Theor Appl Genet (1983) 65: 157–161

Estimation of mating system parameters in plant populations using the EM algorithm

W. M. Cheliak
Petawawa National Forestry Institute, Chalk River, Ontario KOJ 1JO, Canada

K. Morgan, C. Strobeck
Department of Genetics, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

F. C. H. Yeh[1]
Research Branch, British Columbia, Ministry of Forests, Victoria, British Columbia V8W 3E7, Canada

B. P. Dancik
[1] Department of Forest Science, University of Alberta, Edmonton, Alberta T6G 2H1, Canada

Received December 27, 1982
Communicated by R. W. Allard

Summary. An EM algorithm procedure is presented for the maximum-likelihood estimation of mating system parameters of mixed mating system models for both angiosperms and gymnosperms. One advantage of the procedure is the ability to accommodate any number of alleles in the mature population and pollen pool. Estimates of the outcrossing rate ($\hat{t}$) derived from the model are bounded strictly within the natural biological range (i.e., $0 \leqq \hat{t} \leqq 1$).

Key words: Maximum-likelihood estimation – Mating systems – Incomplete data – EM algorithm

Introduction

The mating system of a species is important in determining the amount, distribution and maintenance of genetic variation in the population from generation to generation (Allard et al. 1975). Numerous single locus models have been developed for estimating mating system parameters in plant populations (summarized by Jain 1979). This is related primarily to the great diversity of mating systems in higher plants (Fryxell 1957). Recently, multilocus estimators also have been developed (Brown et al. 1978; Green et al. 1980; Ritland and Jain 1981; Shaw et al. 1981). However, these estimators are formulated specifically for angiosperms and generally for diallelic loci. Shaw and Allard (1982) present maximum-likelihood estimation procedures for gymnosperms, again for a diallelic locus. In this paper, we present generalized equations to estimate system parameters using an Expectation-Maximization (EM) algorithm (Dempster et al. 1977) for both angiosperm and gymnosperm species. We believe that application of the EM algorithm has distinct advantages over other published models for estimation of parameters for a mixed mating system model. These advantages of the model and estimation procedure include:

1. An explicit expression is given to determine the proportion of selfed and outcrossed embryos in phenotypically confounded classes (i.e. classes of observed embryos which contain both selfed and outcrossed embryos).
2. The estimators derived from the EM algorithm are intuitively understandable and give simultaneous maximum-likelihood estimates of the frequency of alleles in the outcrossed pollen pool and the outcrossing rate for an arbitrary number of genotypes in the mature population.
3. The method is easy to program relative to conventional maximum-likelihood methods.

The gene counting methods first published by Ceppellini et al. (1955) and extensions by Smith (1957) are specialized examples of an EM algorithm as applied to problems in genetics.

Differences between reproduction in gymnosperms and angiosperms

In gymnosperms, double fertilization does not occur and the female gametophyte (megagametophyte) remains in a haploid condition. Furthermore, all cells within the seed coat including the egg-cells (ovule), have been derived mitotically from a single megaspore. Thus, barring mutation, the maternal gamete contributed to that embryo can be unambiguously ascertained by co-electrophoresing haploid megagametophytic tissue and the corresponding embryo. This haploid megagametophytic tissue also allows direct probabilistic inference of the maternal genotype (Morris and Speith 1978). Inference of maternal genotype by this method is independent of the mating system (i.e., selfing rate and pollen allele frequencies) if all embryos have equal probability of surviving, and there is no segregation distortion.

Since the endosperm in angiosperm reproductive systems is triploid (2n ♀, 1n ♂) this tissue cannot be used exclusively to infer the maternal genotype. With angiosperm species, it is necessary to infer maternal genotype on the basis of open-pollinated progeny arrays from single plants. The reader is referred to the papers by Brown and Allard (1970); Brown et al. (1975) and Clegg et al. (1978) for details on the methods.

