THE EFFECTIVE PROPORTION OF SELF-FERTILIZATION WITH CONSANGUINEOUS MATINGS IN INBRED POPULATIONS

KERMIT RITLAND[1]

Department of Botany, University of British Columbia, Vancouver, Canada V6T 2B1

ABSTRACT

Allelic segregation at a single locus among offspring derived from matings, including those between inbred relatives, is a combination of two patterns, corresponding to self-fertilization and random outcrossing. The proportion of effective self-fertilization is termed the "effective selfing rate," and it is specified with identity coefficients. The description of the offspring genotypic distribution for a population with mating among relatives requires a set of three independent parameters of genetic and mating structure. One such set is the inbreeding coefficient of parents, the coefficient of kinship between mates and the effective selfing rate. The model used to derive the effective selfing rate distinguishes between the effective selfing rates of inbred *vs.* outbred parents; the mixed mating model does not distinguish between these two rates. As a result, the mixed mating model usually gives biased estimates of effective selfing, if there is mating among inbred relatives. The procedure for estimation of effective selfing, based upon progeny array data distributed according to the "effective selfing model," is presented, and an example is given.

MATINGS between relatives might often contribute significantly to the level of inbreeding in natural populations. Studies of mating systems in plants have primarily assumed mixed self-fertilization and random outcrossing, the "mixed mating" model of HAYMAN (1953). Numerous empirical and theoretical studies of inbreeding (ALLARD, JAIN and WORKMAN 1968; WEIR and COCKERHAM 1973; JAIN 1976), as well as procedures for estimating the degree of random outcrossing in the absence of selection (BROWN and ALLARD 1970; CLEGG, KAHLER and ALLARD 1978), are based on this assumption. However, some studies have found evidence for consanguineous matings in plants. For example, in the self-incompatible plant *Helianthus annuus*, ELLSTRAND, TORRES and LEVIN (1978) inferred the occurrence of a significant amount of apparent selfing, based on the observation of excess homozygosity in progeny derived from known maternal genotypes. These studies relied on the mixed mating model to describe the effects of mating among relatives. However, the circumstances under which this is valid, as well as the appropriate modifications of the model when this is not valid, have not been specified.

An approach to this problem lies in the description of genetic relatedness between all pairs of mates and the elucidation of the effect of this relatedness

upon the genotypic composition of their progeny in terms of effective selfing. Various methods of describing and inferring relatedness are often used for relatively outbred human populations (CROW and KIMURA 1970; THOMPSON 1975; SALMON and BROCTEUR 1978; CANNINGS and THOMPSON 1981). However, populations of many plant species commonly display high levels of inbreeding (BROWN 1979), so that the complicating effects of inbreeding upon the description of genetic relatedness between individuals (HARRIS 1964; JACQUARD 1974) must be taken into consideration. This paper describes the effect of such relatedness upon progeny genotypic composition in terms of effective selfing. When analogies with the mixed mating model are used, a mating parameter termed the "effective selfing rate" is derived, and a model describing mating among relatives, termed the "effective selfing model", is described.

DESCRIPTION OF MATING BETWEEN RELATIVES AT A SINGLE LOCUS IN TERMS OF EFFECTIVE SELFING

Coefficients of identity by descent: The single locus representation of all possible relationships between relatives in an inbred population can be obtained using the concept of alleles identical by descent (MALECOT 1948). This concept allows the systematic description of the possible configurations of identity among the four alleles possessed by two mates at a single locus. Two alleles are *identical by descent* or "ibd" if they are both copies of an ancestral allele (JACQUARD 1974). One measure of inbreeding that can be derived using this concept is WRIGHT'S (1921) *inbreeding coefficient*, which may be defined as the probability that the two alleles at a locus in a diploid individual "A" are ibd (JACQUARD 1974) and in this paper is termed F_A. A second measure is the *coefficient of kinship*, defined as the probability that one allele, drawn at random from a locus in individual A, is ibd with a second allele, drawn at random from the same locus in a second individual B (MALECOT 1948) and is termed here Φ_{AB}.

GILLOIS (1964), in further detailing genetic relatedness between individuals, identified 15 possible configurations of alleles ibd between two diploid individuals, at a single locus. If one does not distinguish maternally from paternally derived alleles, nine "condensed identity modes" are a sufficient description of possible identity configurations. These nine identity modes, denoted as S'_i, $i = 1$ to 9, are diagrammed in Figure 1, as described by JACQUARD (1974). The top two dots of each identity mode in Figure 1 represent the two alleles possessed by the first individual A, and the bottom two dots of each mode represent the alleles of the second individual B. Identical alleles, both within and between individuals, are linked with lines. The corresponding probabilities of observing each identity mode at a given locus are (following the notation of GILLOIS) denoted by Δ_i, for $i = 1$ to 9. Examining Figure 1, one can see that the inbreeding coefficient of individual A is $\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, and the inbreeding coefficient of the individual B is $\Delta_1 + \Delta_2 + \Delta_5 + \Delta_6$. The coefficient of kinship between A and B is $\Delta_1 + (\Delta_3 + \Delta_5 + \Delta_7)/2 + \Delta_8/4$ (JACQUARD 1974).

