Heredity (1979), **43** (1), 107-121

SIMULTANEOUS ESTIMATION OF MALE AND FEMALE LINKAGE FROM A SINGLE INTERCROSS FAMILY

M. J. LAWRENCE,* M. A. CORNISH*‡ and M. D. HAYWARD†

**Department of Genetics, University of Birmingham, Birmingham B15 2TT; †The Welsh Plant Breeding Station, Plas Gogerddan, Aberystwyth, SY23 3EB*

Received 8.ii.79

Summary

Intercross families in which at least three alleles are segregating at each of a pair of linked loci can be used to obtain simultaneous and independent estimates of both male and female recombination frequency. When only two alleles are segregating at one locus, it is still possible to estimate the frequency of recombination both on the male and the female side of the cross, but such estimates are no longer independent.

Numerical examples of each type of intercross family are given and the efficiencies of these families are compared both *inter se* and with those of the more familiar type of family in which only two alleles are segregating at each locus.

The circumstances in which it is worth obtaining such intercross families are discussed.

1. Introduction

Maximum likelihood methods for the detection and estimation of linkage in backcross and intercross families of animals and plants were first introduced some 40 years ago (Mather, 1938, 1951; Bailey, 1961). A number of different situations have been considered in terms of their effect on the efficiency of estimation, such as dominance and progeny testing of double heterozygotes so as to distinguish coupling from repulsion types (Mather, 1936). Hitherto, attention has been confined to cases where there are two alleles only at each of the linked loci.

The discovery that the individuals of many natural populations of animals and plants are frequently polymorphic with respect to many different proteins (see, for example, Ayala, 1976) has, however, also revealed that many of the genes concerned occur in more, sometimes very many more, than two allelic forms. We may add to these recent polymorphisms, the well-known and long-standing case of the genes which determine self-incompatibility in those species of flowering plants with homomorphic systems which are always multi-allelic (see de Nettancourt, 1977).

The purpose of this paper is to deal with the detection and estimation of linkage in intercross families where two or more alleles are segregating at each locus; to compare the efficiencies of these families in this respect; and to give a numerical example of each of the types of family we consider.

‡ Present address: School of Pharmacy and Pharmacology, University of Bath, Claverton Down, Bath BA2 7AY.

2. A COMPLETELY CLASSIFIED INTERCROSS

(i) *Theory*

Consider a cross between two individuals that are A_1B_1/A_2B_2 and A_3B_3/A_4B_4 where A_1, A_2, A_3 and A_4 are four alleles at the first locus and B_1, B_2, B_3 and B_4 are four at the second. Let p_f be the frequency of recombination on the female side and p_m be the corresponding frequency on the male side of the cross ($q_f = 1\text{-}p_f$ and $q_m = 1\text{-}p_m$). Then the expected composition of the progeny from a cross between these two individuals is as shown in table 1. If there is no dominance at either locus, each of the 16 zygotic

TABLE 1

The expected composition of progeny produced by crossing an A_1B_1/A_2B_2 *individual used as a female with an* A_3B_3/A_4B_4 *individual used as a male parent. The* n_{ij}*'s are the numbers observed for each genotype. The grand total,* $n_{..}$ *is written as* n *in the text*

♀/♂	A_3B_3 $q_m/2$	A_3B_4 $p_m/2$	A_4B_3 $p_m/2$	A_4B_4 $q_m/2$	Row totals
A_1B_1 $q_f/2$	A_1B_1/A_3B_3 $q_fq_m/4$ n_{11}	A_1B_1/A_3B_4 $q_fp_m/4$ n_{12}	A_1B_1/A_4B_3 $q_fp_m/4$ n_{13}	A_1B_1/A_4B_4 $q_fq_m/4$ n_{14}	$q_f/2$ $n_{1.}$
A_1B_2 $p_f/2$	A_1B_2/A_3B_3 $p_fq_m/4$ n_{21}	A_1B_2/A_3B_4 $p_fp_m/4$ n_{22}	A_1B_2/A_4B_3 $p_fp_m/4$ n_{23}	A_1B_2/A_4B_4 $p_fq_m/4$ n_{24}	$p_f/2$ $n_{2.}$
A_2B_1 $p_f/2$	A_2B_1/A_3B_3 $p_fq_m/4$ n_{31}	A_2B_1/A_3B_4 $p_fp_m/4$ n_{32}	A_2B_1/A_4B_3 $p_fp_m/4$ n_{33}	A_2B_1/A_4B_4 $p_fq_m/4$ n_{34}	$p_f/2$ $n_{3.}$
A_2B_2 $q_f/2$	A_2B_2/A_3B_3 $q_fq_m/4$ n_{41}	A_2B_2/A_3B_4 $q_fp_m/4$ n_{42}	A_2B_2/A_4B_3 $p_fp_m/4$ n_{43}	A_2B_2/A_4B_4 $p_fq_m/4$ n_{44}	$q_f/2$ $n_{4.}$
Column totals	$q_m/2$ $n_{.1}$	$p_m/2$ $n_{.2}$	$p_m/2$ $n_{.3}$	$q_m/2$ $n_{.4}$	1 $n_{..}$

genotypes can be recognised because in these circumstances there is a one-to-one correspondence between genotype and phenotype. Furthermore, provided that there is no viability disturbance at either locus, the frequency with which each of these 16 classes is expected to occur is obtained as the product of the appropriate gametic frequencies.

