

Lecture Notes in Biomathematics

Managing Editor: S. Levin

60

Population Genetics in Forestry

Proceedings of the Meeting of the IUFRO Working Party
"Ecological and Population Genetics"
held in Göttingen, August 21–24, 1984

Edited by H.-R. Gregorius



Springer-Verlag
Berlin Heidelberg New York Tokyo

MATING SYSTEM ESTIMATION IN FOREST TREES: MODELS, METHODS AND MEANINGS

A.H.D. Brown, S.C.H. Barrett and G.F. Moran

ABSTRACT

Isozyme markers have stimulated efforts to measure the mating system in natural and planted populations of forest trees. When open-pollinated progenies have been surveyed for allozyme variation, the arrays have usually been analyzed in terms of the model of mixed self-fertilization and random outcrossing. This procedure has encountered several problems of which the more difficult are due to heterogeneity in outcrossing rates or in pollen allele frequencies. Recent models which deal with these problems and enrich the study of plant mating systems include multilocus estimation, measurement of 'effective' selfing and of differential male fertility. More reliable estimates of outcrossing, which these methods encourage, will be needed to predict selection gains, management practices and optimal conservation strategies in forest trees.

INTRODUCTION

One topic attracting increased attention among forest geneticists is the analysis of mating systems operating in various populations; be they natural stands, plantations or seed orchards. That mating systems are crucial in determining population genetic structure, and the character of genetic transmission to subsequent generations has long been recognised (e.g. as reviewed by Stern and Roche, 1974). Yet two factors have recently emerged to stimulate this interest further. These are the growth of theoretical work on the evolution of mating systems, and second, the availability of isozyme markers with which to follow mating events after the fact. This paper is concerned with the problem of measuring mating systems. The focus will be on the use of the mixed mating model, primarily in its application to temperate, hermaphroditic forest species. Our aim is to consider the problems encountered in its application, and some of the remedies devised. Finally we outline questions which estimation experiments might address and the importance of the results obtained. More general discussions of estimation are given by Ritland (1983), and citations therein.

MATING SYSTEM OF MIXED SELF-FERTILIZATION AND RANDOM OUTCROSSING

Deviation from purely random mating in plant populations has most commonly been specified by a model assuming a mixture of two types of gametic union. In this model each zygote is assumed to result from either a self-fertilization, with fixed probability s , or as fertilization with a pollen grain chosen at random from the whole population with probability $t=1-s$. This model was first applied to predominantly self-pollinated crops (Jones, 1916; Fyfe and Bailey, 1951), and later used widely in agricultural and natural populations of both inbreeding and outbreeding species including herbaceous plants and forest trees.

In such a model, estimation of the genetic consequences of the mating system is based on the joint behaviour of genotypes at one or more marker loci through the cycle of one mating event. Since codominant isozyme loci have become prevalent as genetic markers, the treatment will be based on such genes. The basic data are the segregation patterns of the frequency of progeny genotypes derived from known or inferred maternal genotypes. In the latter case, progeny size is assumed to be sufficient so that the maternal genotype is reliably known. The simplest transitions can be specified for a single diallelic locus, with alleles A_1 and A_2 as in Table 1. The observed number of progeny genotypes for each maternal genotype is coded as (O_i ; $i=1, \dots, 6$). The expectations are based on assuming the mating system of mixed selfing and random outcrossing, where the allele frequency of A_1 in the pollen is p . In this simplest (diallelic) case, the expected frequency of $A_1 A_2$ progeny from $A_1 A_2$ maternal plants is independent of the mating parameters (s, p) and is omitted here for convenience. However this class of progeny must be added to generalize the scheme to multiple alleles (Cheliak et al, 1983), multiple independent loci (Ritland and Jain, 1981), and for other mixed mating models (such as mixed outcrossing, selfing and apomixis (Marshall and Brown, 1974)).

In gymnosperms, all the progeny of heterozygous maternal plants can yield information on the mating system provided both the maternally derived haploid megagametophyte is assayed with each progeny (Shaw and Allard, 1982). Complete gametic classification is also theoretically possible for loci expressed in the triploid endosperm of angiosperms such as for seed storage protein genes. In these special cases, classes 3 and 4 in Table 1 are redefined such that O_3 includes the endosperm-progeny combinations (A_1 ; $A_1 A_1$) and (A_2 ; $A_1 A_2$) whereas O_4 includes the combinations (A_1 ; $A_1 A_2$) and (A_2 ; $A_2 A_2$). The expectations remain as defined in Table 1.

TABLE 1 Basic maternal/offspring matrix for estimation of the parameters of the mating system of mixed self-fertilization and random outcrossing.