The following data are necessary to estimate mating system parameters using the EM algorithm:

1. A known or inferred maternal genotype. These data can be obtained from either a pedigree or by inference as discussed above.
2. Genotypes of open-pollinated progeny arrays from known or inferred maternal genotypes.
3. At least two different genotypic classes of maternal genotypes.
4. For the gymnosperm formulation only, known maternal contribution to heterozygous embryos from heterozygous maternal plants of the same genotype.

Throughout the development of this estimation procedure it is assumed that the frequency of alleles in the pollen pool is homogeneous over the array of mature plants sampled. It is also assumed that the probability of outcrossing is constant among the array of mature plants sampled. This is equivalent to assuming a constant amount of selfing (s) and random outcrossing $t(=1-s)$ and further that all embryos, regardless of mating event, have equal fitness (Brown and Allard 1970).

For the angiosperm formulation, it is necessary to assume equality of segregation for the two alleles in heterozygous maternal plants. This assumption can be relaxed in the gymnosperm formulation. While it is not usually possible to distinguish between ij versus ji heterozygotes in the mature population, if the distinction can be made, this procedure can accommodate the data. However, the data observed will most often consist of the diagonal homozogotes (ii) and upper trimat of heterozygotes (ij). Thus, in practice we treat ij heterozygotes the same as ji heterozygotes.

Estimating mating system parameters using the EM algorithm

Dempster et al. (1977) have presented a method to derive maximum-likelihood estimates of distributions of incomplete data. They term this iterative process an EM algorithm, since each cycle consists of an expectation step (E) followed by a maximization step (M). Incomplete data in studies of mating systems occur, for example, when an observed class of genotypes contains the results of different genetic events, i.e., different origins of the paternal allele, that are not otherwise distinguishable.

Notation

Let the number of embryos observed in a confounded class from $A_i^*A_i^*$ maternal plants be ${}_{ii}y_{ii}$. The asterisks are used to follow maternal gametes. The subscripts on the left denote the maternal genotype and the subscripts on the right denote the embryo genotype. Let the number of selfed embryos be represented by ${}_{ii}x_{ii}$ and the number of outcrossed embryos be represented by ${}_{ii}x_i^j$. The left-hand subscripts are as previously defined. The inner right-hand subscript denotes the maternal gamete. The superscript or outer right-hand subscript denotes the paternal contribution for outcrossed and selfed embryos, respectively (Table 1). The complete sample space of embryos (X) cannot be observed directly, but can be obtained from the incomplete sample space of observations (Y) by a many-to-one mapping from X to Y.

The expectation step (E)

Our task is to derive the expected number of selfed and outcrossed embryos in each confounded class. For

Table 1. Observed progeny phenotypes and underlying genotypes by origin of mating event

Maternal genotype	Progeny genotype	Incomplete space (obseved no.)	Probability	Complete space (expected no.)	
$A_i^*A_i^*$	$A_i^*A_i^*$	${}_{ii}y_{ii}$	s	${}_{ii}x_{ii}$	
	$A_i^*A_i$		$(1-s)p_i$	${}_{ii}x_i$	
	$A_i^*A_j$	${}_{ii}y_{ij}$	$(1-s)p_j$	${}_{ii}x_i^j$	$j \neq i$
$A_i^*A_j^*$	$A_i^*A_i^*$	${}_{ij}y_{ii}$	$s/4$	${}_{ij}x_{ii}$	
$j \neq i$	$A_i^*A_i$		$(1-s)p_i/2$	${}_{ij}x_i^i$	
	$A_i^*A_j^*$	${}_{ij}y_{ij}$	$s/4$	${}_{ij}x_{ij}$	
	$A_i^*A_j$		$(1-s)p_j/2$	${}_{ij}x_i^j$	
	$A_j^*A_i^*$	${}_{ij}y_{ji}$	$s/4$	${}_{ij}x_{ji}$	
	$A_j^*A_i$		$(1-s)p_i/2$	${}_{ij}x_j^i$	
	$A_j^*A_j^*$	${}_{ij}y_{jj}$	$s/4$	${}_{ij}x_{jj}$	
	$A_j^*A_j$		$(1-s)p_j/2$	${}_{ij}x_j^j$	
	$A_i^*A_l$	${}_{ij}y_{il}$	$(1-s)p_l/2$	${}_{ij}x_i^l$	$l \neq i, j$
	$A_j^*A_l$	${}_{ij}y_{jl}$	$(1-s)p_l/2$	${}_{ij}x_j^l$	$l \neq i, j$

