

## ESTIMATION OF THE LEVEL OF APOMIXIS IN PLANT POPULATIONS

D. R. MARSHALL and A. H. D. BROWN

*Division of Plant Industry, CSIRO, Canberra, Australia*

Received 5.vi.73

### SUMMARY

Models for estimating the level of apomixis in plant species which employ a flexible reproductive pathway are presented. These models depend on segregation data of marker genes within progeny arrays. The estimators of mating system parameters and their variance are derived for various breeding systems using either dominant or co-dominant diallelic loci.

For dominant marker loci, the progeny of recessive homozygotes generally contain the most information about the level of apomixis ( $c$ ). Additional data on heterozygous families are required in order to estimate the gene frequency in the pollen pool ( $p$ ) or the fraction of ovules self-pollinated ( $s$ ). In species capable of self-fertilisation, such data are needed to discriminate between the selfing and apomictic components and thus avoid overestimates of apomixis.

At co-dominant marker loci, the progeny of heterozygous mothers contribute more information about  $c$  than those of the more frequent homozygote. They can also estimate apomixis without any need for an estimate of the pollen gene frequency. All three parameters ( $c$ ,  $s$  and  $p$ ) are estimable when progeny data are available from all three kinds of maternal parents.

A multigenic model depending simply on genotypic comparisons is also discussed.

These progeny testing methods are much more applicable to populations where the level of apomixis is high than to detecting rare apomictic events. The latter are detected more efficiently by cytological procedures. The merits and demerits of estimating apomixis by progeny tests are discussed.

### 1. INTRODUCTION

BREEDING systems play a critical role in determining the population structure and evolutionary potential of plant species. They also have important implications in the development of appropriate breeding procedures for crop improvement (Bashaw, Hovin and Holt, 1970). Consequently quantitative estimates of mating system parameters under field conditions are required by both population biologists and plant breeders.

The classical procedures for the estimation of the level of agamospermy in facultative apomicts involve the cytological screening of sectioned ovules during megasporogenesis. These procedures provide direct and accurate estimates of the relative frequencies of apomictic and sexual embryo sacs. However, they are laborious and time consuming, and this usually precludes the scoring of large numbers of plants. Furthermore the relative frequency of apomictic embryo sacs may not accurately reflect the relative frequency of mature apomictic seed, particularly when pollination is uncertain or there is marked zygotic competition (Barlow, 1958).

The present paper describes alternative procedures for the estimation of the level of agamospermy in facultative apomicts which avoid a number of the deficiencies inherent in the classical cytological methods. These procedures are analogous to those used in the estimation of the level of

outcrossing in predominantly self-pollinating species (Fyfe and Bailey, 1951; Allard and Workman, 1963; Harding and Tucker, 1964), and involve progeny testing known genotypes taken at random from populations polymorphic for specific marker loci. The maximum likelihood method is then used to estimate the level of apomixis in the population from data on the numbers of each genotype appearing in the progeny of the test plants.

## 2. METHODS OF ESTIMATION

In this section the estimation models appropriate for specific mating systems, marker loci and estimable parameters will be discussed as follows:

Mating system	Marker loci	Maternal parent	Estimable parameter
(i) a Apomixis ( $c$ ) and random	Dominant ( $B/b$ )	$bb$	$c$
b outcrossing ( $t$ ) [ $c+t=1$ ]		$Bb$	$c$
c	Co-dominant ( $B_1/B_2$ )	$Bb$ and $bb$	$c$ and $p$
d		$B_1B_1$ or $B_2B_2$	$c$
e		$B_1B_2$	$c$ and $p$
f	Many loci	All differ	$c$
(ii) a Apomixis ( $c$ ), selfing ( $s$ )	Dominant	$Bb$ and $bb$	$c$ and $s$
b and random outcrossing	Co-dominant	$B_1B_2$	$c$ and $s$
c ( $t$ ) [ $c+s+t=1$ ]		$B_1B_1, B_1B_2, B_2B_2$	$c, s$ and $p$

In all models the frequency of apomixis is assumed to be constant for all plants within a population. It is also assumed that autosegregation (Gustafsson, 1947) is absent; that is, all apomictic progeny are genetically identical to their maternal parents. Model specific assumptions and symbols are defined as each model is outlined.