	Genotypes		Frequencies	
	Maternal	Progeny	Observed	Expected
1.	A_1A_1	A_1A_1	O_1	$N_{11}(1-X)$
2.		A_1A_2	O_2	$N_{11}X$
3.	A_1A_2	A_1A_1	O_3	$N_{12}(1-X+Y)/2$
4.		A_2A_2	O_4	$N_{12}(1+X-Y)/2$
5.	A_2A_2	A_1A_2	O_5	$N_{22}Y$
6.		A_2A_2	O_6	$N_{22}(1-Y)$

where

$$\begin{aligned}
 X &= tq \\
 Y &= tp \\
 N_{11} &= O_1 + O_2 \\
 N_{12} &= O_3 + O_4 \\
 N_{22} &= O_5 + O_6
 \end{aligned}$$

As each maternal genotype provides one degree of freedom in the progeny array, at least two maternal types are required to estimate both mating parameters. When progenies from all three maternal genotypes are available, a test for goodness-of-fit is possible. Brown et al (1975) detail the maximum likelihood estimation procedure for this case, and Ritland (1983) provides a more general formulation.

ASSUMPTIONS OF THE BASIC MIXED MATING MODEL

The formulation of the basic model in Table 1 is designed to show its features and assumptions. The formal statistical and genetic assumptions required for estimation can be summarized as follows -

1. Within each maternal parent, the progeny genotypic classes are independent, identically distributed, multinomial random variables.
2. The values of both mating system parameters t and p (and therefore their one-to-one transformations X and Y) are uniform over maternal plants.
3. Segregation of the alleles in heterozygous maternal plants is strictly Mendelian in a 1:1 ratio for both pollen and ovule production.
4. Selection does not occur between fertilization and the assay of progeny genotypes.

These fundamental assumptions are sufficient to specify the construction of the model and the estimation of its parameters. However the assumptions themselves, or the interpretation of the estimates may require several other biological features. In particular, the model assumes inbreeding arises only through self-fertilization. This implies that the pollen involved in each outcrossing event is a random gamete sampled from the entire population. Each outcross is strictly an independent sample from the same uniform population of pollen. The genotype of each outcrossing pollen grain is independent of the maternal genotype, and of that of other outcrossing pollen grains included in the same maternal family. In a similar vein, the probability of outcrossing is assumed to be constant for all maternal plants, unaffected by maternal genotype, and independent of whether any other seed in the sample for assay is an outcross or a self.

PROBLEMS WITH SINGLE-LOCUS ESTIMATES OF OUTCROSSING RATE

As the above model and estimation procedure has been widely applied, a number of problems have come to light. Here we list and discuss several of these.

1. Estimates of outcrossing exceeding unity

When the method is applied to populations of predominantly outcrossing species, estimates of t exceeding unity have occasionally been obtained. Such so-called "biologically unreasonable" values have perturbed several workers. This problem does not arise from Wahlund effects, or necessarily from invalidity of the mixed mating model. Two primary causes are (1) sampling effects within the context of a valid mixed mating model; and (2) disassortative mating.

The first cause can be illustrated with a simple example. Suppose the true parametric value of t is unity and the population strictly adheres to the assumptions of the model. Consider a marker locus with equally frequent alleles. The expected values of X and Y are both 0.5. Consider a simple set of data based on only homozygous maternal plants (A_1A_1 and A_2A_2). On half the occasions such an assay is made, the expected result would be that O_2 would exceed $N_{11}/2$ and hence the estimate of X would exceed 0.5. Likewise on half the occasions (independent of the behaviour of O_2), the estimate of Y would exceed 0.5. Therefore in one half of such experiments the estimate of outcrossing ($t = X+Y$) would exceed unity, just by chance alone, despite the complete adherence of the population to the model.

In attaching a biological interpretation to such estimates, it seems natural to truncate the estimate of t at $t=1.0$. There are however two important dangers with this procedure. First, when a whole series of estimates based on different loci,

populations or species are to be summarized as a mean of estimates, the use of truncated values will bias the overall average estimate of outcrossing downwards. Second, if truncation at $t = 1.0$ is an automatic step of the estimation procedure, such as with the E.M. algorithm (see below), the iteration procedure may yield a biased estimate of the pollen allele frequency (p).

A clear example of the effects of disassortative mating on estimates of outcrossing is that based on the polymorphism governing style-length in tristylous, self compatible Eichhornia paniculata (Barrett and Brown, unpublished data). Open-pollinated progenies from maternal plants of known style-length were classified, and estimates of outcrossing based on the mixed mating model were found to exceed unity (Table 2). However when the progeny expectations were adjusted to allow outcrosses only from legitimate pollinations (which amounts to disassortative matings for style-length), values of t less than unity were obtained.