s = rate of self-fertilization, p_i = frequency of the ith allele in the outcrossed pollen pool homozygous maternal plants this is given by equation [1-a]. For heterozygous maternal plants the appropriate equations are in [1-b].

$$\begin{aligned}
{}_{ii}x_{ii} &= {}_{ii}y_{ii} \quad \{s/[s+p_i(1-s)]\} \\
{}_{ii}x_{i}^{i} &= {}_{ii}y_{ii} \quad \{p_i(1-s)/[s+p_i(1-s)]\} \\
{}_{ii}x_{i}^{j} &= {}_{ii}y_{ij} \qquad\qquad j \neq i \qquad (1\text{-a})
\end{aligned}$$

$$\begin{aligned}
{}_{ij}x_{ii} &= {}_{ij}y_{ii} \quad \{s/[s+2p_i(1-s)]\} \\
{}_{ij}x_{i}^{i} &= {}_{ij}y_{ii} \quad \{2p_i(1-s)/[s+2p_i(1-s)]\} \quad j \neq i \\
{}_{ij}x_{ij} &= {}_{ij}y_{ij} \quad \{s/[s+2p_j(1-s)]\} \\
{}_{ij}x_{i}^{j} &= {}_{ij}y_{ij} \quad \{2p_j(1-s)/[s+2p_j(1-s)]\} \quad j \neq i \\
{}_{ij}x_{ji} &= {}_{ij}y_{ji} \quad \{s/[s+2p_i(1-s)]\} \\
{}_{ij}x_{j}^{i} &= {}_{ij}y_{ji} \quad \{2p_i(1-s)/[s+2p_i(1-s)]\} \quad j \neq i \\
{}_{ij}x_{jj} &= {}_{ij}y_{jj} \quad \{s/[s+2p_j(1-s)]\} \\
{}_{ij}x_{j}^{j} &= {}_{ij}y_{jj} \quad \{2p_j(1-s)/[s+2p_j(1-s)]\} \quad j \neq i \\
{}_{ij}x_{i}^{l} &= {}_{ij}y_{il} \\
{}_{ij}x_{j}^{l} &= {}_{ij}y_{jl} \qquad\qquad l \neq j \neq i \quad (1\text{-b})
\end{aligned}$$

Throughout (1-a) and (1-b), the p's are the frequencies of alleles in the outcrossed pollen pool and s is the selfing rate.

If a heterozygous embryo, say ${}_{ij}y_{ij}$, has the same genotype as the maternal plant, it is not possible to determine whether the embryo received a copy of the i or the j maternal allele except by direct observation. In this case, it is necessary to divide the total number of ${}_{ij}y_{ij}$ embryos into three classes; a selfed class (${}_{ij}x_{++} = {}_{ij}x_{ij} + {}_{ij}x_{ji}$) and two outcrossed classes (${}_{ij}x_{i}^{j}$ and ${}_{ij}x_{j}^{i}$). In doing so, we must assume equal segregation of the i and j alleles in the female gametophytes of heterozygous maternal plants. This changes the expectation for these classes only to:

$$\begin{aligned}
{}_{ij}x_{++} &= {}_{ij}y_{ij} \quad \{s/[s+(1-s)(p_i+p_j)]\} \\
{}_{ij}x_{i}^{j} &= {}_{ij}y_{ij} \quad \{p_j(1-s)/[s+(1-s)(p_i+p_j)]\} \\
{}_{ij}x_{j}^{i} &= {}_{ij}y_{ij} \quad \{p_i(1-s)/[s+(1-s)(p_i+p_j)]\}\,.
\end{aligned}$$