### (i) *Species reproducing by mixed agamospermy and random outcrossing*

#### (a) *Dominant marker loci—recessive homozygotes*

Consider a plant population polymorphic for a dominant diallelic locus with genotypic frequencies  $D(BB)$ ,  $H(Bb)$  and  $R(bb)$  in which there is a constant probability,  $t$ , of random outcrossing and a constant probability,  $c$ , of agamospermy ( $c+t=1$ ). If a random sample of recessive homozygotes is selected from the population and their progeny scored for the numbers of individuals with recessive ( $a_0$ ) and dominant ( $a_1$ ) phenotypes, we have the following expectations:

Class	Observed number	Expected number
Recessive ( $bb$ )	$a_0$	$n(c+ tq)$
Dominant ( $Bb$ )	$a_1$	$\frac{ntp}{n}$
Total	$n$	$n$

where  $p$  is the frequency of the dominant allele in the pollen pool and  $q$  is the frequency of its recessive counterpart ( $p+q=1$ ). When all genotypes contribute equally to the pollen pool,  $p = D + \frac{1}{2}H$  and  $q = R + \frac{1}{2}H$ . The maximum likelihood estimate of  $c$ , obtained by equating either of the observed numbers with their expected values (Bailey, 1951) is:

$$\hat{c} = 1 - a_1/np. \quad (1)$$

If  $p$  is known, as is the case when the population under study is a contrived one, set up specifically to estimate mating system parameters, the variance of  $\hat{c}$  is given by:

$$V_c = t(1 - pt)/np \tag{2}$$

and the information per plant (Mather, 1957):

$$i_f = 1/nV_c = p/t(1 - pt). \tag{3}$$

However, if the true gene frequency is unknown, as is the case, in most studies on natural populations,  $p$  (or  $q$ ) must be estimated from a second random sample of plants selected from the population and this estimate used in (1) above. In this situation, the variance in  $\hat{c}$  includes sampling errors in both  $a_1$  and  $p$  and we have (Kendall and Stuart, 1969):

$$V_c \approx \frac{1}{n^2 p^2} (V_{a_1}) + \left[ \frac{a_1}{np^2} \right]^2 (V_p) - \frac{2a_1}{n^2 p^3} (C_{a_1, p}). \tag{4}$$

Where  $V_{a_1}$ ,  $V_p$  and  $C_{a_1, p}$  represent the variance of  $a_1$ , the variance of  $\hat{p}$  and the covariance between  $a_1$  and  $\hat{p}$ , respectively.

If errors in  $a_1$  and  $p$  are independent, the covariance term can be ignored. On simplification (4) becomes:

$$V_c \approx t(1 - pt)/np + V_p t^2/p^2. \tag{5}$$

It will be noted that the above expectations are identical, taking into account differences in terminology, to those given by Fyfe and Bailey (1951) for species with mixed selfing and random mating. As a consequence, formula (1) can be used to estimate the level of self-pollination in facultative inbreeders as well as the level of agamospermy in facultative apomicts. The fact that apomixis and self-pollination have identical genetic consequences in progeny tests has important practical implications. In particular, it means that application of the above method to facultative apomicts which are wholly or partially self-pollinated will lead to serious overestimates of apomixis. Procedures which avoid this difficulty and permit the estimation of both the level of agamospermy and level of self-pollination in such species are given in Section (ii).

(b) *Dominant marker loci—heterozygotes*

If a random sample of dominants is selected from a polymorphic population and the numbers of recessives ( $a_0$ ) and dominants ( $a_1$ ) scored in the progeny of segregating heterozygotes, we have the following expectations:

Class	Observed number	Expected number
Recessive ( $bb$ )	$a_0$	$ntq/2$
Dominants ( $B-$ )	$a_1$	$n[c + t(1 + p)]/2$
Total	$n$	$n$

In this case the maximum likelihood estimate of  $c$  is:

$$\hat{c} = 1 - 2a_0/nq. \tag{6}$$

If  $p$  is known:

$$V_c = t(2 - qt)/nq. \tag{7}$$

And the information per plant:

$$i_f = q/t(2-qt). \quad (8)$$

Alternatively, if an independent estimate of  $p$  is obtained experimentally

$$V_c \approx t(2-qt)/nq + V_p t^2/q^2. \quad (9)$$

In practice, there is little incentive to use heterozygotes at a dominant marker locus for the estimation of  $c$  alone. First, many of the selected dominants would be homozygotes and these provide no information about the level of agamospermy in the population. Second, for all values of  $c$ ,