2. Heterogeneous estimates from different marker loci

When the same progeny arrays, and hence the same mating events, have been assayed at several marker loci, it is frequently found that the segregations of the separate loci, yield markedly different estimates of outcrossing. This was a common finding for estimates based on morphological marker loci (Harding and Tucker, 1964), and was usually attributed to selection acting differentially on the marker loci. With isozyme markers this explanation may be less attractive. A possible contributing factor to this problem may be that sampling variances of single-locus estimates based on the Cramer-Rao inequality are underestimates of these variances. Further, when the number of maternal plants is restricted, sampling variance among loci may be increased, because each marker locus may test overlapping sets of mating events and with unequal precision. Thus heterogeneity of estimates over loci may not necessarily argue that single loci estimation, or the mixed mating model is unsatisfactory.

TABLE 2. Estimates of outcrossing in tristylous Eichhornia paniculata

<u>Style Morph</u>	Mixed Mating Model	Disassortative Mating Model
Long	1.39	0.90
Mid	1.49	0.93
Short	1.18	0.84

3. Heterogeneity of outcrossing rate - temporal, spatial, population density and maternal genotype

The assumption of uniformity of outcrossing rate in a population must be regarded as unrealistic. Indeed several studies have provided evidence of heterogeneity related to the above four parameters (Hamrick, 1982). The detection of such heterogeneity is obviously important if seed harvests of the desired level of outbreeding are needed. Further, factors influencing variation in outcrossing are of prime importance in understanding the evolution of mating systems.

In reality the experimenter may have to contend with variation in outcrossing arising from more than one source. For example, in a study of a temporal sequence of three seed crops in Eucalyptus delegatensis, Moran and Brown (1980) found that outcrossing estimate was highest in the oldest crop. The difference could merely be year-to-year variation in outcrossing. Alternatively, greater loss of inbred seed in the oldest crop than in the youngest crop, might be the source of the heterogeneity. In an attempt to resolve this difference Moran and Bell (unpublished data) estimated outcrossing in eight crops from the same 33 trees of Eucalyptus stellulata collected over three years. The single-locus estimates are given in Table 3, grouped by year of mating, and show both sources of heterogeneity could be present. Temporal variation was apparent in that the outcrossing rate in 1980 was above average. As selfing was higher in 1979, there was opportunity for detecting elevated estimates of outcrossing when the seed formed in that year was allowed to remain on the tree.

TABLE 3. Temporal variation in outcrossing rate in one population of *Eucalyptus stellulata* (Moran and Bell, unpublished)

Year Fertilized	Year Assayed	Locus				Average
		<u>Aph</u>	<u>Adh</u>	<u>Pgi</u>	<u>Sdh</u>	
1981	+1	0.59	0.98	0.81	0.75	0.78
1980	+1	0.93	0.84	0.91	0.87	0.89
	+2	0.81	0.88	0.84	0.87	0.85
1979	+1	0.65	0.76	0.76	0.67	0.71
	+2	0.73	0.81	0.86	0.83	0.81
	+3	0.87	0.75	0.95	0.66	0.81
1978	+2	0.73	0.72	0.81	0.62	0.72
	+3	0.69	1.04	0.76	0.64	0.78
Total		0.75	0.84	0.84	0.75	0.80

Less attention has been paid to whether the routine estimate of outcrossing based on a model assuming uniformity, is a biased estimate of the population mean, when this assumption is false. A simple example illustrates that this could be a problem. Consider a population with only A_1A_1 and A_2A_2 maternal plants in the ratio $p:q$. If the outcrossing rates are t_1 and t_2 respectively, then the expected value for the outcrossing estimate of a model assuming uniform outcrossing is $t_1q + t_2p$. The actual frequency of outcrossed seed is $t_1p + t_2q$, so that the bias in the estimate is $(t_1 - t_2)(q - p)$. This calculation indicates that marker loci with near equal allele frequencies are to be preferred, as likely to be more robust to variation in outcrossing.

4. Heterogeneity of pollen allele frequencies among and within maternal genotypes

Another problem arises when the outcross pollen differs among maternal plants in allele frequency. This problem was soon encountered when the methods originally developed for agricultural populations were applied to natural populations. Genetic differentiation of subpopulations on a microgeographic scale, coupled with restricted pollen dispersal leads to a partial correlation between pollen genotype and maternal genotype. Estimates of outcrossing overlooking this correlation will be biased downwards (Ennos and Clegg, 1982). Alternatively, disassortative mating may be viewed as another breakdown in this assumption in which outcrossing estimates would be inflated (as in the Eichhornia example). Many of the developments to be discussed below (multilocus procedures, "effective selfing") address the problem of subpopulation structure.