Conditional gene probabilities of mates: The identity modes in Figure 1 can be
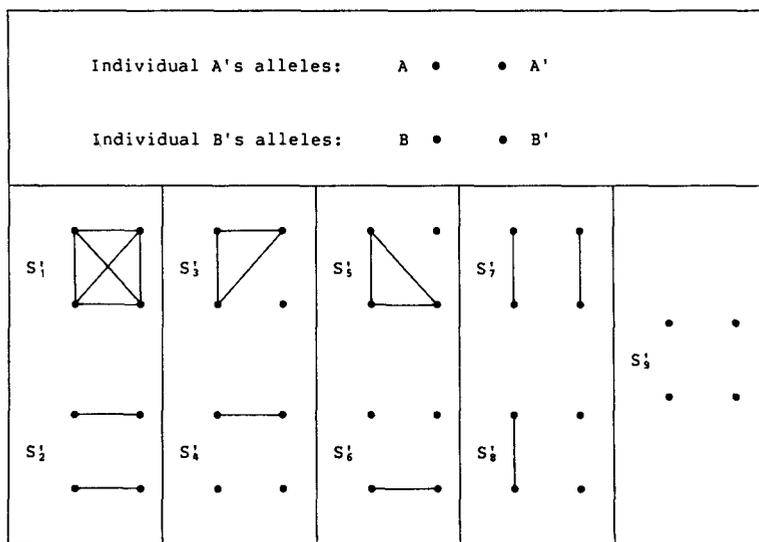
FIGURE 1.—GILLOIS' condensed identity modes. Each S'_i has a corresponding probability or identity coefficient Δ_i. Identical alleles are connected by a line (after JACQUARD 1974).

used to describe the genetic relatedness between mates. Specifying these "modes of mating" constrains the joint genotypic probabilities of pairs of mates. These constraints have an effect upon the genotypic composition of the progeny of mating pairs. Thus, the approach in this paper takes two steps to derive effective selfing: (1) consider the genotype of the first mate as given, and find out how specification of the Δ_i probabilities changes the probabilities of alleles possessed by the second mate; (2) find out how this conditioning of the second mate's gene probabilities affects the first mate's offspring genotypic composition, from the viewpoint of effectively selfing. The first step, which has been previously treated from the viewpoint of genealogy (JACQUARD 1974; CANNINGS and THOMPSON 1981), is now considered.

We start with the genotype of A as given. Denote the two alternative alleles from parent A as A and A', and from parent B as B and B'. Allelic likeness of state is indicated by subscripts, "$=$" denotes alleles ibd and "$\neq$" denotes alleles non-ibd; for example $(A_iA_i, A \neq A')$ denotes a homozygous parent with alleles non-ibd. If the parent A is homozygous for allele A_i and these alleles are ibd, the conditional probabilities for parent B's alleles at the same locus are, assuming no mutational or selectional changes (after JACQUARD 1974, p. 129, but using only Δ_1, Δ_2, Δ_3 and Δ_4 in equation 18),

$$P(B_i|(A_iA_i, A = A')) = \frac{(\Delta_1 + \Delta_3/2) + p_i(\Delta_2 + \Delta_3/2 + \Delta_4)}{F_A},$$
(1)

$$P(B_j|(A_iA_i, A = A'))|_{j \neq i} = \frac{p_j(\Delta_2 + \Delta_3/2 + \Delta_4)}{F_A},$$

in which p_i is the unconditional probability of allele B_i. This probability p_i can

be related to the frequency of A_i in a population defined sufficiently large enough that the descendants of the ancestors of A and B are randomly distributed. If the parent A is homozygous for allele A_i but these alleles are not ibd (only alike in state), the conditional allele probabilities for B are

$$P(B_i|(A_iA_i, A \neq A')) = \frac{(\Delta_5 + \Delta_7 + \Delta_8/2) + p_i(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)},$$ (2)

$$P(B_j|(A_iA_i, A \neq A'))|_{j \neq i} = \frac{p_j(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)}.$$

Finally, if the parent A is heterozygous $(A \neq A')$, the conditional allele probabilities for B are

$$P(B_k|A_iA_j)|_{k=i,j} = \frac{(\Delta_5/2 + \Delta_7/2 + \Delta_8/4) + p_k(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)},$$ (3)

$$P(B_k|A_iA_j)|_{k \neq i,j} = \frac{p_k(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)}.$$

as described by JACQUARD (1974) (equation 17, p. 129).

This dependence of allele probabilities upon the status of parent A with regard to allelic identity by descent will be the source of all differences between the effective selfing model and the mixed mating model.

Arrays of segregating progeny: The second aspect of these mating modes is their influence upon offspring genotypic composition, or the pattern of allelic segregation in terms of effective selfing. This segregation depends on the allele probabilities of parent A's mate(s), which have been shown to be functions of the mating mode probabilities, parent A's genotype and population gene frequencies p_i. To illustrate the segregation pattern, systematically arrange possible progeny genotypes of A in an array. The progeny genotypes of parents A_iA_i or A_iA_j are ordered as A_iB_i, A_iB_j, A_jB_j, A_iB_k and A_jB_k $(i \neq j \neq k)$.