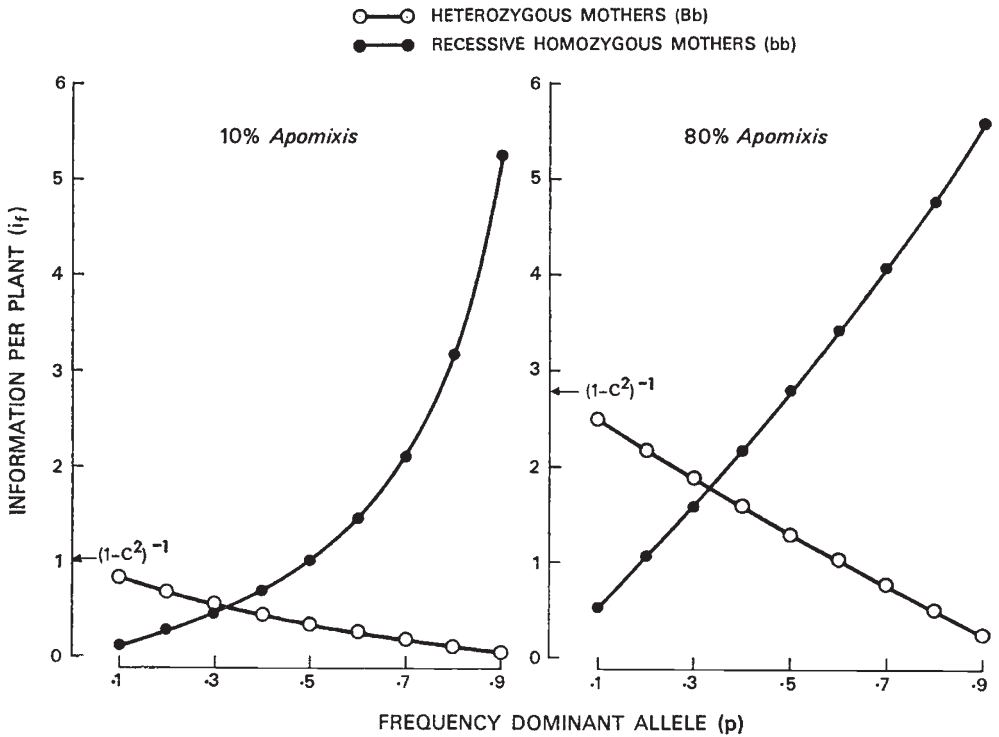


FIG. 1.—Information per plant ( $i_f$ ) for the level of apomixis at a dominant locus in heterozygous families compared with recessive homozygous families.

when  $p > 1/3$  the progenies of heterozygotes give less information per plant and therefore, are less efficient statistically, than progenies of recessive homozygotes (fig. 1). However, as discussed below data obtained from progenies of heterozygotes are required for the joint estimation of  $c$  and  $p$  (section (i) (c)) or  $c$  and the level of selfing (section (ii) (a)) in the population.

(c) *Dominant marker loci—heterozygotes and recessive homozygotes*

Previously, we considered the case where a sample of recessive homozygotes or heterozygotes is taken from the population to estimate  $c$  and a second sample is taken to estimate  $p$ . An alternative procedure is to take a single random sample of plants from the population and score the numbers

of recessive and dominants in the progenies of both the recessive homozygotes and heterozygotes. The two types of progenies provide two degrees of freedom for the joint estimation of  $p$  and  $c$ . The expectations for each progeny type are:

Class	Observed number	Expected number
Recessive homozygote		
Recessives ( $bb$ )	$a_0$	$n_1 [1 - p(1 - c)]$
Dominants ( $Bb$ )	$a_1$	$n_1 [p(1 - c)]$
Total	$n_1$	$n_1$
Heterozygotes		
Recessives ( $bb$ )	$a_2$	$n_2 [(1 - c)(1 - p)]/2$
Dominants ( $B-$ )	$a_3$	$n_2 [1 + c + p(1 - c)]/2$
Total	$n_2$	$n_2$

Setting the observations equal to their expectations and solving gives the maximum likelihood estimates,

$$\begin{aligned} \hat{c} &= 1 - (2a_2n_1 + a_1n_2)/n_1n_2 \\ \hat{p} &= a_1n_2/(2a_2n_1 + a_1n_2). \end{aligned} \tag{10}$$

The variance of the estimates are:

$$\begin{aligned} V_p &= pq(qx + py)/t \\ V_c &= t(px + qy) \end{aligned} \tag{11}$$

where  $x = (1 - pt)/n_1$  and  $y = (2 - qt)/n_2$ . The above procedure has the advantage that only a single random sample of plants needs to be drawn from the population. Further, it provides a direct estimate of the gene frequency in the male gametophytic population (pollen pool) and this can be compared with gene frequency in the parental population to test whether all genotypes do in fact contribute equally to the pollen pool (Harding and Tucker, 1969).

The variance formulae (11) are useful also when some initial estimates of  $p$  and  $t [= 1 - c]$  have been made, but the experimenter would like to increase the size of the progeny test, to obtain a more precise estimate of  $c$ . The question arises concerning what proportion of the total experimental effort ( $N = n_1 + n_2$ ) should be devoted to the progeny of heterozygous compared with those of recessive mothers ( $k = n_2/n_1$ ). The value of  $k$  for which  $V_c$  is minimum for a fixed value of  $N$  is given by:

$$k = \sqrt{\frac{q(2 - tq)}{p(1 - tp)}} \tag{12}$$

Table 1 gives the ratios of total progenies of heterozygous to homozygous mothers which are maximally efficient. These values depend far more on gene frequency than on the mating system parameter, and in general show that the infrequent class in the original population sample should be increased.