Examination of the row and column totals of table 1 shows that this fully classified intercross family is equivalent to two double backcross families in one. For this reason, it is possible to carry out simple tests on these totals to determine whether A_1 to A_2 and B_1 to B_2 are each as 1 : 1 (row totals); whether A_3 to A_4 and B_3 to B_4 are as 1 : 1 (column totals), and whether there is evidence of linkage between A and B on the female (row totals) or on the male side (column totals) in the same way as for backcross data. We note, furthermore, that these tests are orthogonal.

The test for linkage on the female side is:

$$\chi^2_{(1)} = (n_{1.} - n_{2.} - n_{3.} + n_{4.})^2/n \qquad (1)$$

where n is written for the $n_{..}$ of table 1; and the corresponding test on the male side is:

$$\chi^2_{(1)} = (n_{.1}-n_{.2}-n_{.3}+n_{.4})^2/n \tag{2}$$

In practice, however, it is more useful to partition the sum of these χ^2's to provide an overall or joint test for linkage and a test for heterogeneity between male and female linkage. The χ^2's corresponding to these tests are:

Joint test

$$\chi^2_{(1)} = [(n_{1.}+n_{.1})-(n_{2.}+n_{.2})-(n_{3.}+n_{.3})+(n_{4.}+n_{.4})]^2/2n \tag{3}$$

Heterogeneity

$$\begin{aligned}\chi^2_{(1)} &= [(n_{1.}-n_{.1})-(n_{2.}-n_{.2})-(n_{3.}-n_{.3})+(n_{4.}-n_{.4})]^2/2n \\ &= \text{Equ. (1)}+\text{Equ. (2)}-\text{Equ. (3)}\end{aligned} \tag{4}$$

Turning next to the estimation of linkage, since a fully classified intercross is equivalent to two backcross families in one, it follows that the maximum likelihood estimates of the recombination frequencies are:

$$\hat{p}_f = \frac{n_{2.}+n_{3.}}{n} \quad \text{and} \quad \hat{p}_m = \frac{n_{.2}+n_{.3}}{n} \tag{5}$$

and that the variances of these estimates are

$$V_{p_f} = p_f q_f/n \quad \text{and} \quad V_{p_m} = p_m q_m/n \tag{6}$$

respectively. The estimates $\hat{p}_f$ and $\hat{p}_m$ are independent because the cross information, $I_{p_f p_m} = 0$.

We have not felt it worthwhile to consider the effects of differential viability in any great detail because it is not obvious how this could be realistically specified in respect of four alleles at each locus. On the other hand, since it is possible to carry out simple tests on the segregation ratios, it is perhaps rather unlikely that we should remain unaware of a disturbance due to this cause were it to be present. Furthermore, if only one ratio is disturbed in either one or both parents, the detection and estimation procedures we have given require no amendment. In more complex cases it would be necessary to examine the data for guidance as to how best to specify the disturbance. Lastly, it is worth pointing out that in appropriate circumstances it might be possible to distinguish between differential viability at the gametic level, on the one hand, from that at the zygotic level, on the other hand in this type of family.

In general three alleles at each locus also give a complete classification of the progeny, provided that both parents are heterozygous at both loci.

(ii) *A numerical example*

Cornish, Hayward and Lawrence (1979) have shown that self-incompatibility in perennial ryegrass (*Lolium perenne*) is controlled by two multiallelic loci, S and Z, determination of the pollen phenotype being gametophytic. As is usual in gametophytic systems, there is no dominance of the incompatibility alleles in the stigma. One of the families, H, on which this conclusion was based is shown in table 2. The data in this table are

TABLE 2

Family H. Plants classified according to their incompatibility genotype

♀/♂	$S_3\mathcal{Z}_3$	$S_3\mathcal{Z}_4$	$S_4\mathcal{Z}_3$	$S_4\mathcal{Z}_4$	Row totals
$S_1\mathcal{Z}_1$	$S_1\mathcal{Z}_1/S_3\mathcal{Z}_3$ 4	$S_1\mathcal{Z}_1/S_3\mathcal{Z}_4$ 3	$S_1\mathcal{Z}_1/S_4\mathcal{Z}_3$ 5	$S_1\mathcal{Z}_1/S_4\mathcal{Z}_4$ 1	$n_{1.}$ 13
$S_1\mathcal{Z}_2$	$S_1\mathcal{Z}_2/S_3\mathcal{Z}_3$ 1	$S_1\mathcal{Z}_2/S_3\mathcal{Z}_4$ 0	$S_1\mathcal{Z}_2/S_4\mathcal{Z}_3$ 2	$S_1\mathcal{Z}_2/S_4\mathcal{Z}_4$ 0	$n_{2.}$ 3
$S_2\mathcal{Z}_1$	$S_2\mathcal{Z}_1/S_3\mathcal{Z}_3$ 3	$S_2\mathcal{Z}_1/S_3\mathcal{Z}_4$ 0	$S_2\mathcal{Z}_1/S_4\mathcal{Z}_3$ 1	$S_2\mathcal{Z}_1/S_4\mathcal{Z}_4$ 3	$n_{3.}$ 7
$S_2\mathcal{Z}_2$	$S_2\mathcal{Z}_2/S_3\mathcal{Z}_3$ 2	$S_2\mathcal{Z}_2/S_3\mathcal{Z}_4$ 3	$S_2\mathcal{Z}_2/S_4\mathcal{Z}_3$ 1	$S_2\mathcal{Z}_2/S_4\mathcal{Z}_4$ 2	n_4 8
Column totals	$n_{.1}$ 10	$n_{.2}$ 6	$n_{.3}$ 9	$n_{.4}$ 6	$n_{..}$ 31