The probability array of progeny derived from parental genotype A_iA_i with alleles ibd is, from equation 1,

$$\begin{bmatrix} P(A_iB_i) \\ P(A_iB_j) \\ P(A_jB_j) \\ P(A_iB_k) \\ P(A_jB_k) \end{bmatrix} = \begin{bmatrix} P(B_i|(A_iA_i, A = A')) \\ P(B_j|(A_iA_i, A = A')) \\ 0 \\ P(B_k|(A_iA_i, A = A')) \\ 0 \end{bmatrix}$$ (4)

$$= \frac{(\Delta_1 + \Delta_3/2)}{F_A} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \frac{(\Delta_2 + \Delta_3/2 + \Delta_4)}{F_A} \begin{bmatrix} p_i \\ p_j \\ 0 \\ p_k \\ 0 \end{bmatrix}$$

The probability array of progeny derived from parental genotype A_iA_i with alleles non-ibd is, from equation 2,

$$
\begin{bmatrix} P(A_iB_i) \\ P(A_iB_j) \\ P(A_jB_j) \\ P(A_iB_k) \\ P(A_jB_k) \end{bmatrix} = \begin{bmatrix} P(B_i|(A_iA_i,\ A \neq A')) \\ P(B_j|(A_iA_i,\ A \neq A')) \\ 0 \\ P(B_k|(A_iA_i,\ A \neq A')) \\ 0 \end{bmatrix}
$$

(5)

$$
= \frac{(\Delta_5 + \Delta_7 + \Delta_8/2)}{(1 - F_A)} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \frac{(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)} \begin{bmatrix} p_i \\ p_j \\ 0 \\ p_k \\ 0 \end{bmatrix}
$$

Finally, the probability array of progeny derived from the heterozygous parental genotype A_iA_j is, from equation 3,

$$
\begin{bmatrix} P(A_iB_i) \\ P(A_iB_j) \\ P(A_jB_j) \\ P(A_iB_k) \\ P(A_jB_k) \end{bmatrix} = (1/2) \begin{bmatrix} P(B_i|A_iA_j) \\ P(B_i|A_iA_j) + P(B_j|A_iA_j) \\ P(B_j|A_iA_j) \\ P(B_k|A_iA_j) \\ P(B_k|A_iA_j) \end{bmatrix}
$$

(6)

$$
= \frac{(\Delta_5 + \Delta_7 + \Delta_8/2)}{(1 - F_A)} \begin{bmatrix} 1/4 \\ 1/2 \\ 1/4 \\ 0 \\ 0 \end{bmatrix} + \frac{(\Delta_6 + \Delta_8/2 + \Delta_9)}{(1 - F_A)} \begin{bmatrix} p_i/2 \\ (p_i + p_j)/2 \\ p_j/2 \\ p_k/2 \\ p_k/2 \end{bmatrix}
$$

In equations 4, 5 and 6, the first array of the rightmost pair of arrays (the array with the quantities 1, ½ or ¼) contains the expected proportions of progeny if the specified genotype of parent A self-fertilizes. The second array (containing unconditional allele probabilities) contains the expected proportions of progeny if the specified genotype of parent A mates at random (outcrosses). Thus, the segregation pattern of offspring derived from a parental genotype is reducible to a linear combination of two segregational patterns: that expected under complete selfing, and that expected under complete outbreeding. The amount of this effective selfing is a function of the pedigree probabilities Δ_i, $i = 1, 9$; these probabilities fall into four groups in equations 4–6.

Effective selfing rates: A basic premise is that progeny that segregate according to equations 4–6 segregate as if some were derived from random outcrossing, and the remainder from self-fertilization, regardless of whether they actually were produced by self-fertilization. First, we define two indices of "conditional" effective selfing, so called because they premultiply the effective selfing arrays in equations 4–6 and have been conditioned by the identity of parental alleles: (1) the effective self-fertilization rate given that parent A has alleles ibd

$$
s_A^i = \frac{\Delta_1 + \Delta_3/2}{F_A}
$$

(7)

and (2) the effective self-fertilization rate given that parent A has alleles non-ibd

$$s_A^o = \frac{\Delta_5 + \Delta_7 + \Delta_8/2}{1 - F_A}. \tag{8}$$

The effective random outcrossing rate given that parent A has alleles ibd is $t_A^i = (\Delta_2 + \Delta_3/2 + \Delta_4)/F_A$, and the effective random outcrossing rate given that parent A has alleles non-ibd is $t_A^o = (\Delta_6 + \Delta_8/2 + \Delta_9)/(1 - F_A)$. Both are entirely determined by their corresponding conditional selfing rates: $t_A^i = 1 - s_A^i$ and $t_A^o = 1 - s_A^o$.

Since the probability that parent A has alleles ibd is F_A, and likewise alleles not-ibd is $(1 - F_A)$, the effective selfing rate of individual A, termed "E_A," is the weighted mean of s_A^i and s_A^o; i.e.,

$$E_A = F_A s_A^i + (1 - F_A)s_A^o.$$

The F_A terms in s_A^i and s_A^o cancel, giving E_A solely in terms of identity coefficients, i.e.,

$$E_A = \Delta_1 + \Delta_3/2 + \Delta_5 + \Delta_7 + \Delta_8/2. \tag{9}$$

The effective selfing rate depends on whether it is based on parent A or parent B in Figure 1. One can verify through verbal arguments, presented at the end of this section, that the effective selfing rate of parent B is $E_B = \Delta_1 + \Delta_3 + \Delta_5/2 + \Delta_7 + \Delta_8/2$, which differs from A's rate by $\Delta_3/2 - \Delta_5/2$. Thus, in the absence of any selection, it is possible for the female effective selfing rate to differ from the male effective selfing rate. This asymmetrical property does not occur with the coefficient of consanguinity between mates $\Phi_{AB} = \Delta_1 + (\Delta_3 + \Delta_5 + \Delta_7)/2 + \Delta_8/4$.