(d) *Co-dominant marker loci—homozygotes*

The estimation procedures for single homozygotes at co-dominant

marker loci are identical with those given in (a) above for recessive homozygotes at dominant marker loci.

(e) *Co-dominant marker loci—heterozygotes*

If we select a random sample of heterozygotes from a population polymorphic for a co-dominant diallelic locus with genotypic frequencies

TABLE 1  
*Optimum ratio of the total number of progeny from heterozygotes to the number from recessive homozygotes to minimise the variance of the estimated rate of apomixis ( $V_c$ )*

$c$	Frequency of dominant allele ( $p$ )				
	0.1	0.3	0.5	0.7	0.9
0.95	4.21	2.16	1.42	0.94	0.48
0.50	3.83	2.13	1.51	1.10	0.63
0.10	3.43	2.09	1.68	1.42	1.06
$2p/q^*$	0.22	0.86	2.0	4.7	18

\*  $2p/q$  represents the anticipated frequency ratio of maternal parents in the population, assuming no selection.

$D(B_1B_1)$ ,  $H(B_1B_2)$  and  $R(B_2B_2)$  and score the numbers of homozygotes ( $a_0$ ) and heterozygotes ( $a_1$ ) in their progeny, we have the following expectations:

Class	Observed number	Expected number
Homozygotes ( $B_1B_1, B_2B_2$ )	$a_0$	$tn/2$
Heterozygotes ( $B_1B_2$ )	$a_1$	$\frac{(c+t/2)n}{n}$
Total	$n$	$n$

Because we are considering the combined numbers of homozygotes, we have one degree of freedom to estimate the level of agamospermy. In this case, the maximum likelihood estimate of  $c$  is:

$$= \hat{c} (a_1 - a_0)/n \quad (13)$$

with variance:

$$V_c = (1 - c^2)/n \quad (14)$$

and information per plant:

$$i_f = 1/(1 - c^2). \quad (15)$$

It will be noted that in this case the maximum likelihood estimate of  $c$  and its variance are independent of the gene frequency in the pollen pool. Thus, the use of this procedure avoids the problem of estimating  $p$ . Moreover, progenies of heterozygotes at co-dominant loci provide more information per plant than those of the more frequent homozygote. This can be seen by substituting  $p = 0.5$  into (3) to obtain (15); that is, the information per plant from heterozygotes and homozygous families are equal. Fig. 1 depicts (3) as an increasing function of  $p$ . Hence at low values of  $p$ , (15)

exceeds (3). It follows that the progeny of heterozygous mothers provide more information about  $c$  than those of the more frequent homozygote.

However, in many circumstances, one wishes to test the assumption that all genotypes contribute equally to the male gametophytic population, and an estimate of the gene frequency in the pollen pool is therefore required. In this situation, it is necessary to modify the above procedure slightly and score the numbers of two homozygotes appearing in the progenies of the parental heterozygotes separately. The expectations in this case are:

Class	Observed number	Expected number
Homozygote $B_1B_1$	$a_0$	$\frac{1}{2}tpn$
Homozygote $B_2B_2$	$a_1$	$\frac{1}{2}tqn$
Heterozygote $B_1B_2$	$a_2$	$\frac{(c + \frac{1}{2}t)n}{n}$
Total	$n$	$n$

The joint maximum likelihood estimates of  $c$  and  $p$  are:

$$\begin{aligned}\hat{c} &= 2a_2/n - 1 \\ \hat{p} &= a_0/(a_0 + a_1)\end{aligned}\quad (16)$$

and the variances of the estimates:

$$\begin{aligned}V_c &= (1 - c^2)/n \\ V_p &= 2p(1 - p)/nt.\end{aligned}\quad (17)$$

#### (f) *A plurality of segregating loci*

When individual plants of predominantly outcrossing species such as *Zea mays* (Brown and Allard, 1969), *Phalaris tuberosa* (McWilliam *et al.*, 1971 and unpublished) and *Arrhenantherum elatius* (Brown, unpublished) are subjected to electrophoresis and assayed for a series of non-specific enzymes such as esterases or peroxidases, the zymograms obtained usually display each individual in the population as genetically unique. In other words, sufficient loci are segregating in the population that inspection of half-sib arrays for genetically identical individuals can be undertaken without a knowledge of the genetical basis of the allozyme variants (Smith, 1972). This procedure is less rigorous than that based on single specific marker loci, but may be the only practical alternative where allozyme patterns are so complex as to preclude a Mendelian analysis. A similar procedure can be applied to progeny counts, where the classification into sexual or apomictic progeny depends upon a diversity of phenotypic comparisons (Pommer, 1972).