arranged in the same way as table 1, it being assumed that the progeny arose from the cross $S_1\mathcal{Z}_1/S_2\mathcal{Z}_2(♀) \times S_3\mathcal{Z}_3/S_4\mathcal{Z}_4(♂)$. Three of the expected genotypic classes are empty, an outcome which is perhaps hardly surprising in view of the fact that only 31 plants were classified in this family.

The combined χ^2 analysis of the row and column totals of table 2 is shown in table 3. All four of the single factor ratios are in good agreement

TABLE 3

The χ^2 analysis of family H; the S, Z *data. The numbers in parenthesis refer to equations* 1-4 *on p.* 109

Item	$\chi^2_{(1)}$	P
$S_1 : S_2$	0·032	0·90—0·80
$S_3 : S_4$	0·032	0·90—0·80
$\mathcal{Z}_1 : \mathcal{Z}_2$	2·613	0·20—0·10
$\mathcal{Z}_3 : \mathcal{Z}_4$	1·581	0·30—0·20
♀ linkage	3·903 (1)	0·05—0·02*
♂ linkage	0·032 (2)	0·90—0·80
Joint	2·323 (3)	0·20—0·10
Heterogeneity	1·612 (4)	0·30—0·20

with the expected 1 : 1 segregation. With the exception of the $♀\chi^2_{(1)}$ the data are also homogeneous with respect to linkage. However, the test on the female arrays need not be taken very seriously, both because the χ^2 in question is only just significant and because we have not detected linkage between S and $\mathcal{Z}$ in any of the six other families that have been analysed. Ordinarily, therefore, the analysis would stop at this point. We proceed to estimate the male and female linkage parameters together with their variances purely by way of illustration.

From equations 5:

$$\hat{p}_f = 0{\cdot}3226 \quad \text{and} \quad \hat{p}_m = 0{\cdot}4839$$

From equations 6:

$$V_{\hat{p}_f} = 0{\cdot}007049 \quad \text{and} \quad V_{\hat{p}_m} = 0{\cdot}008056.$$

Thus the estimates and their standard errors are:

$$\hat{p}_f = 0{\cdot}3266 \pm 0{\cdot}0840 \quad \text{and} \quad \hat{p}_m = 0{\cdot}4839 \pm 0{\cdot}0898$$

Finally, since, on the evidence of the analysis of the data, the male and female recombination frequencies are homogeneous, we conclude this example by obtaining the joint estimate of linkage $\hat{p}$. In the present case, $\hat{p}$ is simply the average of $\hat{p}_f$ and $\hat{p}_m$, so that $\hat{p} = 0{\cdot}4033$. The amount of information about this estimate, I_p, is

$$I_p = I_{p_f p_f} + I_{p_m p_m} = 2n/pq \quad \text{since} \quad I_{p_f p_m} = 0.$$

Hence $I_p = 257{\cdot}6365$, $V_p = 0{\cdot}003881$ and $s_p = 0{\cdot}0623$. Thus the joint estimate is $\hat{p} = 0{\cdot}4033 \pm 0{\cdot}0623$.

3. An incompletely classified intercross

(i) *Theory*

We consider next an intercross family produced by crossing two individuals that are A_1B_1/A_2B_2 and A_1B_3/A_2B_4 respectively. This differs from the previous cross in that while there are still four alleles at one locus (B) there are only two at the other (A). Because of this reduction from four to two alleles at one of the loci, it is no longer possible to deduce whether an A_1 allele, say, in the progeny has descended from the male or the female parent. The consequence of this incomplete classification is that we can recognise only 12, rather than 16, genotypic classes in the progeny. In particular, it is no longer possible to distinguish the genotypes corresponding

Table 4

The expected composition of the progeny of a cross between an A_1B_1/A_2B_2 *individual used as a female and an* $A_1B_3A_2B_4$ *individual used as a male parent. All frequencies in the table should be divided by 4*