It is difficult to tell whether heterogeneity arises from variation in outcrossing rate, or from variation in pollen frequencies. For example, a chi-square contingency test may be used to check for uniformity of progeny frequencies among the individual maternal plants of the same genotype. When this test detects heterogeneity, for frequency of progeny genotypes, it is not apparent whether outcrossing rates differ, or pollen allele frequencies differ (or both).

5. Multiple alleles

The occurrence of multiple alleles at a locus although providing more statistical information for mating system parameters, leads directly to numerical complications in computing the estimate. Programs and procedures are available to use the added information. Alternatively the researcher may lump the rarer alleles and reduce the dimensions of the task. No rules have been established to guide such lumping. It can be shown that for random samples from populations in inbreeding equilibrium, the pollen allele frequencies enter the expression for the theoretical variance for the maximum likelihood estimate of t , solely as powers of the product pq . Computation of the variance for a fixed value of t , and various values of p , shows that the variance is a minimum for $p=q=0.5$. Therefore lumping of alleles should proceed to yield the two-allele case with synthetic allele frequencies nearest to 0.5. However, for more complex models which invoke a greater number of parameters, multiallelic data may be needed. (Ritland, 1984.)

6. Segregation distortion

In heterozygous maternal plants, meiotic drive at micro - or mega sporogenesis or gametophytic competition thereafter could seriously distort the 1:1 expected gametic production, and hence affect genotypic frequencies in the progeny array. Yet the arrays of maternal heterozygotes generally yield limited statistical information in the estimates of the mating system parameters. Therefore biases to estimates arising from this source are likely to be minor. However the model can be modified to include the estimation of such effects if this is deemed necessary.

In angiosperms, distortion of either ovule or pollen segregation (or both) will affect the expectations in Table 1. In particular, the expected frequency of A_1A_2 progeny on A_1A_2 females will either exceed or decline from 1/2 and this may be used as a chi-square test for distortion (Clegg, 1983). In gymnosperms, when the relevant megagametophytic data are available, distortion for ovule frequencies would not affect the estimates of X and Y, as is seen in the more complex model below.

Let us assume that the ratio of $A_1:A_2$ ovules matured on A_1A_2 maternal plants is $k:1-k$, and the ratio of $A_1:A_2$ self-fertilizing pollen grains is $\ell:1-\ell$. In angiosperms, the observed and expected numbers of the three progeny genotypes are shown in Table 4A. With the extra degree of freedom (four altogether), the four parameters (k , ℓ , X and Y) can be estimated. The maximum likelihood estimators are obtained by equating observed with expected frequencies (Bailey, 1951). The four independent equations are provided by classes 2 and 5 in Table 1, and 3' and 4' in Table 4A. More restricted models are possible by substituting $\ell=1/2$ for segregation distortion only for ovules, or $k=\ell$ for equal distortion in both gametes.

For gymnosperms, there is an extra class and hence an extra degree of freedom. The expectations are given in Table 4B, and it is clear that the maximum likelihood estimator of k is

$$k = (O_{31} + O_{41}) / N$$

The M.L. estimates of ℓ , X and Y would have to be obtained by numerical procedures. As noted above, this also implies that the departure of k from 1/2 in the gymnosperm case would not affect the estimation of X and Y if it were based on the basic mixed mating model.

MORE COMPLEX PROCEDURES AND MODELS

Along with the growth in attempts at measuring outcrossing rates in plant populations have emerged several procedures to handle the problems they encountered.

TABLE 4 Expected numbers of progeny genotypes in maternal heterozygotes when segregation distortion is present.

A. <u>Angiosperms</u>			
Class	Progeny	Observed Number	Expected Number
3'	A_1A_1	O_3	$N[kls + ky]$
	A_1A_2	$N-N_{12}$	$N[(1-k)(ls+Y) + k(s(1-l)+X)]$
4'	A_2A_2	O_4	$N[(1-k)(1-l)s + (1-k)X]$
Total		N	N

B. <u>Gymnosperms</u>			
Mega-gametophyte genotype	Embryo genotype	Observed Number	Expected Number
A_1	A_1A_1	O_{31}	$N[kls + kY]$
A_1	A_1A_2	O_{41}	$N[k(1-l)s + kX]$
A_2	A_1A_2	O_{32}	$N[l(1-k)s + (1-k)Y]$
A_2	A_2A_2	O_{42}	$N[(1-k)(1-l)s + (1-k)X]$
Total		N	N

1. The E-M algorithm

Recently Cheliak et al., (1983) have used the Expectation-Maximization (EM) algorithm for obtaining the maximum likelihood estimates of mating system parameters. The major advantage of this procedure is its facility in coping with a high number of alleles at the marker loci. In principle it can be extended to handle multiple loci. As formulated, it yields estimates of outcrossing bounded strictly between 0 and 1. In any given set of data, it yields identical estimates for outcrossing with the strictly comparable maximum likelihood analysis (M.L) provided the latter is less than unity. However, in the case cited above, where M.L. gives the estimate of t exceeding 1.0; the value from the algorithm is $t=1.0$. Caution is thus needed in interpreting E.M. estimates of $t=1.0$ because the simultaneous E.M. estimates of pollen allele frequencies will be biased. The E.M. procedure does not provide values for the variance of the estimates, except in the case of a single parameter.