A multigenic model is also relevant to another situation; that is where many Mendelian genes are indeed scored separately but the estimates of apomixis thus obtained from each locus differ significantly. This would arise if one of the markers was non-randomly associated with a gene predisposing to apomixis (Harlan *et al.*, 1964). This problem is analogous to that encountered in the estimation of outcrossing (Harding and Tucker, 1964). It seemed appropriate therefore to develop a model to estimate the level of apomixis based on multigenic data.

There are two kinds of experimental cases; the first where the maternal genotype is assayed together with each half-sib progeny array, and the second where only the half-sibs are assayed and direct information on the

mother is lacking. In the first case where a total of  $n$  half-sibs are each compared with their maternal parent we have the following expectations.

	Observed number	Expected number
Parental—apomitic	$a_0$	$nc$
Non-parental—sexual	$a_1$	$n(1-c)$
Total	$n$	$n$

This is the standard binomial estimation problem and thus:

$$\hat{c} = a_0/n \tag{18}$$

and

$$V_c = c(1-c)/n. \tag{19}$$

However, when the comparisons which detect genetic identity must be made between the progeny themselves, at least two apomictic seeds must occur within a progeny array for the event to be detected. For this model let us suppose that  $r$  progeny are compared among themselves in each of  $m$  families. Then the families can be classified on the basis of how many apomictic seeds are detected:

Number of parentals	Number of families	
	Observed	Expected
0, 1	$a_1$	$m(1-c)^{r-1}[1+c(r-1)]$
2	$a_2$	$mK_{2,r} c^2(1-c)^{r-2}$
3	$a_3$	$mK_{3,r} c^3(1-c)^{r-3}$
:	:	:
$k$	$a_k$	$mK_{k,r} c^k(1-c)^{r-k}$
:	:	:
$r$	$a_r$	$mc^r$
Total	$m$	$m$

where  $K_{i,r} = r!/[i!(r-i)!]$ .

The maximum likelihood estimator is:

$$\hat{c} = \frac{A + \sqrt{A^2 + B}}{2r(r-1)} \tag{20}$$

where  $A = \sum_{i=2}^r [(r-1)i-r]a_i/m$  and  $B = 4r(r-1) \sum_{i=2}^r ia_i/m$ , and the variance of  $\hat{c}$  is given by:

$$V_c = \frac{ct(cr+t)}{mr(cr+t-t^{r-1})}. \tag{21}$$

(ii) *Species reproducing by mixed agamospermy, selfing and random outcrossing*

(a) *Dominant marker loci—recessive homozygotes and heterozygotes*

Consider a plant population polymorphic for a dominant diallelic locus in which there are constant probabilities  $t$ ,  $s$  and  $c$  of random mating, selfing and agamospermy, respectively. If a random sample of genotypes is drawn from the population and the numbers of individuals with recessive and dominant phenotypes scored in the progenies of the recessive homozygotes, we have the following expectations:



Class	Observed number	Expected number
Recessive homozygotes		
Recessives ( <i>bb</i> )	$a_0$	$n_1(c + s + tq)$
Dominants ( <i>Bb</i> )	$\frac{a_1}{n_1}$	$\frac{n_1 t p}{n_1}$
Total	$n_1$	$n_1$
Heterozygotes		
Recessives ( <i>bb</i> )	$a_2$	$n_2[s/4 + tq/2]$
Dominants ( <i>B-</i> )	$\frac{a_3}{n_2}$	$\frac{n_2[c + 3s/4 + t(1 + p)/2]}{n_2}$
Total	$n_2$	$n_2$

We have two degrees of freedom and two parameters (*s* and *c*) to estimate. Equating observations to their expectations yields:

$$\hat{s} = 2[\rho(2n_1a_2 + n_2a_1) - n_2a_1]/n_1n_2\rho \tag{22}$$

$$\hat{c} = [a_1n_2(1 - 2\rho) - n_1\rho(3a_2 - a_3)]/n_1n_2\rho$$

If  $\rho$  is known, the variances of the estimates are:

$$V_s = [n_1\rho x + 4n_2(1 - \rho)^2y]/n_1n_2\rho \tag{23}$$

$$V_c = [n_1\rho x + n_2(1 - 2\rho)^2y]/n_1n_2\rho$$

Where  $x = (s + 2tq)(4 - s - 2tq)$  and  $y = t(1 - \rho t)$

If  $\rho$  is estimated experimentally, the approximate variances for  $\hat{s}$  and  $\hat{c}$  are:

$$V_s \approx \left[ \frac{2(1 - \rho)}{n_1\rho} \right]^2 (V_{a_1}) + \left( \frac{4}{n_2} \right)^2 (V_{a_2}) + \left( \frac{2a_2}{n_1\rho^2} \right)^2 (V_p) \tag{24}$$

$$V_c = \left( \frac{1 - 2\rho}{n_1\rho} \right)^2 (V_{a_1}) + \left( \frac{4}{n_2} \right)^2 (V_{a_2}) + \left( \frac{a_1}{n_1\rho^2} \right)^2 (V_p)$$