Furthermore, if there are only two alleles at this locus in the population (i.e., $p_i + p_j = 1$) the ${}_{ij}y_{ij}$ embryos are apportioned to the selfed and outcrossed classes according to the current estimate of the selfing rate (s) and outcrossing rate ($t = 1 - s$). That is, this class contains no information about the parametric values of p_i and s, as the partial differential of the log-likelihood equation with respect to these parameters for heterozygous plants is zero (Brown et al. 1975).

The maximization step (M)

Maximum-likelihood estimators of p_i and s can now be derived simultaneously from the complete space (X) by partial differentiation of the log-likelihood equation and setting to zero. The derivation for r alleles in the pollen pool, n alleles in the mature population for the general case (gymnosperms) is given in Appendix. The maximum-likelihood estimators for the frequency of p_i in outcrossed pollen pool and the selfing rate (s) are (from Appendix):

$$\hat{p}_i = \frac{{}_{..}x_{.}^{i}}{{}_{..}x_{.}^{.}} \tag{2}$$

and

$$\hat{s} = \frac{{}_{..}x_{..}}{{}_{..}x_{.}^{.} + {}_{..}x_{..}} \tag{3}$$

where the . indicates summation over all alleles.

Maximum-likelihood estimates of the frequency of each of the pollen alleles and the selfing rate are obtained by cycling between equations (1-a), (1-b) and (2), (3) until successive estimates of the parameters converge to a specified criterion. Thus one cycle of the EM algorithm consists of estimating the number of selfed and outcrossed embryos in a confounded class at step v, and obtaining the maximum-likelihood estimates of the parameters based on the provisional estimates of the complete sample space at step v + 1. Initially, estimates or, arbitrary values of the p's and s are provided to begin the iterative procedures. The robustness of the algorithm to initial values of p_i and s is discussed in the next section.

The estimators of the EM-algorithm procedure are easy to understand. The maximum-likelihood estimate of the outcrossed pollen-pool frequency of allele i is the sum of the numbers of outcrossed embryos containing pollen allele i (${}_{..}x_{.}^{i}$) divided by the total number of outcrossed embryos (${}_{..}x_{.}^{.}$). Similarly, the maximum-likelihood estimate of the selfing rate (s) is the sum of the numbers of selfed embryos (${}_{..}x_{..}$) divided by the total number of embryos (${}_{..}x_{..} + {}_{..}x_{.}^{.}$). Estimates of the numbers in these various classes of embryos are obtained from the expectation step.

Analysis of data using the EM algorithm

We have reanalyzed the electrophoretic data for a breeding population of *Lupinus albus* L. presented by Green et al. (1980). With the exception of the pollen allele frequency of locus Pgm-1, the EM algorithm gives results identical to those published (Table 2). We have also re-analyzed the data for the Pgm-1 locus using the method of Brown et al. (1975). These results are in complete agreement with those obtained by the EM-algorithm procedure.

A large number of initial values of the set {p, s} for the Pgm-1 locus ranging from {0.00001, 0.99999} to

Table 2. Comparison of estimates of mating system parameters in *Lupinus albus* L. (Source of data: Green et al. 1980)

Parameter[a]	Locus		
	Pgm-1	6Pgd-2	Aat-2
$\hat{t}_G$	0.12	0.09	0.09
$\hat{t}_{EM}$	0.12	0.09	0.09
$\hat{p}_G$	0.62	0.62	0.15
$\hat{p}_{EM}$	0.64	0.62	0.15