2. Multilocus procedures

Several of the above problems (notably 1,2,4,5 and 6) have led to the suggestion that a single estimate of outcrossing based on the joint behavior of several marker loci may be more robust than separate single-locus estimates (Shaw et al., 1981). In addition, the formulation of standard errors for the average estimate of outcrossing strictly requires a multilocus approach, rather than by combining the standard errors of single-locus estimates. This is because the scoring of an increased number of loci per individual only increases the probability of detecting an outcrossing event. The precision of the estimate of the number of such events is limited by the total number of zygotes sampled.

This leads us to the concept of detection probability. Each progeny of a maternal array can be classified as to whether it is a genetically marked outcross or not. The probability that an outcross will be detected depends on whether the pollen grain carries a non-maternal allele. When only a small number of marker loci are employed, this probability may vary considerably over maternal genotypes. Green et al. (1980) used a procedure allowing different detection probabilities for each maternal genotype in a three-locus analysis of outcrossing in Lupinus albus. The main problem with this method is that computing the detection probabilities requires making assumptions about the frequencies of pollen multilocus genotypes.

Shaw et al. (1981) developed and simplified this approach further to cope with many loci and hence multilocus maternal genotypes. Their estimator essentially assumes that the same detection probability applies to all outcrosses irrespective of maternal genotype. Although this assumption is strictly incorrect, the error introduced is negligible when the number of loci is large, so that the detection probability is high. They investigated the effect of gametic disequilibrium on their estimator. Correlation among marker loci in effect lowers the detection probability of outcrosses. If estimation proceeds ignoring the correlation then the estimates of outcrossing are biased downwards. The simplest means of avoiding this problem would be to test for the presence of disequilibrium, and use in the mating analysis, the scores of only one locus from any correlated pair. With these caveats in mind, the method of Shaw et al., although approximate, and statistically inefficient, should prove particularly useful in obtaining estimates from data on multiple loci.

Ritland and Jain (1981) present a complete M.L. estimation procedure based on many independent loci. They demonstrated numerically that the multilocus estimate is less affected by selection and non-random outcrossing than are single locus estimates. Data from a large number of loci would require considerable computer storage and iteration time. However, their analysis indicates that in most circumstances three to four loci would approach the minimum variance possible.

3. Subpopulation differentiation

A significant complicating factor in the analysis of natural populations, as distinct from agricultural populations or plantations, is the occurrence in the former of subpopulation structure. In natural plant populations, genetically alike individuals occur together because of restricted pollen dispersal or seed migration, and because of localized selection. Estimates of outcrossing based on the simple mixed mating model will usually be biased downwards. Some contrived types of subpopulation structure, such as over-dispersed plantations of genotypes can lead to upward biases in estimates (Eilstrand and Foster 1983). Such mating designs are relevant for seed orchards attempting to maximize disassortative mating.

Recently Ritland (1984) has developed a model aimed at estimating the unbiased "effective selfing" rate. The model derives estimates of selfing rates of inbred parents as opposed to outbred parents. The concept of "effective selfing" is designed to include both true selfing and crossing with related individuals. The estimation procedure includes maternal genotypic frequencies, apparently assumes equilibrium and yields higher standard errors because of the contribution of parental sampling.

4. Outcrosses sharing paternity

Schoen and Clegg (1984) have analyzed another departure from the assumptions of the mixed model. Suppose there is correlation between the pollen genotypes of the outcross progeny within a single progeny array. Such correlation can arise when the progeny come from a restricted number of pollen parents, as when a single pollen load is deposited by an insect. The determination of the maternal genotype by likelihood ratio from such an array may be biased if it assumed that outcrosses are half sibs rather than full sibs. This misclassification leads to estimates of outcrossing biased downwards. The correlation differs from that due to population structure in that the misclassification yields excessive numbers of heterozygous progeny in the apparently heterozygous maternal class. Schoen and Clegg (1984) developed a model for families with outcrosses from a single male parent for each progeny array, and the modified procedure for estimating the fraction of self pollinated seed in such cases. For tree species, it may be simpler to avoid if possible the complexities of shared paternity by collecting fruits from widely spaced branches, since these are likely to arise from separate pollinations.