This derivation of E_A and E_B was based upon one mating pair. The effective rate of selfing in the population would be the average of E_A and E_B over all mating pairs. However, for purposes of data analysis, the effective selfing rate in the population will henceforth be considered, in this paper, as the average of only E_A, where A is the female parent. This is because the common parent of open pollinated plants is female and relatively easy to determine, whereas male parentage of such plants is difficult to ascertain. This average effective selfing rate of all individuals in a class A is termed "E." Likewise, subscripts will be dropped from s_A^i, s_A^o and F_A.

The effective selfing rate of A may be more directly derived by examining Figure 1 and using the following arguments. First, note these two principles: (1) if self-fertilized, a parent with alleles ibd will always produce progeny with alleles ibd and (2), if self-fertilized, a parent with alleles non-ibd will produce progeny with alleles ibd one-half of the time. Progeny of the mating defined by identity mode S_1' always have alleles ibd; hence this mode corresponds to complete selfing. A randomly chosen allele from B in mode S_3' is ibd to a randomly chosen allele from A half the time. Hence, their progeny will have alleles ibd half the time, and since parent A has alleles ibd, mode S_3' corre-

sponds to selfing parent A half of the time. Progeny of matings defined by identity mode S'_5 and S'_7 also will have alleles ibd half the time, but now parent A has alleles non-ibd, so that these two modes correspond to complete selfing. Progeny of mode S'_8 have alleles ibd one-fourth of the time, and parent A has alleles non-ibd, so that mode S'_8 corresponds to selfing half the time. The other modes (S'_2, S'_4, S'_6 and S'_9) correspond to complete outbreeding since they define no relatedness between the two parents.

Examination of Figure 1 reveals that the effective selfing rate E can rigorously be defined as "the probability that a randomly chosen allele from an individual's mate is identical by descent with *either* of that individual's alleles at the same locus, averaged over all individuals under consideration."

ESTIMATION OF EFFECTIVE SELFING WITH THE EFFECTIVE SELFING MODEL

The effective selfing model that will be used for estimation is summarized in Table 1, where the probabilities of observing progeny, given the maternal genotype, are given. If $s^i = s^o$ or $F = 0$, this description is identical with the mixed mating model, thus demonstrating that the mixed mating model is a special case of the effective selfing model. This additional consideration of allelic identity by descent warrants a slightly different approach for estimating E, than the one commonly used for estimating the actual self-fertilization rate s. It also results in genotypic frequencies alone not providing information about effective selfing, even at inbreeding equilibrium [when $F = s/(2 - s)$]. This is because the effective selfing model has three independent inbreeding parameters (F, Φ_{AB} and E; or alternatively F, s^i and s^o), and genotype data provide just one degree of freedom for one inbreeding parameter (F), leaving two parameters undetermined. At inbreeding equilibrium, $F = \Phi_{AB}$ or $F(1 - s^i) = (1 - F)s^o/2$, thus determining a second inbreeding parameter, but one parameter is still unknown. Estimation of E, using information provided by progeny arrays, will be examined following the next section.

Bias of estimate when the mixed mating model is used for a population with mating among relatives: If the mixed mating model is used to obtain estimates of self-fertilization in a population in which mating with relatives (other than self-fertilization) also occurs, the estimated selfing rate approximates the true effective selfing rate. If the estimation procedure is based on fitting expected to observed frequencies of parent-progeny pairs (BROWN and ALLARD 1970), an approximation for the deviation may be derived by using the relation $\Phi_{AB} = s(1 + F_A)/2$ (CROW and KIMURA 1970, p. 93), where s is the actual self-fertilization rate, F_A is the inbreeding coefficient of parents and Φ_{AB} the inbreeding coefficient of progeny. Then, $\hat{s} - E \simeq 2\Phi_{AB}/(1 + F_A) - E$, or

$$\hat{s} - E \simeq \frac{F_A(1 - F_A)(s^i - s^o)}{1 + F_A}. \tag{10}$$

Alternatively, if inbreeding equilibrium is assumed ($F_A = \Phi_{AB}$ assumed) and one uses $\hat{s} = 2F_A/(1 + F_A)$, the bias is

TABLE 1

Single locus probabilities of observing progeny genotypes, given the maternal parent genotype, in the effective selfing model

Progeny genotype	Maternal parent genotype		
	A_iA_i $(A = A')$	A_iA_i $(A \neq A')$	A_iA_j $(A \neq A')$
A_iA_i	$s^i + t^i p$	$s^o + t^o p$	$s^o/4 + t^o p/2$
A_iA_j	$t^i q$	$t^o q$	$s^o/2 + t^o(p + q)/2$
A_jA_j	0	0	$s^o/4 + t^o q/2$
A_iA_k	$t^i r$	$t^o r$	$t^o r/2$
A_jA_k	0	0	$t^o r/2$

The effective selfing rate of inbred parents $(A = A')$ is s^i $(= 1 - t^i)$ and their frequency is F, and the effective selfing rate of outbred parents $(A \neq A')$ is s^o $(= 1 - t^o)$ and their frequency is $1 - F$. The frequencies of alleles A_i, A_j and A_k are p, q and r, respectively.

$$\hat{s} - E = \frac{2F_A(1 - s^i) - (1 - F_A)s^o + F_A(1 - F_A)(s^i - s^o)}{1 + F_A} \qquad (11)$$

$(2F_A(1 - s^i) - (1 - F_A)s^o$ is zero when $F_A = \Phi_{AB})$.