	B_1B_3	B_1B_4	B_2B_3	B_2B_4
A_1A_1	$A_1A_1B_1B_3$ $q_f q_m$ n_{11}	$A_1A_1B_1B_4$ $q_f p_m$ n_{12}	$A_1A_1B_2B_3$ $p_f q_m$ n_{13}	$A_1A_1B_2B_4$ $p_f p_m$ n_{14}
A_1A_2	$A_1A_2B_1B_3$ $p_f q_m + q_f p_m$ n_{21}	$A_1A_2B_1B_4$ $p_f p_m + q_f q_m$ n_{22}	$A_1A_2B_2B_3$ $p_f p_m + q_f q_m$ n_{23}	$A_1A_2B_2B_4$ $p_f q_m + q_f p_m$ n_{24}
A_2A_2	$A_2A_2B_1B_3$ $p_f p_m$ n_{31}	$A_2A_2B_1B_4$ $p_f q_m$ n_{32}	$A_2A_2B_2B_3$ $q_f p_m$ n_{33}	$A_2A_2B_2B_4$ $p_f q_m$ n_{34}

to the coupling and repulsion double heterozygotes of the classical F_2 family in this progeny, so that it is immediately apparent that the present design must be less efficient in respect of the estimation of recombination frequencies than the previous one.

There are two further general points worth making about this intercross family. Firstly, whereas the previous family was equivalent to a pair of double backcrosses, the present family is equivalent to a pair of single backcrosses. Thus while the expected frequencies in the first and third row of table 4 are clearly similar to those in table 1, the corresponding

entries in the middle row of this table resemble the frequencies of the classes in an F_2 family. It is possible, therefore, to estimate male and female recombination frequency from a family of this type using the first and third rows only. However, since it is clear that the frequencies of the classes in the middle row of the table are equal only if either $p_f = 0{\cdot}5$ or $p_m = 0{\cdot}5$, it is desirable that the information from the classes of the middle row should be used as well as that from the remaining classes.

The second point is that the row and column totals of the present table, in which the classes are necessarily arranged in a different way to the previous one (table 1), provide no information about linkage, though they may again be used to carry out independent tests on the segregation ratios at each locus. We shall consider the detection of linkage in a family of this type after we have dealt with the question of its estimation.

The logarithm of the likelihood, l, obtaining an observed family is:

$$\begin{aligned} L = {} & (n_{13}+n_{14}+n_{31}+n_{32}) \log p_f + (n_{11}+n_{12}+n_{33}+n_{34}) \log q_f \\ & + (n_{12}+n_{14}+n_{31}+n_{33}) \log p_m + (n_{11}+n_{13}+n_{32}+n_{34}) \log q_m \\ & + (n_{21}+n_{24}) \log (p_f q_m + q_f p_m) + (n_{22}+n_{23}) \log (p_f p_m + q_f q_m) \\ & + \text{constants.} \end{aligned} \tag{7}$$

Differentiating L in turn with respect to p_f and p_m and thereby obtaining the score, S_{p_f}, for each parameter we find

$$\left.\begin{aligned} S_{p_f} = {} & \frac{n_{13}+n_{14}+n_{31}+n_{32}}{p_f} - \frac{n_{11}+n_{12}+n_{33}+n_{34}}{q_f} \\ & + \frac{(p_m - q_m)}{\theta(1-\theta)}((n_{22}+n_{23})(1-\theta) - (n_{21}+n_{24})\theta) \\ \text{and} \\ S_{p_m} = {} & \frac{n_{12}+n_{14}+n_{31}+n_{33}}{p_m} - \frac{n_{11}+n_{13}+n_{32}+n_{34}}{q_m} \\ & + \frac{(p_f - q_f)}{\theta(1-\theta)}((n_{22}+n_{23})(1-\theta) - (n_{21}+n_{24})\theta) \end{aligned}\right\} \tag{8}$$

where $\theta = (p_f p_m + q_f q_m)$ and $1-\theta = (p_f q_m + q_f p_m)$.

It is worth pointing out that the first line of each of these equations involves terms from the first and third rows of table 4 only and thus is the contribution of the backcross portion of the data to the scores, the remaining part of each equation being the contribution of the F_2 portion of the family. Furthermore, since S_{p_f} is a function of p_m, as well as p_f and (vice versa), it is not possible in general to obtain independent estimates of these parameters from this type of family. It follows, therefore, that numerical solutions to these equations cannot be found directly as in the previous family, it being necessary to obtain these solutions by iteration. We note also that, as expected, the correlation between the estimates arises solely from the F_2 portion of the data.

The amounts of information about the estimates are obtained in the usual way as

$$-E\left(\frac{\partial^2 L}{\partial p_i \partial p_j}\right)$$

These equations are:

$$\left.\begin{aligned} I_{p_f p_f} &= \frac{n}{2}\left(\frac{1}{p_f q_f} + \frac{(p_m - q_m)^2}{\theta(1-\theta)}\right) \\ I_{p_m p_m} &= \frac{n}{2}\left(\frac{1}{p_m q_m} + \frac{(p_f - q_f)^2}{\theta(1-\theta)}\right) \end{aligned}\right\} \quad (9)$$

and

$$I_{p_f p_m} = \frac{n}{2}\left(\frac{2\theta - 1}{\theta(1-\theta)}\right)$$

Then if I is the information matrix, where

$$\boldsymbol{I} = \begin{bmatrix} I_{p_f p_f} & I_{p_f p_m} \\ I_{p_f p_m} & I_{p_m p_m} \end{bmatrix} \quad (10)$$