5. Differential male fertility

Considerable theoretical interest has recently emerged in the extent to which plants differ in their contribution through male gametes, to the next generation (see Ross and Gregorius, 1983). Aside from the importance of such differentials to the evolution of breeding systems, is the practical significance of such variation. Thus in seed orchards, if only one clone were to provide all the successful pollen in seed production, the progeny generation would be subject to biased gene frequencies and elevated levels of inbreeding. (The progeny in this case would be a mixture of half- and full-sibs). A similar problem can arise in conservation of rare species such as Eucalyptus caesia, (Moran and Hopper, 1983), where natural populations are of very restricted size. Indeed the degree of inbreeding arising from such biases may be greater than that due to the self-fertilized component of predominantly outcrossed species.

A major issue concerns the extent to which such variation is related to genotype. Male fertility variance among individuals in natural populations might be anticipated; just as individual plants can show conspicuous variation in seed production. Presumably the genetic component for such variance is under intense selective pressure and could be rapidly exhausted unless opposed by countervailing pressures, as considered by sex allocation theory (Charnov, 1982). In natural populations, such pressures may be finely adjusted. In contrast, plantations and seed orchards may consist of individuals of disparate origin and phenology for which fertility variance may be extreme.

In relating variation to genotype three experimental situations present themselves. In populations with a finite number of distinct known genotypes, it is possible to disentangle the paternity of an array of seeds, given a sufficiently large battery of polymorphic marker loci (Ellstrand, 1984). The increased power of combined megagametophyte-seed scores in gymnosperms makes them ideal for this purpose. Genetic determination of paternity proceeds by exclusion, locus-by-locus, ideally so that all but one possible genotype is left as the potential source of pollen. Even then isolation has to be assumed. However, ambiguous paternity may be a frequent problem. Ritland (1983) formulated a likelihood approach for an estimate for any individual progeny.

The next problem is how to compute the male contributions summed over all progeny. It may be simplest to attribute ambiguous progeny fractionally to all possible male parents in proportion to their likelihoods. Such a routine would not take account of an observation that a genotype achieves a nonzero estimated contribution without ever being proved as a male parent by being involved in a unique attribution.

A second procedure, pioneered by Horovitz and Harding (1972), is to use morphological polymorphisms (such as flower colour or heterostyly) in natural or contrived populations, to test for differential male fertility, among the different morphs (Barrett *et al*, 1983).

Finally one may hope that differential male fertility might be reflected in hermaphrodites as a shift in allele frequency in the pollen as compared with parental population (Allard *et al*, 1977; Clegg, 1983). It should be noted however that relatively intense selection can occur at the diploid level without a noticeable shift in allele frequencies.

CURRENT TOPICS FOR RESEARCH

The problem of measuring mating systems with genetic markers in progeny arrays can be viewed as how to transform the frequencies of observable outcrosses (X,Y) into meaningful quantities or processes - conventionally into outcrossing rates and gamete allele frequencies. We can call these the t-effects and the p-effects. Given that a number of studies have been published, one may well ask what are the more worthwhile questions upon which future work should focus.

The t-effects

Notwithstanding the published estimates of t in forest trees, it is still a real issue as to whether outbreeders show appreciable fractions of self-pollinated progeny in nature. Alternatively, is it valid to assume that selfing is essentially zero, and any nonzero estimate arises from bias (such as Wahlund effects)? Further, what circumstances lead to higher levels of selfing?

Moran and Bell (1983) summarized the several outcrossing estimates for Eucalyptus species (Table 5). In sum, these estimates are remarkably consistent in showing that partial selfing occurs in these animal-pollinated species. The outcrossing estimates are consistently less than those found for wind-pollinated conifers.

If appreciable selfing is found to occur, the important question becomes, can we pinpoint the stages of the life-cycle at which selection against inbreds is evident (Harding, 1975). Studies of seed harvest effects (such as Table 3), of embryo competition in gymnosperms, and of genotype changes during the life cycle could uncover the effect of selection. Stern and Roche (1974) have set out a detailed scheme for the gymnosperm life cycle which is a useful focus.