When $s^i > s^o$ (inbred parents effectively self more than outbred parents) the mixed mating model, using either estimation procedure, erroneously infers more effective selfing than has actually occurred. The bias term $F_A(1 - F_A)/(1 + F_A)$ is greatest at $F_A = \sqrt{2} - 1$, and tapers off to zero near $F_A = 0$ or $F_A = 1$. If $s^i < s^o$, this error is reversed in sign, but there is reason to suspect $s^i > s^o$ is usually the case (see next section). If the inbreeding equilibrium estimate is used but equilibrium has not yet been attained by the population ($\Delta H < 0$, or $\Phi_{AB} > F_A$), the estimate of self-fertilization tends to suggest that there is less effective selfing than there actually is. But at inbreeding equilibrium, equation 11 shows bias still can exist if $s^i \neq s^o$. If $s^i = s^o$ (an assumption of the mixed mating model) or $F_A = 0$, both methods, with their assumptions, give an unbiased estimate of E. There probably are circumstances when a population with mating among relatives, other than self-fertilization, approximate $s^i = s^o$.

Estimation using progeny arrays: The maximum likelihood method was applied to progeny array data from RITLAND and JAIN (1981) as an illustration of estimation, with three objectives: to (1) determine the relative values of $\hat{s}^i$ and $\hat{s}^o$ for each of the three loci assayed, (2) calculate $\hat{E}$, $\hat{\Phi}_{AB}$ and $\hat{F}$, and check the deviation of $\hat{s}$ from $\hat{E}$ given by equation 10, for each locus, and (3) compare the effective selfing rate with the multilocus selfing rate (which is closer to the actual selfing rate; SHAW, KAHLER and ALLARD 1981). The likelihood equation and estimation procedure are outlined in the APPENDIX. Two populations of the annual plant *Limnanthes alba* were electrophoretically assayed: (1) "Mather," with 37 progeny arrays (families) assayed for two glutaminoxaloacetic transaminase loci (*Got-2* and *Got-3*) and esterase (*Est*), and (2) "Ingot," with 36 arrays assayed for peroxidase (*Prx*), *Got-3* and *Est*.

For the Mather population, the effective selfing rate estimate was found to equal the estimate of selfing given by the single locus mixed mating model at all three loci, due to the inference of $F = 0$ at all three loci, so these estimates

TABLE 2

Estimates of effective selfing model parameters, and comparisons with mixed mating model selfing estimate ŝ, at three loci for the Ingot population of Limnanthes alba

		Effective selfing model								Difference	
		Conditional self-ing rates		Inbreeding parameters			Gene frequencies		Mixed mating model	$\hat{s} - \hat{E}$	
Locus	N	$\hat{s}^i$	$\hat{s}^o$	$\hat{E}$	$\hat{\Phi}_{AB}$	$\hat{F}_A$	$\hat{p}$	$\hat{o}$	$\hat{s}$	Observed	Predicted
Prx	392		0.22	0.22	0.12	0.00	0.65	0.61	0.22	0.0	0.0
C.I.			0.24	0.24	0.15	0.33	0.09	0.11	0.14		
Got-3	367	0.64	0.16	0.39	0.35	0.47	0.81	0.79	0.46	0.07	0.08
C.I.		0.24	0.16	0.18	0.20	0.36	0.07	0.12	0.16		
Est	392	0.06	0.65	0.51	0.26	0.24	0.57	0.48	0.43	−0.08	−0.09
C.I.		*a*	*a*	*a*	0.13	0.32	0.10	0.13	0.13		

Predicted difference of $\hat{s} - \hat{E}$ is based on equation 10. The 95% confidence interval (C.I.) is given below each estimate. Data from RITLAND and JAIN (1981).

a Large due to gene frequencies near ½.

are not given here. For the Ingot population, estimates of effective selfing differed at two of three loci (Table 2): at *Got-3* $\hat{E} = 0.39$, and at *Est* $\hat{E} = 0.51$. This compares to $\hat{s} = 0.46$ and $\hat{s} = 0.43$, respectively, given by the mixed mating model. These differences agree with predictions given by equation 10 (Table 2). *Got-3* gave $\hat{s}^i = 0.64$ and $\hat{s}^o = 0.16$, whereas *Est* gave nearly the reverse, $\hat{s}^i = 0.06$ and $\hat{s}^o = 0.65$, but the confidence intervals for *Est* are large due to gene frequencies at this diallelic locus being near 0.5 (see APPENDIX). The significantly greater s^i at *Got-3* would be expected in a subdivided, inbred population, wherein patches of alleles ibd are maintained by greater effective selfing, relative to outbred patches with the same actual selfing rate. It is not unreasonable for s^i to be less than s^o at the *Est* locus, despite $s^i > s^o$ at *Got-3*, because the rate of effective selfing depends on the population structure, which may vary from locus to locus due to chance historical factors.

