS	0.87	0.89	0.89
p_{her}	S21	0.91	0.90 (0.02)	0.93 (0.02) ns	0.91	0.91 (0.02)
	S22	0.90	0.89 (0.02)	0.86 (0.02)*	0.90	0.91 (0.02)
	S23	0.90	0.90 (0.02)	0.92 (0.02) ns	0.91	0.87 (0.02)
	mean	0.90	0.90 NS	0.90	0.91	0.90
p_{hec}	S31	0.91	0.76 (0.02)	0.63 (0.02) ns	0.75	0.90 (0.04)
	S32	0.91	0.74 (0.02)	0.62 (0.02) ns	0.73	0.86 (0.03)
	S33	0.91	0.78 (0.02)	0.65 (0.02) ns	0.78	0.87 (0.03)
	Mean	0.91	0.76 NS	0.63	0.75	0.88

[a] Estimates of t_m were based on nine loci; t_s on four loci

[b] Minimum variance mean

[c] Arithmetic mean

NS Not significant heterogeneity among replications ($P>0.05$)

* Significant ($P<0.05$) heterogeneity of t_s among loci; ns, not significant ($P<0.05$) heterogeneity of t_s among loci

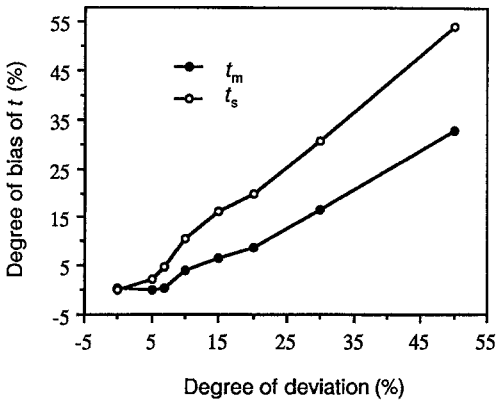

Fig. 1. Degree of bias of outcrossing estimate with the deviation level of consanguineous pollen pool from the outcross pollen pool

average of t_s is expected to be lower than the corresponding t_m when cross-pollination occurs among family members. The results obtained indicated that the consanguineous pollen pool existed in GLFP but not in CP11. Overall, the results obtained using three methods, although slightly different, are generally similar to those of previous studies of jack pine (Cheliak et al. 1985; Snyder et al. 1985). A study of the family structure in these same stands using spatial autocorrelation analysis of genetic data (C.Y. Xie, unpublished results) found that a very weak family structure existed, which indirectly supports the conclusion about the existence of p(c).

Third, comparison of the results between the experimental and simulation studies (Tables 2, 3, and 6) also supports the conclusion about the existence of consanguineous pollen pools in the experimental populations. We expect that the average of single-tree multilocus outcrossing estimates is approximately equal to population multilocus or single-locus outcrossing estimates [i.e., $\bar{t}_{f2} = t_{(m \text{ or } s)}$] if there is no consanguineous mating. In contrast, if there is consanguineous mating, $\bar{t}_{f2} > t_{(m \text{ or } s)}$. The simulation study confirmed these two expectations (Table 6). The results in either studied population (Tables 2 and 3) are similar to the latter expectation, which suggests that the consanguineous pollen pool existed in both populations. However, the differences between $\bar{t}_{f2}$ and t_s (0.08 for GLFP and 0.15 for CP11) are much smaller than that (0.25) in the simulations of p_{hec} model, implying that there were more variations of $p(o)_f$ in the simulated populations. This may be due to a high deviation of 0.3 used in the simulations. Moreover, natural pollen pools are probably more complex than the simulation models employed, and may involve factors such as selection and linkage disequilibrium which were not considered. Thus, this comparison can only provide an indirect evidence for the existence of p(c).

The mean population estimate of outcrossing obtained in this study is 89%, which is similar to those of previous studies using either isozyme markers (Cheliak et al. 1985; Snyder et al. 1985) or morphological markers (Fowler 1965, Sittman and Tyson 1971; Rudolph 1979). However, the estimates of outcrossing differed significantly between GLFP and CP11. The difference in stand age may result in the difference in outcrossing (Cheliak et al. 1985; Synder et al. 1985; Shea 1987). CP11 was much older (104 years) than GLFP (16 years). This might explain the observed difference in outcrossing. GLFP and CP11 also showed a difference in density, which might result in the observed difference in outcrossing. Some evidence for a positive relationship between stand density and outcrossing has been found in ponderosa pine (*Pinus ponderosa*) (Farris and Mitton 1984) and tamarack (*Larix laricina*) (Knowles et al. 1987). However, the density in either population was very high. Moreover, various generation methods could result in the differences both in stand structure and in average self-fertility of trees, which may in turn produce the difference in outcrossing. We expect that more stand structure could develop in CP11 with natural seeding after fire than in GLFP originating via seeding from remnant slash after logging. However, the study of family structure in the same stands (C. Y. Xie, unpublished results) produced results opposite to this expectation. Whether the difference in self-fertility existed between two populations is not known. No report has been found relating regeneration history to variation of outcrossing.