Table 5. Mean estimates of the outcrossing rate (\hat{t}) in 10 Eucalyptus species (from Moran and Bell, 1983)

Species	No. of populations	No. of loci	t	Source
<u>E. obliqua</u>	4	3	0.76	Brown <u>et al.</u> (1975)
<u>E. pauciflora</u>	3	7	0.70	Phillips & Brown (1977)
<u>E. delegatensis</u>	4	3	0.79	Moran and Brown (1980)
<u>E. regnans</u>	1	4	0.69	Moran (unpubl.)
<u>E. stellulata</u>	1	3	0.77	Moran (unpubl.)
<u>E. stoatei</u>	1	3	0.82	Hopper & Moran (1981)
<u>E. kitsoniana</u>	2	3	0.77	Fripp (1982)
<u>E. citriodora</u>	1	3	0.86	Yeh <u>et al.</u> (1983)
<u>E. grandis</u>	2	6	0.84	Bell (unpubl.)
<u>E. saligna</u>	1	6	0.77	Bell (unpubl.)

Differences among individuals in outcrossing rate are an important component of variation which has received little attention. Such variation should not be confused with variance in female fecundity (or seed production).

The p-effects

One important source of p-effects is that of subpopulation structure. This may masquerade in estimates as increased levels of apparent selfing, yet it is useful to hold to a basic difference in conception between real selfing (including geitonogamy, or even crosses between plants of the same clone) versus 'effective' selfing. In general p-effects are concerned with the source of the effective non-self pollen - to what extent is it contaminant, restricted, or biased. A particularly interesting focus of research is the question to what extent is there variation among individual plants and among genotypes in populations of forest trees, in their male fertility or contribution to the outcrossing pollen? At seed harvest, the female contributions to a seed bulk can be readily controlled - in marked contrast to the lack of control over male fertility variance. Further, in

seed orchards such variance could be more noticeable than in natural populations because planted populations may consist of selected clones of disparate origin, unlikely to be in equilibrium. As already discussed, the careful analysis of p-effects requires the construction of more complex, less general models, incorporating more features of the reproductive biology of the species.

The t-effects as opposed to the p-effects

Further we need to know the relative importance of variation in t-effects and p-effects, and their degree of interrelationship. What is the variance in selfing, compared with the variance in male fertility? Do the individuals or genotypes with higher rates of selfing have lower male fertility?

Significance of mating system studies

We are seeing a growth in studies of plant mating systems, particularly in forest trees. Studies based on markers may be descriptive (aimed at detecting what has happened in one mating cycle), or manipulative (testing what can happen in contrived populations). The overall aim is to understand the dynamic processes which result in the progeny generation, and to specify this in parameters attributable to the parental generations. The practical significance of such studies cannot be overemphasized in forest genetics - where the long term cost of producing poor seedling generations is substantial. The information is fundamental to the management of seed orchards so that the hard-won gains of selection will not be lost on inferior seed quality.

There are also important implications for the management of natural stands. Selective logging has been mainly debated in terms of dysgenic selection. Yet the effects on variation in mating system induced by selective logging may also be significant. Thus when outcrossing rates are density dependent, isolated individuals left after logging may be highly self-fertilized.

Third, an appreciation of the mating system is needed for defining optimal conservation strategies. We need to know in more quantitative terms, what are effective population sizes, neighborhood sizes, and migration rates, to answer questions about strategies (Brown and Moran, 1980). Only when these parameters are known with greater precision can a truly genetic management of biological resources be undertaken.