The confidence interval given for $\hat{s}$ at the *Prx* locus is smaller than for $\hat{E}$ because the inverted information matrix included only s and p in the work by RITLAND and JAIN (1981), whereas pollen and ovule gene frequencies, F, and conditional effective selfing rates s^i and s^o were included in the information matrix here. However, only one progeny per family was assumed in its computation. Despite the confidence interval for $\hat{F}$ being wide due to the "small" sample size of 36 (progeny arrays), $\hat{F}$ at *Got-3* was significant. More arrays would have been desirable, but this decreases the probability of correctly classifying progeny arrays, for a fixed total sample size. The coefficient of kinship between mates Φ_{AB} was also significant at both *Got-3* and *Est*, but the sample size, assumed to be the geometric mean of number of families and total number of individuals, may be incorrect.

Finally, the multilocus estimate of selfing for Ingot, based on the multilocus mixed mating model of RITLAND and JAIN (1981), was 0.32; the mean effective selfing rate for all three loci was 0.37, indicating that at least 5% of the

effective selfing on this population was due to mating with relatives beyond self-fertilization.

<div align="center">DISCUSSION</div>

To characterize the genotypic distribution of offspring resulting from matings with relatives in an inbred population, three independent parameters of single locus genetic and mating structure are required. The first two are the inbreeding coefficient of parents, F, and the coefficient of kinship between mates, Φ_{AB}. The third is derived in this paper and named the effective selfing rate. The effective selfing rate, E, reflects the degree of relatedness between mates, as it measures the amount of genetically equivalent selfing incurred by mating with relatives. It is defined using GILLOIS' (1964) condensed identity modes, which are used to specify genetic relationship, as $E_A = \Delta_1 + \Delta_3/2 + \Delta_5 + \Delta_7 + \Delta_8/2$ for the parent A in Figure 1 and as $E_B = \Delta_1 + \Delta_3 + \Delta_5/2 + \Delta_7 + \Delta_8/2$ for parent B in Figure 1. In this paper its use was primarily with reference to parent A. The effective random outcrossing rate is $(1 - E)$. The effective selfing rate E is rigorously defined as "the probability that an allele chosen at random from an individual's mate is identical by descent with *either* allele at the same locus in that individual, averaged over all individuals under consideration." This qualification of identity with either allele makes E dependent upon the parent referred to in Figure 1.

The three independent parameters of genetic and mating structure may alternatively be formulated as the inbreeding coefficient F and two indices of conditional effective selfing: s^i, the effective selfing rate of parents with alleles identical by descent, and s^o, the effective selfing rate of parents with alleles not identical by descent. This latter formulation provides a clearer picture of the transmission process (Table 1) and demonstrates that the primary difference between the model presented in this paper (the effective selfing model) and the mixed mating model is the distinction of the effective selfing rates of inbred *vs.* outbred parents. This difference is ultimately due to the dependence of allele frequencies of a parent's mate, when there is mating among inbred relatives, upon the status of that parent with regard to allelic identity by descent.

The coefficient of kinship between mates A and B, Φ_{AB}, is a parameter of the mating system, as it describes how gametes from mating pairs give rise to zygotes, and in many cases it is the correlation between the gametic values of mates. JACQUARD'S (1975) coefficient of deviation from random mating, δ, in the context of mating with relatives, is this Φ_{AB}. The effective selfing rate E is close to Φ_{AB}, as they are are both functions of the conditional selfing rates s^i and s^o, but they differ by $(1 - F)s^o/2$. The rate of effective selfing is, thus, not directly related to the population's overall change in heterozygosity from parents to progeny, $F - \Phi_{AB}$, which equals $(F - E) + (1 - F)s^o/2$. However, E is directly related to ΔH *among* parents with alleles ibd (where $\Delta H = 1 - s^i$) and among parents with alleles non-ibd (where $\Delta H = -s^o/2$). Thus, E does mediate the change in heterozygosity, but this change is conditional on allelic

identity by descent. This indicates that values of E, in nature, may involve ibd-dependent balances between the costs and benefits of inbreeding $vs.$ outbreeding. In particular, the fitness characteristics of s^i may differ from those of s^o because a locus with alleles ibd should show no further deleterious effects in progeny derived by effective selfing.

The mixed mating model model does not distinguish between the effective selfing rates of inbred parents and outbred parents. When $F = 0$ or $s^i = s^o$, the mixed mating model is equivalent to the effective selfing model. However, the values of of s^i and s^o in natural populations are as yet largely unknown, so only those estimates of self-fertilization based on completely outbred parental populations are reliable estimates of E. Consideration of s^i and s^o should be of help in understanding the patchiness of pollen gene frequencies sometimes found (BROWN, MATHESON and ELDRIDGE 1975) and the relative roles of pollen and seed flow in promoting microgeographical gene differentiation (LEVIN and KERSTER 1974; RAI and JAIN 1982), as well as the frequency of effective selfing beyond self-fertilization, which is best documented in known self-incompatible species (LEVIN 1977; ELLSTRAND, TORRES and LEVIN 1978).