Where  $V_{a_1}$ ,  $V_{a_2}$ , and  $V_p$  are the variances of  $a_1$ ,  $a_2$  and  $\hat{\rho}$  respectively.

(b) *Co-dominant marker loci—heterozygotes*

If we select a random sample of heterozygotes from a population polymorphic for a co-dominant diallelic locus and score the numbers of each genotype appearing in their progeny, the expectations are:

Class	Observed number	Expected number
Homozygote, $B_1B_1$	$a_0$	$n(s + 2tp)/4$
Heterozygote, $B_1B_2$	$a_1$	$n(1 + c)/2$
Homozygote, $B_2B_2$	$\frac{a_2}{n}$	$\frac{n(s + 2tq)/4}{n}$
Total	$n$	$n$

In this case, the maximum likelihood estimates of *s* and *c* are:

$$s = 4(a_2\rho - a_0q)/n(\rho - q) \tag{25}$$

$$c = 2a_1/n - 1.$$

It will be noted that if  $p = q$ ,  $s$  is undefined. This is due to the fact that if gene frequencies are equal, selfing and outcrossing have identical genetic consequences.

If  $p$  is known, the variances of the above estimates are given by:

$$V_s = [s(4-s)(1-2p)^2 + 8(1-c)p(1-p)]/n(p-q)^2$$

$$V_c = (1-c^2)/n$$
(26)

while if  $p$  is estimated experimentally, the variances are approximately:

$$V_s = \left[ \frac{4q}{n(p-q)} \right]^2 (V_{a_0}) + \left[ \frac{4q}{n(p-q)} \right]^2 (V_{a_2}) + \left[ \frac{4(a_0 - a_2)}{n(p-q)^2} \right]^2 (V_p)$$

$$V_c \approx 4V_{a_1}/n^2$$
(27)

(c) *Co-dominant markers—heterozygotes and homozygotes*

If a single random sample of plants is taken from a population polymorphic for a co-dominant diallelic locus and the numbers of genotypes in the progenies of three kinds of mothers is scored, then the expectations are:

Maternal	Progeny	Observed number	Expected number
$B_1B_1$	$B_1B_1$	$a_0$	$n_1(c + s + tq)$
	$B_1B_2$	$a_1$	$n_1tp$
$B_1B_2$	$B_1B_1$	$a_2$	$n_2(s + 2tp)/4$
	$B_1B_2$	$a_3$	$n_2(1 + c)/2$
	$B_2B_2$	$a_4$	$n_2(s + 2tq)/4$
$B_2B_2$	$B_1B_2$	$a_5$	$n_3tq$
	$B_2B_2$	$a_6$	$n_3(c + s + tp)$

In this general model  $n_1$ ,  $n_2$  and  $n_3$  are the total numbers of progeny scored from each of kind mother. There are four degrees of freedom and three parameters ( $p$ ,  $s$  and  $c$ ) to estimate. Maximum likelihood estimators cannot be formulated explicitly so that the so-called scoring method must be used. The scoring method is outlined in the Appendix I for this case.

### 3. DISCUSSION

The procedures described here permit the estimation of the level of agamospermy in facultative apomicts under a wide variety of circumstances. These procedures offer a number of advantages over the classical cytological methods. First, and most important, they permit the scoring of large numbers of individual plants (Tinney and Aamodt, 1940). Consequently, these procedures are ideal for the extensive surveys required to study the impact of environmental factors on the incidence of sexuality in facultative apomicts. Such studies have previously been restricted by limitations of cytological methods (Knox and Heslop-Harrison, 1963; Knox, 1967; Evans and Knox, 1969). Second, progeny test procedures, when used in conjunction with morphological marker loci, reduce the need for specialised equipment and expertise. As a result, these procedures are of special value for field stations and experiment farms where both equipment and expertise are often lacking. Finally, progeny tests provide a direct measure of the

relative frequency of apomictic seed. Such an estimate is of greater populational significance than the prezygotic estimate afforded by cytological techniques.