$\hat{t}$ is the maximum-likelihood estimate of the outcrossing rate, $\hat{p}$ is the maximum likelihood estimate of the pollen allele frequency in the outcrossed pollen pool; G stands for estimates from Green et al. (1980) and EM stands for estimates obtained by the EM-algorithm procedure of this paper

{0.99999, 0.00001}, respectively were used to test convergence of the iterative procedure to a stationary value or local or global maximum (Wu 1981). Regardless of the initial set, the procedure always converged to $\{\hat{p}, \hat{s}\} = \{0.64449, 0.11787\}$ for the estimate set. These results suggest that the present procedure is robust with respect to initial starting values, and that the values of $\hat{p}$ and $\hat{s}$ are likely to be the global maximum-likelihood estimates.

Extension of the procedure to a multiple allelic locus is straightforward and can be illustrated by an analysis of the data for the Adh locus of *Eucalyptus citriodora* Hook (Yeh et al. 1982). The results are compared to those obtained by the method of Brown et al. (1975). A total of four alleles was observed at the Adh locus in the mature and filial generations with six inferred maternal genotypes and up to seven different filial genotypes per maternal genotype (Yeh et al. 1982). The two procedures give different estimates for the frequency of the pollen allele 100 and the outcrossing rate (Table 3). The present method utilizes all of the information in the Adh data. Reducing the dimensionality of the data by collapsing to a diallelic system (Brown et al. 1975) might well have resulted in less efficient estimates of mating system parameters. The effect of such statistical inefficiency on the parameter estimates will depend on the number and frequency distribution of the alleles at a particular locus.

Table 3. Estimation of mating system parameters for the Adh locus of *E. citriodora* using the EM-algorithm procedure and method of Brown et al. (1975)

Allele	Frequency			
	Mature	Filial	Pollen (EM)	Pollen (B)
1 (90)	0.09	0.11	0.19	–
2 (100)	0.57	0.60	0.61	0.66
3 (108)	0.33	0.28	0.19	–
4 (95)	0.01	0.01	0.01	–
$\hat{t}$[a]			0.86	0.80

[a] $\hat{t}$ is the maximum-likelihood estimate of the outcrossing rate and EM and B refer to estimates from the EM-algorithm procedure and the method of Brown et al. (1975)

Conclusions

We have developed an iterative procedure for the maximum-likelihood estimation of mating system parameters for a mixed mating system model for both angiosperms and gymnosperms. This procedure is an application of the EM-algorithm. For diallelic loci with equal segregation of gametes in heterozygous maternal plants the EM-algorithm procedure and conventional maximum-likelihood estimators give identical results. However, the real power of the EM-algorithm procedure is the ability to estimate mating system parameters for multi-allelic loci in either the mature population or outcrossed pollen pool or both. Furthermore, the estimates of the mating system parameters obtained by this procedure may approach closer to the global maximum-likelihood estimates with each iteration regardless of initial starting values.

Acknowledgements. The authors, especially K.Morgan, would like to thank Dr. Joseph Felsenstein for drawing our attention to the EM algorithm. We would also like to thank Dr. A. H. D. Brown, Dr. B. S. Weir and two anonymous reviewers for suggestions on an earlier version of this paper. This work was supported by NSERC operating grants NSERC A7756 to K.M., NSERC A0502 to C.S. and NSERC A8348 to F.C.Y. Financial assistance to W.M.C. and B.P.D. by the Canadian Forestry Service and to B.P.D. by the Alberta Forest Service is gratefully acknowledged.