the variances of p_f and p_m are, as usual, the appropriate elements in the inverse of the information matrix, $\boldsymbol{I}^{-1}$ which is

$$\boldsymbol{I}^{-1} = \begin{bmatrix} I_{p_m p_m}/\Delta & -I_{p_f p_m}/\Delta \\ -I_{p_f p_m}/\Delta & I_{p_f p_f}/\Delta \end{bmatrix} \quad (11)$$

where

$$\Delta = I_{p_f p_f} I_{p_m p_m} - I^2_{p_f p_m}$$

Thus

$$V_{\hat{p}_f} = \left(I_{p_f p_f} - \frac{I^2_{p_f p_m}}{I_{p_m p_m}}\right)^{-1}$$

and

$$V_{\hat{p}_m} = \left(I_{p_m p_m} - \frac{I^2_{p_f p_m}}{I_{p_f p_f}}\right)^{-1}; \quad (12)$$

and the covariance of the estimates is:

$$W_{\hat{p}_f \hat{p}_m} = \left(I_{p_f p_m} - \frac{I_{p_f p_f} I_{p_m p_m}}{I_{p_f p_m}}\right)^{-1}$$

Turning now to the detection of linkage in a family of this type, inspection of the equation of estimation for p_f (equation 8) shows that when $p_m = 0{\cdot}5$, the third term on the right hand side equals zero and similarly for the third term of S_{p_m} when $p_f = 0{\cdot}5$. It follows, therefore, that when testing the null hypothesis of no linkage ($p_f = p_m = 0{\cdot}5$) the entries in the second row of table 4 contribute no information to the appropriate χ^2's. A similar reduction is also obtained in the amounts of information (equations 9) and $I_{p_f p_m} = 0$.

Writing:

$$C = n_{11} + n_{34}$$

$$D = n_{12} + n_{33}$$

$$E = n_{13} + n_{32}$$

$$F = n_{14} + n_{31}$$

when $p_f = p_m = 0{\cdot}5$

$$S_{p_f} = -2(C + D - E - F) \quad (13)$$

$$S_{p_m} = -2(C-D+E-F) \tag{14}$$

and

$$I_{p_f p_f} = I_{p_m p_m} = 2n$$

Hence the test for linkage on the female side of this cross is:

$$\chi^2_{(1)} = S^2_{p_f}/I_{p_f p_f} = [2(C+D-E-F)]^2/2n \tag{15}$$

and the corresponding test on the male side is:

$$\chi^2_{(1)} = S^2_{p_m}/I_{p_m p_m} = [2(C-D+E-F)]^2/2n \tag{16}$$

As in the case of the completely classified intercross, it is convenient to partition the sum of these χ^2's to obtain a joint or overall test for linkage and a test for heterogeneity between male and female linkage. The comparison for the joint test may be obtained as the sum of and that for the test of heterogeneity as the difference between the female and male linkage comparisons (equations 13 and 14). Thus the comparison for the joint test is:

$$\text{Equ. (13)} + \text{Equ. (14)} = S_p = -4(C-F); \tag{17}$$

and that for the heterogeneity is

$$\text{Equ. (13)} - \text{Equ. (14)} = -4(D-E). \tag{18}$$

Since

$$I_p = I_{p_f p_f} + I_{p_m p_m} = 4n,$$

$$\text{the Joint } \chi^2_{(1)} = [4(C-F)]^2/4n \tag{19}$$

and the

$$\text{Heterogeneity } \chi^2_{(1)} = [4(D-E)]^2/4n \tag{20}$$

Three further points deserve mention before we consider a numerical example. Firstly, equations 13 and 14, on the one hand, and equations 17 and 18, on the other, are two alternative sets of orthogonal comparisons which can be made between the totals C, D, E and F. Since there are four of the latter, it is clear that we have yet to account for the third degree of freedom. The comparison associated with this degree of freedom, which is orthogonal to each of the alternative sets, is (C−D−E+F) which is a measure of the departure from zero of the quantity $(p_f - q_f)(p_m - q_m)$. Though the expected value of this quantity is obviously zero on the null hypothesis, its value in other circumstances is not particularly informative.

Secondly, in practice the joint and heterogeneity χ^2's are of greater interest than the other pair, since if either is significant, the presence of linkage has been detected in the data. If the joint item alone is significant linkage is homogeneous on the male and female side of the cross and we would wish to obtain a joint estimate of linkage. Where, however, the heterogeneity item is significant, the intensity of linkage is not the same on each side of the cross and we would wish to obtain estimates of both p_f and p_m from the data. Indeed, the female, male and heterogeneity χ^2's given above are strictly valid only when the null hypothesis holds. In principle, if any one of the four tests is significant, it is necessary to recalculate these three χ^2's, though in practice it is worth doing so for the heterogeneity item only, which should be recalculated on the joint estimate

of linkage, $\hat{p}$ (see later). For these reasons we have calculated only the joint and heterogeneity items in the following numerical example.

Thirdly, as in the previous family, only three alleles are required at the second locus, *B*, to obtain the same classification, provided that each parent is heterozygous for this locus.