Although the significant difference between pollen and ovule allele frequencies at *Aco* in CP 11 (Table 1) could be due to chance, it is not clear why the *Aco* locus displayed abnormally high selfing rates relative to the other loci both in two natural stands (Table 2) and in one simulated replication (S22, not shown). A phenomenon similar to that observed for the *Aco* locus has been reported for the esterase locus (*Est*) in a few other conifers (see Yeh and Morgan 1987 for discussion). An explanation previously proposed for *Est* in loblolly pine (*Pinus taeda* L.) is that selection might have been the contributing factor by acting either directly on the *Est* or at correlated loci (Roberds and Conkle 1984). The significant negative $r_{s(t)}$ values (Table 5) suggest that epistatic selection favoring combination of genotypes (Epperson and Allard 1989) might act on *Gdh* and *Mdh1* but not on *Aco*. When *Aco* is excluded from the analysis, population estimates of t_m and the minimum variance mean of t_s are very high in both populations, which are contradictory to those ($t = 0.88$) of previous studies in jack pine.

In this study, we also found that the heterogeneity *G*-test was not sensitive in detecting pollen pool heterogeneity caused by consanguineous mating (Table 4). This is not unexpected since: (1) the pollen pool of like genotypes was similarly affected when the variation in p(o) was correlated with the maternal genotypes (as in consanguineous mating), and (2) the *G*-test only tested for

heterogeneity among like genotypes. However, it is not clear why the results of the *G*-test should differ between p_{hec} and p_{ho} (Table 4). For this reason 11 further replications were performed for each of these two models. In these replications, significant results were obtained for 7 of the 99 tests (9 loci × 11 replications) in p_{hec} and for 10 of the 99 tests in p_{ho}. Thus, the results of *G*-tests in the p_{hec} model presented in Table 4 were misleading and the discrepancy between p_{hec} and p_{ho} may be due to the small numbers of replications. It is also apparent that more significant test results were obtained in p_{ho} than expected by chance at the 0.05 level. These spurious test results occurred mainly at loci with extremely uneven allele frequencies (e.g., *Me*) and may thus be due to large numbers of empty cells in the matrices tested.

In conclusion, heterogeneity of outcross pollen pools existed in two natural populations of jack pine. It did not have a major effect on the population outcrossing estimates, since the consanguineous pollen pool probably was a relatively minor component of the outcross pollen pool in both populations. Computer simulations indicated that the consanguineous pollen pool was capable of biasing outcrossing estimates. Therefore, the consanguineous pollen pool should be considered when the mixed-mating model is used. The method for examining the consanguineous pollen pool reported here using the Wilcoxon signed-rank test seems to be very conservative. While the other methods with similar limitations are still available, an efficient, direct method is required. Investigations of the genetic basis of pollen pool variations and of their causal factors will be a difficult challenge. Such studies will not only provide information for evaluating the reliability of estimates of mating system parameters, but may also provide insights into the evolution of plant mating systems.

Acknowledgements. We are greatly indebted to the many people who contributed directly and indirectly to this work. These include M. Berzowski, K. M. Brown, J. R. Coates, H. Foster, R. Foster, D. W. Ip, M. Lesyk, A. McDonald, W. H. Parker, and C. Y. Xie. We thank Drs. W. T. Adams, K. Ritland, A. H. D. Brown, and M. T. Clegg for their critical comments on the manuscript. This work was supported by NSERC Grant A5222 to P. K.

References

Barrett SCH, Husband BC (1990) Variation in outcrossing rates in *Eichhornia paniculata*: the role of demographic and reproductive factors. Plant Species Biol 5:41–55

Bijlsma R, Allard RW, Kahler AL (1986) Nonrandom mating in an open-pollinated maize population. Genetics 112:669–680

Brown AHD (1990) Genetic characterization of plant mating systems. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland/MA, pp 145–162

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

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

Brown AHD, Barrett SCH, Moran GF (1985) Mating system estimation in forest trees: models, methods and meaning. In: Gregorius H (ed) Population genetics in forestry. Springer, Göttingen pp 32–49

Cheliak WM, Pitel JA (1984) Techniques for starch gel electrophoresis of enzymes from forest tree species. Inf Rep PI-X-42, Can For Serv, Environment Canada, Ottawa

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

Cheliak WM, Dancik BP, Morgan K, Yeh FCH, Strobeck C (1985) Temporal variation of the mating system in a natural population of jack pine. Genetics 109:569–584

Clegg MT (1980) Measuring plant mating systems. Bioscience 30:814–818

El-Kassaby YA, Parkinson J, Devitt WJB (1985) The effect of crown segment on the mating system in a Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) seed orchard. Silvae Genet 35(4):149–155