Appendix

Equations (1-a) and (1-b) provide estimates of the expected numbers of outcrossed and selfed embryos from homozygous and heterozygous maternal plants. Using this information and the probabilities of observing that sample from column 4 of Table 1, the likelihood equation for the complete sample space is

$$\begin{aligned} L \propto \prod_{i=1}^{r} \Big\{ (s)^{{}_{ii}x_{ii}} [(1-s)\,p_i]^{{}_{ii}x_i^i} \prod_{j \neq i} [(1-s)\,p_j]^{{}_{ii}x_i^j} \Big\} \\ \cdot \prod^{r} \prod^{r}_{i \neq j} \Big\{ (s/4)^{{}_{ij}x_{ii}} [(1-s)\,p_i/2]^{{}_{ij}x_i^i} (s/4)^{{}_{ij}x_{jj}} [(1-s)\,p_j/2]^{{}_{ij}x_j^j} \\ \cdot (s/2)^{{}_{ij}x_{ij}} [(1-s)\,p_j/2]^{{}_{ij}x_i^j} [(1-s)\,p_i/2]^{{}_{ij}x_j^i} \\ \cdot \prod_{l \neq i,j} [(1-s)\,p_l/2]^{{}_{ij}x_i^l} [(1-s)\,p_l/2]^{{}_{ij}x_j^l} \Big\} \end{aligned}$$

where p_i, $i = 1, \ldots, r$ is the frequency of the i^{th} allele in the outcrossed pollen pool, s is the selfing rate and x is an expected observation.

The log-likelihood function is given by:

$$\ln L = L^* = \text{const.} + \{\sum_i [{}_{ii}x_{ii} + \sum_{j \neq i} ({}_{ij}x_{ii} + {}_{ij}x_{jj} + {}_{ij}x_{ij})]\} \ln(s)$$

$$+ \{\sum_i [{}_{ii}x^i_i + \sum_{j \neq i} ({}_{ii}x^j_i + {}_{ij}x^i_i + {}_{ij}x^j_j + {}_{ij}x^j_i + {}_{ij}x^i_j)$$

$$+ \sum_{j \neq l \neq i} \sum ({}_{ij}x^l_i + {}_{ij}x^l_j)]\} \ln(1-s)$$

$$+ \{\sum_i [{}_{ii}x^i_i + \sum_{j \neq i} ({}_{ij}x^i_i + {}_{ij}x^i_j)] \ln p_i\}$$

$$+ \{\sum_{i \neq j} \sum [({}_{ii}x^j_i + {}_{ij}x^j_j + {}_{ij}x^j_i) \ln p_j]\}$$

$$+ \{\sum_{i \neq j \neq l} \sum \sum [({}_{ij}x^l_i + {}_{ij}x^l_j)] \ln p_l\}$$

$$= \text{const.} + {}_{..}x_{..} \ln(s) + {}_{..}x^{\cdot}_{\cdot} \ln(1-s)$$

$$+ \{\sum_i [{}_{ii}x^i_i + \sum_{j \neq i} ({}_{jj}x^i_j + {}_{ij}x^i_i + {}_{ij}x^i_j + {}_{ji}x^i_i + {}_{ji}x^i_j)$$

$$+ \sum_{j \neq l \neq i} \sum ({}_{jl}x^i_j + {}_{jl}x^i_l] \ln p_i\}$$

$$= \text{const.} + {}_{..}x_{..} \ln(s) + {}_{..}x^{\cdot}_{\cdot} \ln(1-s) + \sum_i {}_{..}x^i_{\cdot} \ln p_i$$

where

$${}_{..}x^i_{\cdot} = {}_{ii}x^i_i + \sum_{j \neq i} ({}_{ij}x^i_i + {}_{ij}x^i_j + {}_{ji}x^i_i + {}_{ji}x^i_j)$$

$$+ \sum_{j \neq i} [{}_{jj}x^i_j + \sum_{l \neq i,j} ({}_{jl}x^i_j + {}_{jl}x^i_l)]$$

is the number of outcrossed embryos containing pollen allele i,

$${}_{..}x^{\cdot}_{\cdot} = \sum_i [{}_{ii}x^i_i + \sum_{j \neq i} ({}_{ii}x^j_i + {}_{ij}x^i_i + {}_{ij}x^j_j + {}_{ij}x^j_i + {}_{ij}x^i_j)$$