(ii) *A numerical example*

(a) *The detection of linkage*

All but one of the individuals of family H, together with their parents were also scored with respect to their phosphoglucoisomerase (PGI) phenotype. Though four alleles are known at this locus in ryegrass (Hayward and McAdam, 1977), both parents turned out to be *a b* heterozygotes for the PGI locus. Family H is thus an F_2 for this gene. Inspection of the combined isozyme and incompatibility data from this family suggested that the PGI locus might be linked to one of the incompatibility loci, *S*, and that the genotypes of the parents were *a* S_1/*b* S_2(♀) and *b* S_3/*a* S_4 (♂) respectively. The relevant results from this family, arranged in the same way as in table 4, are shown in table 5.

TABLE 5

Family H. Plant classified according to their genotype at the PGI-2 *and the* S *locus*

	S_1S_4	S_1S_3	S_2S_4	S_2S_3	Row totals
aa	aaS_1S_4	aaS_1S_3	aaS_2S_4	aaS_2S_3	
	6	1	0	0	7
ab	abS_1S_4	abS_1S_3	abS_2S_4	abS_2S_3	
	2	5	6	2	15
bb	bbS_1S_4	bbS_1S_3	bbS_2S_4	bbS_2S_3	
	0	2	0	6	8
Column totals	8	8	6	8	30

The combined χ^2 analysis of the data shown in table 5 is given in table 6. The single factor ratios $S_1 : S_2$, $S_3 : S_4$ and *aa* : *ab* : *bb* are in good agreement with the expected 1 : 1 and 1 : 2 : 1 ratios respectively. There is, however, very little doubt of the presence of linkage in these data, for the joint item in this analysis is highly significant. Furthermore, since the heterogeneity χ^2 is very small, the intensity of linkage between *S* and PGI-2 appears to be the same in each parent. However, for reasons given earlier, since we have been led to reject the null hypothesis of no linkage in these data, this item will have to be recalculated later, though the difference between this and the correct heterogeneity χ^2 is unlikely to be very great.

(b) *The estimation of linkage*

Since linkage appears to be homogeneous on each side of the cross, we proceed to obtain the joint estimate, $\hat{p}$ from the data which is accomplished by the process of iteration.

Now

$$S_p = S_{pf} + S_{pm}$$

$$= \frac{n_{12}+n_{13}+2n_{14}+2n_{31}+n_{32}+n_{33}}{p}$$

$$- \frac{2n_{11}+n_{12}+n_{13}+n_{32}+n_{33}+2n_{34}}{q}$$

$$+ \frac{2(p-q)}{\theta(1-\theta)}[(n_{22}+n_{23})(1-\theta)-(n_{21}+n_{24})\theta] \quad (21)$$

and

$$I_p = I_{p_f p_f} + I_{p_m p_m} + 2I_{p_f p_m} = n\left[\frac{1}{pq} + \frac{2(2\theta-1)}{\theta(1-\theta)}\right] \quad (22)$$

where $\theta = (p^2+q^2)$ and $(1-\theta) = 2pq$.
Using the backcross portion of the equation of estimation, S_p, to obtain an initial, trial value of $\hat{p}$ we find:

$$\hat{p} = \frac{D+E+2F}{2(C+D+E+F)} = \frac{1+2+0}{2(12+1+2+0)} = 0{\cdot}1$$

At this juncture in the procedure, the maximum likelihood estimate of p may be obtained either by using an appropriate computer library subroutine or manually. Since in most circumstances only two rounds of iteration will be required to obtain a satisfactory estimate we shall use the latter method. Inserting $\hat{p}_0 = 0{\cdot}1$ in equations 21 and 22 we find:

$$S_p = 14{\cdot}09214092 \quad \text{and} \quad I_p = 593{\cdot}49560163$$

Since S_p is positive, the trial value is smaller than the maximum likelihood estimate of p. Following Mather (1951, p. 134), a new trial value is found from:

$$\delta_p = S_p / I_p$$

where δ_p is the adjustment to the estimate of p we are seeking.
Thus

$$\delta_p = \frac{+14{\cdot}09214092}{593{\cdot}49560163} = +0{\cdot}02374430$$

and the new trial value of the linkage parameter is

$$\hat{p}_1 = 0{\cdot}10+0{\cdot}02374430 = 0{\cdot}12374430$$

Two further rounds of iteration yield a joint estimate

$$\hat{p} = 0{\cdot}12384976$$

The variance of this estimate is

$$V_{\hat{p}} = 1/I_p = 1/476{\cdot}7290 = 0{\cdot}0021$$

so that its standard deviation is $s_{\hat{p}} = 0{\cdot}0458$.
The joint linkage estimate together with its standard error is thus:

$$\hat{p} = 0{\cdot}1238 \pm 0{\cdot}0458$$

(c) *The correct heterogeneity* χ^2

Having rejected the null hypothesis of no linkage in these data, we need to complete this analysis by recalculating the heterogeneity χ^2 about the joint value, $\hat{p}$, rather than about $p = 0{\cdot}5$ as we did earlier. Following Bailey (1961, p. 279) this may be accompanied by calculating