Nevertheless, progeny tests are not without their disadvantages. In fact, two major reservations need to be stated concerning the use of segregation patterns for marker genes among progeny as a measure of apomixis. First, this kind of data quantifies the proportions of progeny according to some preconceived model. Before a particular model can be used it must be established by independent tests that the model is valid. For example, the fact that a particular species can reproduce by agamospermy should be demonstrated by the examination of ovules. Tests should also be carried out to determine whether a species can self-fertilise and whether apomictic autosegregation occurs.

Second, progeny test procedures are relatively inefficient in detecting rare apomictic events. This arises from the following consideration of the models graphed in fig. 1. The curves show that the minimum value of  $i_j$ , the information per plant, for estimating the level of agamospermy obtains when  $p = 1/3$  (given that one could choose either homozygous or heterozygous families). At this point  $V_c = t(3-t)/n$ . Therefore the sample size required to demonstrate  $c > 0$  with 95 per cent certainty is:

$$n \approx 4(1-c)(2+c)/c^2$$

For low values of  $c$ , extremely large progeny sizes are required to demonstrate  $c > 0$ . For example, if  $c = 0.05$  approximately 3000 progeny must be scored to achieve this objective. In sharp contrast, the screening of  $4(1-c)/c = 76$  ovules cytologically would achieve the same end. Consequently, the use of progeny tests should be restricted to those species or populations where  $c > 0.20$ . For lower values of  $c$ , cytological procedures are to be preferred in most circumstances. It is interesting to note that to demonstrate  $t > 0$  with 95 per cent certainty when  $p = 1/3$  under the same model requires:

$$n \approx 4(3-t)/t.$$

When  $t = 0.05$  a progeny size of  $n = 236$  is required to demonstrate  $t > 0$ . It follows that in most situations progeny testing methods can detect rare outcrossing events more efficiently than rare apomictic events.

In all models, it has been assumed that the frequency of agamospermy is constant among all plants in a population. This assumption will be unrealistic in many species and this may affect the accuracy of estimates obtained. Fortunately, the validity of this assumption can be easily tested by estimating  $c$  for each individual family and testing the estimates for heterogeneity using standard procedures (Mather, 1957), and the appropriate adjustments made to the estimation procedure.

#### 4. REFERENCES

- ALLARD, R. W., AND WORKMAN, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution*, 17, 470-480.
- BAILEY, N. T. J. 1951. Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics*, 7, 268-274.
- BARLOW, B. A. 1958. Heteroploid twins and apomixis in *Casuarina nana*. *Aust. J. Bot.*, 6, 204-219.

- BASHAW, E. C., HOVIN, A. W., AND HOLT, E. C. 1970. Apomixis, its evolutionary significance and utilisation in plant breeding. *Proc. XI Intern. Grassl. Cong.*, 245-248.
- BROWN, A. H. D., AND ALLARD, R. W. 1969. Further isozyme differences among the inbred parents of a reciprocal recurrent selection population in maize. *Crop. Sci.*, 9, 643-644.
- ELANDT-JOHNSON, R. C. 1971. *Probability Models and Statistical Methods in Genetics*. Wiley, New York.
- EVANS, L. T., AND KNOX, R. B. 1969. Environmental control of reproduction in *Themeda australis*. *Aust. J. Bot.*, 17, 375-389.
- 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.
- GUSTAFSSON, A. 1947. Apomixis in higher plants. III. Biotype and species formation. *Lund. Univ. Arsskr.*, 43, 183-370.
- HARDING, J., AND TUCKER, C. L. 1964. Quantitative studies on mating systems. 1. Evidence for the non-randomness of outcrossing in *Phaseolus lunatus*. *Heredity*, 18, 369-381.
- HARDING, J., AND TUCKER, C. L. 1969. Quantitative studies on mating systems. III. Methods for the estimation of male gametophytic selective values and differential outcrossing rates. *Evolution*, 23, 85-95.
- HARLAN, J. R., BROOKS, M. H., BORGAONKAR, D. S., AND DE WET, J. M. J. 1964. Nature and inheritance of apomixis in *Bothriochloa* and *Dichanthium*. *Bot. Gaz.*, 125, 41-46.
- KENDALL, M. G., AND STUART, A. 1969. *The Advanced Theory of Statistics*. Vol. 1. *Distribution Theory*. Charles Griffin, London.
- KNOX, R. B. 1967. Apomixis: seasonal and population differences in a grass. *Science*, 157, 325-326.
- KNOX, R. B., AND HESLOP-HARRISON, J. 1963. Experimental control of aposporous apomixis in a grass of the *Andropogoneae*. *Bot. Notiser.*, 116, 127-141.
- MATHER, K. 1957. *The Measurement of Linkage in Heredity*. Methuen, London.
- MCWILLIAM, J. R., SCHROEDER, H. E., MARSHALL, D. R., AND ORAM, R. N. 1971. Genetic stability of Australian phalaris (*Phalaris tuberosa* L.) under domestication. *Aust. J. agric. Res.*, 22, 895-908.
- POMMER, G. 1972. Untersuchungen zur Züchtung der Wiesenripen, *Poa pratensis* L., unter besonderer Berücksichtigung der Art der Samenbildung. *Z. Pflanzenzüchtung*, 67, 279-304.
- SMITH, R. L. 1972. Sexual reproduction in *Panicum maximum* Jacq. *Crop Sci.*, 12, 624-626.
- TINNEY, F. W., AND AAMODT, G. S. 1940. The progeny test as a measure of the type of seed development in *Poa pratensis* L. *J. Hered.*, 31, 457-464.