REFERENCES

- Allard, R.W., Kahler, A.L., and Clegg, M.T. 1977. Estimation of mating cycle components of selection in plants. In: Measuring Selection in Natural Populations. (ed. Christiansen, F.B. and Fenchel, T.M.) pp. 1-19, Springer-Verlag, Berlin.
- Bailey, N.T.J. 1951. Testing the solubility of maximum likelihood equations in the routine application of scoring methods. Biometrics 7: 268-274.
- Barrett, S.C.H., Price, S.D. and Shore, J.S. 1983. Male fertility and anisoplethic population structure in tristylous Pontederia cordata. Evolution 37:745-759.
- Brown, A.H.D., Matheson, A.C., and Eldridge, K.G. 1975. Estimation of the mating system of Eucalyptus obliqua L'Herit. using allozyme polymorphisms. Aust. J. Bot. 23: 931-949.
- Brown, A.H.D., and Moran, G.F. 1980. Isozymes and the genetic resources of forest trees. In: Isozymes of North American Forest Trees and Forest Insects. (ed. Conkle, M.T.) pp. 1-10, U.S.D.A.
- Charnov, E.L. 1982. The Theory of Sex Allocation. Princeton Univ. Press., N.J.
- Cheliak, W.M., Morgan, K., Strobeck, C., Yeh, F.C.H., and Danick, B.P. 1983. Estimation of mating system parameters in plant populations using the EM algorithm. Theor. Appl. Genet. 65: 157-161.
- Clegg, M.T. (1983). Detection and measurement of natural selection. In: Isozymes in Plant Genetics and Breeding, Part A. (ed. Tanksley, S.D. and Orton, T.J.) pp. 241-255. Elsevier, Amsterdam.
- Ellstrand, N.C. 1984. Multiple paternity within the fruits of the wild radish Raphanus sativus. Am. Nat. 123: 819-828.
- Ellstrand, N.C., and Foster, K.W. 1983. Impact of population structure on the apparent outcrossing rate of grain sorghum (Sorghum bicolor). Theor. Appl. Genet. 66: 323-327.
- Ennos, R.A., and Clegg, M.T. 1982. Effect of population substructuring on estimates of outcrossing rate in plant populations. Heredity 48: 283-292.
- Fripp, Y.J. 1982. Allozyme variation and mating system in two populations of Eucalyptus kitsoniana (Leuhm) Maiden. Aust. For. Res. 13: 1-10.
- Fyfe, J.L., and Bailey, N.T.J. 1951. Plant breeding studies in leguminous forage crops. 1. Natural cross-breeding in winter beans. J. Agric. Sci. 41: 371-378.
- Green, A.G., Brown, A.H.D., and Oram, R.N. 1980. Determination of outcrossing in a breeding population of Lupinus albus L. Z. Pflanzenzucht. 84: 181-191.
- Hamrick, J.L. 1982. Plant population genetics and evolution. Amer. J. Bot. 69: 1685-1693.
- Harding, J. and Tucker, C.L. 1964. Quantitative studies on mating systems. 1. Evidence for the non-randomness of outcrossing in Phaseolus lunatus. Heredity 19: 369-381.
- Harding, J. 1975. Models for gametic competition and self fertilization as components of natural selection in populations of higher plants. In: Gametic Competition in Plants and Animals. (ed. Mulcahy, D.L.) pp. 243-255. North Holland Publ. Co., Amsterdam.
- Hopper, S.D., and Moran, G.F. 1981. Bird pollination and the mating system of Eucalyptus stoatei. Aust. J. Bot. 29: 625-638.
- Horovitz, A., and Harding, J. 1972. The concept of male outcrossing in hermaphrodite higher plants. Heredity 29: 223-236.

- Jones, D.F. 1916. Natural cross-pollination in the tomatoe.
Science 43: 509-510.
- Marshall, D.R., and Brown, A.H.D. 1974. Estimation of the level of apomixis in plant populations.
Heredity 32: 321-333.
- Moran, G.F., and Bell, J.C. 1983. Eucalyptus. In: Isozymes in Plant Genetics and Breeding, Part B (ed. Tanksley, S.D. and Orton, T.J.) pp. 423-441 Elsevier, Amsterdam.
- Moran, G.F. and Brown, A.H.D. 1980. Temporal heterogeneity of outcrossing rates in alpine ash (Eucalyptus delegatensis R.T. Bak.).
Theor. Appl. Genet. 57: 101-105.
- Moran, G.F. and Hopper, S.D. 1983. Genetic diversity and the insular population structure of the rare granite rock species, Eucalyptus caesia Benth.
Aust. J. Bot. 31:161-172.
- Phillips, M.A., and Brown, A.H.D. 1977. Mating system and hybridity in Eucalyptus pauciflora.
Aust. J. Biol. Sci. 30: 337-344.
- Ritland, K. 1983. Estimation of mating systems. In: Isozymes in Plant Genetics and Breeding, Part A. (ed. Tanksley, S.D. and Orton, T.J.) pp. 289-302. Elsevier, Amsterdam.
- Ritland, K. 1984. The effective proportion of self-fertilization with consanguineous matings in inbred populations.
Genetics 106: 139-152.
- Ritland, K., and Jain, S.K. 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci.
Heredity 47: 35-52.
- Ross, M.D. and Gregorius, H.-R. 1983. Outcrossing and sex function in hermaphrodites: a resource-allocation model.
Am. Nat. 121: 204-222.
- Schoen, D.J. and Clegg, M.T. 1984. Estimation of mating system parameters when outcrossing events are correlated.
Proc. Natl. Acad. Sci. U.S. 81:5258-5262.
- Shaw, D.V., and Allard, R.W. 1982. Estimation of outcrossing rates in Douglas-fir using isozyme markers.
Theor. Appl. Genet. 62: 113-120.
- Shaw, D.V., Kahler, A.L., and Allard, R.W. 1981. A multilocus estimator of mating system parameters in plant populations.
Proc. Natl. Acad. Sci. U.S.A. 78: 1298-1302.
- Stern, K., and Roche, L. 1974. Genetics of Forest Ecosystems. Springer-Verlag, Berlin.
- Yeh, F.C., Brune, A., Cheliak, W.M., and Chipman, D.C. 1983. Mating system of Eucalyptus citriodora in a seed production area.
Can. J. For. Res. 13:1051-1055.