$$\chi^2_{(1)} = \boldsymbol{S}'\boldsymbol{I}^{-1}\boldsymbol{S}$$

where $\boldsymbol{S}'$ is the row vector and $\boldsymbol{S}$ the corresponding column vector of the scores (equations 8); and $\boldsymbol{I}^{-1}$ is the inverse of the information matrix (equation 11). Both $\boldsymbol{S}$ and $\boldsymbol{I}^{-1}$ are calculated at $p_f = p_m = \hat{p} = 0{\cdot}12384976$. Inserting this value in equation 8 we find

$$S_{p_f} = +4{\cdot}60783132 \quad \text{and} \quad S_{p_m} = -4{\cdot}60782450$$

so that $\boldsymbol{S}' = [4{\cdot}60783132 \; -4{\cdot}60782450]$.
Similarly, substituting this value of $\hat{p}$ in equations 9 gives

$$I_{p_f p_f} = I_{p_m p_m} = 188{\cdot}19462902 \quad \text{and} \quad I_{p_f p_m} = 49{\cdot}95979248$$

so that

$$\boldsymbol{I} = \begin{bmatrix} 188{\cdot}19462902 & 49{\cdot}95979248 \\ 49{\cdot}95979248 & 188{\cdot}19462902 \end{bmatrix}$$

and

$$\boldsymbol{I}^{-1} = \begin{bmatrix} 0{\cdot}00571651 & -0{\cdot}00151756 \\ -0{\cdot}00151756 & 0{\cdot}00571651 \end{bmatrix}$$

Hence $\chi^2_{(1)} = \boldsymbol{S}'\boldsymbol{I}^{-1}\boldsymbol{S} = 0{\cdot}307$. As anticipated this χ^2 is a little bigger than the one that we calculated earlier on the assumption that the null hypothesis of no linkage held, the value of the latter being 0·133. Since, however, $P(\chi^2_{(1)} \geqq 0{\cdot}307) = 0{\cdot}70-0{\cdot}50$, we conclude, as before, that the intensity of linkage appears to be the same on each side of the cross.

TABLE 6

The χ^2 *analysis of family H; the* S, PGI-2 *data. The subtotals on which the linkage* χ^2*'s are calculated are:* $C = 12$, $D = 1$, $E = 2$ *and* $F = 0$. *The numbers in parenthesis refer to equations* (19) *and* (20) *on p.* 114

Item	d.f.	χ^2	P
$S_1 : S_2$	1	0·133	0·80—0·70
$S_3 : S_4$	1	0·133	0·80—0·70
aa : *ab* : *bb*	2	0·067	0·80—0·70
Joint	1	19·200 (19)	<0·001***
Heterogeneity	1	0·133 (20)	0·80—0·70

4. DISCUSSION

The chief and indeed most obvious advantage of the intercross families that we have discussed is that they allow the simultaneous estimation of male and female recombination frequencies from a single family. Hitherto, it has been necessary to raise two families for this purpose, such as a pair of reciprocal backcrosses. With organisms which are small and which have a short life-cycle and a high reproductive rate, the advantage of these intercross families over a pair of double backcross families may be insufficient to justify the extra labour in assembling a cross in which three or

four alleles are segregating at each locus. In other circumstances, however, particularly where the amount of effort that has to be expended in scoring phenotypes is considerable, as in the case of incompatibility, the advantage of being able to estimate male and female recombination frequencies from a single family may be decisive. In order to work out the most efficient procedure in any particular case we clearly need to calculate the amount of information that we could expect to obtain from the family in question with respect to the male and female linkage parameters.

Before, however, we turn to this matter it is convenient to compare the value of these intercross families with others that we might use in respect of the joint estimate of recombination frequency; that is, by assuming that $p_f = p_m = p$.

Now we saw earlier that for the completely classified intercross

$$I_p = \frac{2n}{pq}$$

and that for the incompletely classified intercross

$$I_p = n\left(\frac{1}{pq} + \frac{2(2\theta-1)}{\theta(1-\theta)}\right)$$

Following Mather (1936), it is convenient to compare the efficiencies of different types of family in terms of the amount of information yielded by a single individual of a family, i_p, where $i_p = I_p/n$.

Then for the completely classified intercross

$$i_p = \frac{2}{pq} = \frac{2}{p(1-p)}$$

and for the incompletely classified intercross

$$i_p = \left[\frac{1}{pq} + \frac{2(2\theta-1)}{\theta(1-\theta)}\right] = \frac{2(1-3p+3p^2)}{p(1-p)(1-2p+2p^2)}$$

Taking the amount of information given per individual of a backcross progeny as a standard, where $i_p = 1/p(1-p)$, the relative value of a completely classified intercross is 2; and that for the incompletely classified intercross is

$$\frac{2(1-3p+3p^2)}{1-2p+2p^2}.$$

In terms of the joint estimate of linkage, therefore, the completely classified intercross has the same value as a completely classified F_2; and the incompletely classified intercross has the same value as Mather's incompletely classified F_2. In principle, therefore, these new families are no more efficient than the others. In practice, however, the completely classified intercross is in fact more efficient than the completely classified F_2 because whereas in the latter it is necessary to progeny test the class of double heterozygotes in order to establish whether they are of the coupling or repulsion type, in the former it is not.