## APPENDIX

*The scoring method for estimating outcrossing, apomixis and gene frequency in the general case*

This formulation follows that given by Elandt-Johnson (1971) in which the approximation procedure is based on the Taylor expansion. First, the

TABLE 2  
*Expectations and differentials in the general model for a co-dominant locus*

Class		<i>i</i>	Expectation ( $E_i$ )	Differentials		
Maternal	Progeny			$(\partial E_i / \partial X)$	$(\partial E_i / \partial Y)$	$(\partial E_i / \partial s)$
$B_1 B_1$	$B_1 B_1$	1	$n_1(1-X)$	$-n_1$	—	—
	$B_1 B_2$	2	$n_1 X$	$n_1$	—	—
$B_1 B_2$	$B_1 B_1$	3	$n_2(s+2X)/4$	$n_2/2$	—	$n_2/4$
	$B_1 B_2$	4	$n_2(2-s-X-Y)/2$	$-n_2/2$	$-n_2/2$	$-n_2/2$
	$B_2 B_2$	5	$n_2(s+2Y)/4$	—	$n_2/2$	$n_2/4$
$B_2 B_2$	$B_1 B_2$	6	$n_3 Y$	—	$n_3$	—
	$B_2 B_2$	7	$n_3(1-Y)$	—	$-n_3$	—

problem is simplified by substituting  $X = tp$  and  $Y = tq$ . We now require estimates of  $c$ ,  $X$  and  $Y$ . Table 2 gives the expectations and partial differentials with respect to each parameter for each data class.

The initial approximations are:

$$\begin{aligned} \hat{X}_1 &= a_1/n_1 & \hat{Y} &= a_5/n_3 \\ \hat{s}_1 &= 2[a_2 + a_4]/n_2 - (\hat{X} + \hat{Y}). \end{aligned} \tag{A.1}$$

The information matrix ( $I$ ) entries are:

$$\begin{aligned} I_{XX} &= \sum_{i=1}^7 (\partial E_i / \partial X)^2 / E_i & I_{XY} &= \sum_i (\partial E_i / \partial X)(\partial E_i / \partial Y) / E_i \\ I_{YY} &= \sum_i (\partial E_i / \partial Y)^2 / E_i & I_{Xs} &= \sum_i (\partial E_i / \partial X)(\partial E_i / \partial s) / E_i \\ I_{ss} &= \sum_i (\partial E_i / \partial s)^2 / E_i & I_{Ys} &= \sum_i (\partial E_i / \partial Y)(\partial E_i / \partial s) / E_i \end{aligned} \tag{A.2}$$

The vector of scores ( $S$ ) contains the entries:

$$\begin{aligned} S_X &= \sum_i (\partial E_i / \partial X) E_i \\ S_Y &= \sum_i (\partial E_i / \partial Y) / E_i \\ S_s &= \sum_i (\partial E_i / \partial s) / E_i \end{aligned} \tag{A.3}$$

The iterative approximation formulae are:

$$\begin{bmatrix} \hat{X}_{i+1} \\ \hat{Y}_{i+1} \\ \hat{s}_{i+1} \end{bmatrix} = \begin{bmatrix} \hat{X}_i \\ \hat{Y}_i \\ \hat{s}_i \end{bmatrix} + I^{-1} \cdot S \tag{A.4}$$

Successive estimates of  $X$ ,  $Y$  and  $s$  are obtained until a satisfactory level of approximation is reached; for example when the difference between two successive estimates of a parameter is less than 1 per cent of its value. The variance and covariances of these estimates are contained in  $I^{-1}$ .

To obtain estimates of  $t$  and  $p$  from  $X$  and  $Y$  we use the formulae:

$$\begin{aligned} t &= \hat{X} + \hat{Y} \\ \hat{p} &= Y / (\hat{X} + \hat{Y}) \\ V_t &= V_X + V_Y + 2C_{XY} \\ V_p &= [X^2 V_Y + Y^2 V_X - 2XY C_{XY}] / (X + Y)^4. \end{aligned}$$