LITERATURE CITED

ALLARD, R. W., S. K. JAIN and P. L. WORKMAN, 1968 The genetics of inbreeding populations. Adv. Genet. 14: 55–131.

BROWN, A. H. D., 1979 Enzyme polymorphism in plant populations. Theor. Pop. Biol. 15: 1–42.

BROWN, A. H. D. and R. W. ALLARD, 1970 Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. Genetics 72: 489–503.

BROWN, A. H. D., A. C. MATHESON and K. G. ELDRIDGE, 1975 Estimation of the mating system of Eucalyptus oblique L'Hérit. using allozyme polymorphisms. Aust. J. Bot. 23: 931–949.

CANNINGS, C. and E. A. THOMPSON, 1981 Genealogical and Genetic Structure. Cambridge University Press, Cambridge.

CEPPELLINI, R., M. SINISCALCO, and C. A. B. SMITH, 1955 The estimation of gene frequencies in a random-mating population. Ann. Hum. Genet. 20: 97–115.

CLEGG, M. T., A. L. KAHLER, and R. W. ALLARD, 1978 Estimation of life cycle components of selection in an experimental plant population. Genetics 89: 765–792.

CROW, J. F. and M. KIMURA, 1970 An Introduction to Population Genetics Theory. John Wiley & Sons, New York.

ELLSTRAND, N. C., A. M. TORRES, and D. A. LEVIN, 1978 Density and apparent rate of outcrossing in Helianthus annuus (Asteraceae). Syst. Bot. 3: 403–407.

GILLOIS, M., 1964 La relation d'identité en génétique. Thèse Fac. Sciences, Paris.

HARRIS, P. L., 1964 Genotypic covariances between inbred relatives. Genetics 50: 1319–1348.

HAYMAN, B. I., 1953 Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity 7: 185–192.

JACQUARD, A., 1974 *The Genetic Structure of Populations.* Springer-Verlag, New York.

JACQUARD, A., 1975 Inbreeding: one word, several meanings. Theor. Pop. Biol. **7:** 338–363.

JAIN, S. K., 1976 The evolution of inbreeding in plants. Annu. Rev. Ecol. Syst. **7:** 469–495.

LEVIN, D. A., 1977 The organization of genetic variability in *Phlox drummondii.* Evolution **31:** 477–494.

LEVIN, D. A. and H. W. KERSTER, 1974 Gene flow in seed plants. Evol. Biol. **7:** 139–220.

MALECOT, G., 1948 *Les Mathematiques de l'heredite.* Masson et Cie, Paris.

RAI, K. N. and S. K. JAIN, 1982 Population biology of *Avena.* IX. Gene flow and neighborhood size in relation to microgeographic variation in *Avena barbata.* Oecologia **53:** 399–405.

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

SALMON, D. and J. BROCTEUR, 1978 Probability of paternity exclusion when relatives are involved. Am. J. Hum. Genet. **30:** 65–75.

SHAW, D. V., A. L. KAHLER and R. W. ALLARD, 1981 A multilocus estimator of mating system parameters in plant populations. Proc. Natl. Acad. Sci. USA **78:** 1298–1302.

THOMPSON, E. A., 1975 The estimation of pairwise relationships. Ann. Hum. Genet. **39:** 173–188.

WEIR, B. S. and C. C. COCKERHAM, 1973 Mixed selfing and random mating at two loci. Genet. Res. **21:** 247–262.

WRIGHT, S., 1921 Coefficients of inbreeding and relationship. Am. Nat. **56:** 330–338.

Corresponding editor: C. F. WEHRHAHN

APPENDIX

Maximum likelihood estimation: The data consists of many progeny arrays (families); each array consists of many progeny plants derived from an open pollinated common mother of unknown genotype, and each progeny has been assayed for an n-allele marker locus ($n \geq 2$). Denote the sample progeny array for family k as $\mathbf{x}_k$ with elements x_{jk} containing observed numbers of genotypes, ordered like the genotypes in equations 4–6, and allowing n alleles. Denote $\mathbf{s}_m$ with elements s_{jm} as the array corresponding to self-fertilization of the parent of genotype m ($m = 1, (n^2 + n)/2$) in equations 4–6, and $\mathbf{t}_m$ with elements t_{jm} as the array corresponding to random outcrossing of genotype m in equations 4–6. To begin the estimation procedure, we need provisional estimates of s^i, s^o, maternal inbreeding coefficient F, pollen gene frequencies p_i, $i = 1, n$, and maternal (ovule) genotypic frequencies $f(m|A = A')$, $f(m|A \neq A')$. Denote this set of parameters as $\mathbf{a}$.

The procedure used to analyze the data in this paper assumed gene frequencies of inbred maternal plants equaled those of outbred maternal plants, and genotypic frequencies of outbred maternal plants were in Hardy-Weinberg. Other assumptions include: no selection between zygote formation and census, s^i and s^o are independent of maternal genotypic *state* and pollen gene frequency is uniform over maternal genotypes that effectively outcross. As this procedure is based upon open pollinated plants, E can be found for only the maternal population. This approach most closely follows that of CLEGG, KAHLER and ALLARD (1978).