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

Ennos RA, Clegg MT (1982) Effect of population substructuring on estimates of outcrossing rate in plant populations. Heredity 48:283–292

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of genotypes within populations of lodgepole pine. Genetics 121:369–377

Farris MA, Mitton JB (1984) Population density, outcrossing rate, and heterozygote superiority in ponderosa pine. Evolution 38:1151–1154

Fowler DP (1965) Natural self-fertilization in three jack pines and its implications in seed orchard management. For Sci 11(1):56–58

Fripp YJ, Griffin AR, Moran GF (1987) Variation in allele frequencies in the outcross pollen pool of *Eucalyptus regnans* F. Muell. throughout a flowering season. Heredity 59:161–171

Fyfe JL, Bailey NTJ (1951) Plant breeding studies in leguminous forage crops. I. Natural crossbreeding in winter beans. J Agric Sci 41:371–378

Hamrick JL, Schnabel A (1985) Understanding the genetic structure of plant populations: some old problems and a new approach. In: Gregorius H (ed) Population genetics in forestry. Springer, Göttingen, pp 50–70

Hedrick PW, Ritland K (1990) Gametic disequilibrium and multilocus estimation of selfing rates. Heredity 65:343–347

Jones DF (1916) Natural cross-pollination in the tomato. Science 43:509–510

King JN, Dancik BP, Dhir NK (1984) Genetic structure and mating system of white spruce (*Pinus glauca*) in a seed production area. Can J For Res 14:639–643

Knowles P (1985) Comparison of isozyme variation among natural stands and populations: jack pine and black spruce. Can J For Res 15:902–908

Knowles P, Furnier GR, Aleksiuk MA, Perry DJ (1987) Significant levels of self-fertilization in natural populations of tamarack. Can J Bot 65:1087–1091

Merzeau D, Di Giusto F, Comps B, Thiebaut B, Letouzey J, Cuguen J (1989) Genetic control of isozyme systems and

heterogeneity of pollen contribution in beech (*Fagus sylvatica* L.). Silvae Genet 38 (5–6):195–201
Müller G (1976) A simple method of estimating rates of self-fertilization by analyzing isozymes in tree seeds. Silvae Genet 25:15–17
Muona O (1990) Population genetics in forest tree improvement. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland/MA, pp 282–298
Perry DJ, Knowles P (1990) Evidence of high self-fertilization in natural populations of eastern white cedar (*Thuja occidentalis* L.) Can J Bot 68(3):663–668
Rao CR (1973) Linear statistical inference and its applications. Wiley and Sons, New York
Ritland K (984) The effective proportion of self-fertilization with consanguineous matings in inbred populations. Genetics 106:139–152
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
Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity 47:35–52
Roberds JH, Conkle MT (1984) Genetic structure in loblolly pine stands: allozyme variation in parents and progeny. For Sci 30:319–329
Rudolph TD (1979) Seed production in the first eight years and frequency of natural selfing in a simulated jack pine seedling seed orchard. In: Proc 13th Lake States For Tree Improv Conf, St. Paul, MN, USA, pp 33–47
Sampson JF, Hopper SD, James SH (1990) Temporal variation in allele frequencies in the pollen pool of *Eucalyptus rhodantha*. Heredity 65:189–199
Schoen DJ, Clegg MT (1984) Estimation of mating system parameters when outcrossing events are correlated. Proc Natl Acad Sci USA 81:5258–5262
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
Shea KL (1987) Effects of population structure and cone production on outcrossing rates in Engelmann spruce and subalpine fir. Evolution 41:124–136
Sittman K, Tyson H (1971) Estimates of inbreeding in *Pinus banksiana*. Can J Bot 49:1241–1245
Snyder TP, Stewart DA, Strickler AF (1985) Temporal analysis of breeding structure in jack pine (*Pinus banksiana*). Can J For Res 15:1159–1166
Sokal RR, Rohlf FJ (1981) Biometry. Freeman and Co., New York
Tobolski JJ (1979) Inheritance of polymorphic isozymes in jack pine (*Pinus banksiana* Lamb.) In: Proc 1st North Central Tree Improv Conf, Madison, Wisconsin, USA, pp 147–154
Waller DM, Knight SE (1989) Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. II. Outcrossing rates and genotypic correlations. Evolution 43(4):860–869
Yeh FC, O'Malley D (1980) Enzyme variations in natural populations of Douglas-fir, *Pseudotsuga menziessi* (Mirb.) Franco., from British Columbia. I. Genetic variation patterns in coastal populations. Silvae Genet 29:83–92
Yeh FC, Morgan K (1987) Mating system and multilocus associations in a natural population of *Pseudotsuga menziesii* (Mirb.) Franco. Theor Appl Genet 73:799–808
Zar JH (1974) Biostatistical analysis. Prentice-Hall, Englewood Cliffs/NJ