$$+ \sum_{j \neq l \neq i} \sum ({}_{ij}x^l_i + {}_{ij}x^l_j)]$$

$$= \sum_i \{{}_{ii}x^i_i + \sum_{j \neq i} ({}_{ij}x^i_i + {}_{ij}x^i_j + {}_{ji}x^i_i + {}_{ji}x^i_j)$$

$$+ \sum_{j \neq i} [{}_{jj}x^i_j + \sum_{l \neq i,j} ({}_{jl}x^i_j + {}_{jl}x^i_l)]\} = \sum_i {}_{..}x^i_{\cdot}$$

is the number of all outcrossed embryos and

$${}_{..}x_{..} = \sum_i [{}_{ii}x_{ii} + \sum_{j \neq i} ({}_{ij}x_{ii} + {}_{ij}x_{jj} + {}_{ij}x_{ij})]$$

is the number of all selfed embryos.

To constrain the allele frequencies to sum to 1, we wish to maximize L^* subject to the contraint.

$$\sum_{i=1}^{r} p_i = 1$$

by the method of Lagrangian multipliers. The log-likelihood function with the constraint (L^+) is given by

$$L^+ = L^* + \lambda(1 - \sum_i p_i)\,.$$

Maximum likelihoood estimates of p_i and s are obtained by the solution of the following equations;

$$0 = \frac{\partial L^+}{\partial \lambda} = 1 - \sum_i p_i$$

$$0 = \frac{\partial L^+}{\partial s} = {}_{..}x_{..}/s - {}_{..}x^{\cdot}_{\cdot}/(1-s)$$

$$0 = \frac{\partial L^+}{\partial p_i} = {}_{..}x^i_{\cdot}/p_i - \lambda\,.$$

Thus, the maximum likelihood estimators of λ, p_i and s are

$\hat{\lambda} = {}_{..}x^{\cdot}_{\cdot}$

$\hat{p}_i = {}_{..}x^i_{\cdot}/{}_{..}x^{\cdot}_{\cdot}$

$\hat{s} = {}_{..}x_{..}/{}_{..}x$

where ${}_{..}x = {}_{..}x_{..} + {}_{..}x^{\cdot}_{\cdot}$ is the total number of all embryos.

References

Allard RW, Kahler AL, Clegg MT (1975) Isozymes in plant population genetics. In: Market CL (ed) Isozymes IV. Academic Press, New York, pp 261–272

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 the mating system in *Eucalyptus obliqua* L'Herit using allozyme polymorphisms. Aust J Bot 23:931–949

Brown AHD, Zohary D, Nevo E (1978) Outcrossing rates and heterozygosity in natural populations *Hordeum spontaneum* Koch. in Israel. Heredity 41:49–62

Ceppellini R, Siniscalco R, Smith CAB (1955) The estimation of gene frequencies in a random mating population. Ann Hum Genet 20:97–115

Clegg MT, Kahler AL, Allard RW (1978) Estimation of life cycle components of selection in an experimental plant population. Genetics 89:765–792

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

Fryxell PA (1957) Mode of reproduction in higher plants. Bot Rev 23:135–233

Green AG, Brown AHD, Oram RN (1980) Determination of outcrossing rate in a breeding population of *Lupinus albus* L. (white lupin). Z Pflanzenzücht 84:181–191

Jain SK (1979) Estimation of outcrossing rates: Some alternative procedures. Crop Sci 19:23–26

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

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

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

Wu C-H (1981) On the convergence of the EM algorithm. Tech Rep No 642. University of Wisconsin, Dept of Statistics

Yeh FC, Brune A, Cheliak WM, Chipman DC (1982) Mating system of *Eucalyptus citriodora* in a seed orchard. Can J For Res (submitted)

A program listing, documentation, and test data may be obtained from the senior author on request.