The likelihood of progeny array x_k with maternal parent genotype m whose alleles are ibd is (omitting the combinatorial):

$$L_{mk}^i(\mathbf{x}_k; \mathbf{a}) = Ff(m|A = A') \prod_j [s^i s_{jm} + (1 - s^i)t_{jm}]^{x_{jk}}$$

and whose alleles are non-ibd is

$$L_{mk}^o(\mathbf{x}_k; \mathbf{a}) = (1 - F)f(m|A \neq A') \prod_j [s^o s_{jm} + (1 - s^o)t_{jm}]^{x_{jk}}.$$

The likelihood of the progeny array with maternal parent m is the sum of inbred and outbred likelihoods, i.e.,

$$L_{mk}(x_k; a) = L_{mk}^i(x_k; a) + L_{mk}^o(x_k; a).$$ (12)

The overall likelihood of progeny array k, taking into account alternative maternal genotypes, is

$$\sum_m L_{mk}(x_k; a),$$ (13)

so that the likelihood of the matrix of all sampled progeny arrays X is (assuming mutual independence) the product over all arrays k,

$$L(X; a) = \prod_k \sum_m L_{mk}(x_k; a).$$ (14)

Maximum likelihood estimates are obtained by finding the a maximizing $L(X; a)$.

Since information about maternal parentage (needed to estimate the other parameters) is provided by the segregation pattern of the progeny array, the number of plants in each array must be large enough to sufficiently distinguish between alternative parents ($\geq\approx 9$ for outbred populations and $\geq\approx 5$ for inbred populations under mixed mating, BROWN and ALLARD 1970). If maternal parentage is known by previous assay, equation 13 is not needed, but $L(X, a)$ still requires the $f(m|A = A')$ and $f(m|A \neq A')$ for joint estimation, since the variances of the other parameters (obtained by inverting the information matrix for a) includes a component due to sampling of maternal genotypes. Since family size was sufficiently large, the estimation procedure in this paper discarded the less likely $L_{mk}(x_k; a)$ [$m = 1$, $(n^2 + n)/2$ for each k] and in this aspect, resembled BROWN and ALLARD (1970), although they did not estimate parental genetic parameters.

Except for maternal genotypic frequencies when maternal genotypes are previously assayed, all parameters maximizing equation 14 must be solved by iterative methods. For example, $\hat{F}$ satisfies

$$0 = \frac{d}{dF} \log\{L(X; a)\}|\hat{F} = \sum_k \frac{\sum_m \{(L_{mk}^i(x_k; \hat{a})/\hat{F}) - (L_{mk}^o(x_k; \hat{a})/(1 - \hat{F}))\}}{\sum_m \{L_{mk}^i(x_k; \hat{a}) + L_{mk}^o(x_k; \hat{a})\}},$$

which multiplied by $\hat{F}(1 - \hat{F})/n_f$ (n_f = number of families) and rearranged is

$$\hat{F} = \frac{1}{n_f} \sum_k \frac{\sum_m L_{mk}^i(x_k; \hat{a})}{\sum_m \{L_{mk}^i(x_k; \hat{a}) + L_{mk}^o(x_k; \hat{a})\}},$$ (15)

which must be iterated ($\hat{a}$ includes $\hat{F}$) but converges nicely. This is analogous to the iterative gene counting method of CEPPELLINI, SINISCALCO and SMITH (1955). Interestingly, when $s^i \neq s^o$, progeny array data provide information about F that deviations from Hardy-Weinberg cannot give: information about allelic ibd of specific maternal genotypes. The iterative equations for maximum likelihood estimates of s^i, s^o, and pollen and ovule gene frequencies were found with expressions similar to equation 15. $\hat{E}$ and $\hat{\Phi}_{AB}$ were then found as $\hat{E} = \hat{F}\hat{s}^i + (1 - \hat{F})\hat{s}^o$ and $\hat{\Phi}_{AB} = \hat{F}\hat{s}^i + (1/2)(1 - \hat{F})\hat{s}^o$.

Information about parameters: Parameters are not jointly estimable if they index identical distribution functions and would display high statistical variance and covariance if they index relatively similar functions. For certain values of a, these problems occur for a diallelic locus. Inversions of the information matrix of a, assuming maternal genotype is known and family size is 1, have shown that for a diallelic locus, the variance of E as a function of gene frequency p, assuming equal pollen and ovule frequencies, is a minimum of $p = 0.75$ or 0.25, but at $p = 0.5$, E is not estimable, apparently due to the lack of information about selfing provided by heterozygous maternal plants. Triallelic loci with equal gene frequencies give much better estimates, with variances about 1/5 the minimum obtainable from a diallelic locus. With known maternal genotype and family size of one, data from a diallelic locus has 5 degrees of freedom (2 among maternal genotype

classes, plus 3 within their progeny genotype distributions). But for larger family sizes when $s^i \neq s''$, the degrees of freedom are nearly doubled for a diallelic locus (d.f. = 9: 3 given the maternal parent has alleles ibd, plus 5 given alleles non-ibd, plus 1 between). This is because, when $s^i \neq s^o$, progeny arrays allow the separate classification of maternal parents with alleles ibd from maternal parents with alleles non-ibd.