The full value of these intercross families is not realised, of course, until we consider their efficiencies in respect of the simultaneous estimation of male and female linkage. Thus in the first place it is unlikely in practice that the frequency of recombination in the male is the same as that in the

female, for there is an increasing amount of evidence which suggests that the former is in many species lower than the latter (Callan and Perry, 1977). Secondly, no other type of family is capable of yielding unique estimates of these parameters. Thus while it is possible to obtain two estimates of recombination frequency from a completely classified F_2 family, it is not at the same time possible to recognise which of these refers to the male and which to the female parent. In short, these estimates are completely interchangeable and the value of this family in this respect is therefore zero.

The minimum requirement that has to be fulfilled if unique estimates of the male and female linkage parameters are to be obtained from one and the same family is that it must be at least partly possible to trace the ancestry of the alleles in the progeny. As we have seen, this cannot be done with less than three alleles at one locus and the classification of ancestry is not complete until there are at least three alleles at each locus, with both parents being heterozygous in this respect.

For these reasons there are only two comparisons that we can usefully make between the various types of family. The first of these is a comparison between the completely classified intercross and a backcross. The amount of information yielded by an individual of a completely classified intercross in respect of the frequency of recombination in the female parent is $i_{p_f} = 1/p_f q_f$, which is the same, of course, as that yielded by an individual of a backcross family when the female parent is the double heterozygote. Since, however, we also obtain an estimate of the frequency of recombination in the male parent from a completely classified intercross family, it follows that this family is twice as valuable as a backcross, for in the latter, the cross has to be made in reciprocal if estimates of both linkage parameters are to be obtained. A completely classified intercross has thus twice the value of a backcross in respect of both the joint estimate as well as the individual estimates of male and female recombination frequency—a consequence of the fact that in the former, the estimates of p_f and p_m are independent.

The second comparison of interest concerns the two intercross families only. Now the variance of the estimate of p_f in an incompletely classified intercross is (from equation 12)

$$V_{p_f} = \left(I_{p_f p_f} - \frac{I^2_{p_f p_m}}{I_{p_m p_m}} \right)^{-1}$$

so that, in this case,

$$i_{p_f} = \left(I_{p_f p_f} - \frac{I^2_{p_f p_m}}{I_{p_m p_m}} \right) / n$$

this being the amount of information per individual analogous to Mather's i (note that this is *not* $I_{p_f p_f}/n$ here). Hence the efficiency of an incompletely classified intercross relative to a completely classified intercross in respect of p_f is:

$$e = 100 . \left(I_{p_f p_f} - \frac{I^2_{p_f p_m}}{I_{p_m p_m}} \right) . p_f q_f / n$$

$$= \frac{V_{p_f} \text{ for completely classified intercross} \times 100}{V_{p_f} \text{ for incompletely classified intercross}}$$

Since i_{p_f} for an incompletely classified intercross is a function of p_m as well as p_f, it is necessary to consider a range of values for the former as well as for the latter when considering the effects of the intensity and phase of linkage on the value of this type of family, as has been done in table 7.

TABLE 7

The value of an incompletely classified intercross relative to that of a completely classified intercross in respect of the precision of an estimate of p_f. *Entries in the table are efficiencies in per cent. Values of* p_f *and* p_m *greater than 0·5 refer to linkage in the repulsion phase*

		p_m								
		0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
p_f	0·1	**64·0**	55·5	52·0	50·5	50·0	50·5	52·0	55·5	64·0
	0·2	**72·8**	**60·5**	54·1	51·0	50·0	51·0	54·1	60·5	72·8
	0·3	78·1	**64·5**	**56·1**	51·5	50·0	51·5	56·1	64·5	78·1
	0·4	81·1	67·1	**57·5**	**51·9**	50·0	51·9	57·5	67·1	81·1
	0·5	82·0	68·0	58·0	**52·0**	**50·0**	52·0	58·0	68·0	82·0
	0·6	81·1	67·1	57·5	51·9	**50·0**	**51·9**	57·5	67·1	81·1
	0·7	78·1	64·5	56·1	51·5	50·0	**51·5**	**56·1**	64·5	78·1
	0·8	72·8	60·5	54·1	51·0	50·0	51·0	**54·1**	**60·5**	72·8
	0·9	64·0	55·5	52·0	50·5	50·0	50·5	52·0	**55·5**	**64·0**

We note first that in this table no entry is greater than 82·0 per cent; that is, even in the most favourable case we should need to score 122 individuals in an incompletely classified intercross to obtain an estimate of p_f which had the same precision as one from 100 individuals of a completely classified intercross. But we also note that the highest values in table 7 occur for rather unlikely combinations of p_m and p_f. Assuming, therefore, that, while p_m and p_f may not in practice be identical, they are nevertheless likely to be similar (with $p_f > p_m$), we see that at best, incomplete classification is only 72·8 per cent as efficient as complete classification in these circumstances ($p_f = 0{\cdot}2$; $p_m = 0{\cdot}1$); that is, we would require 137 individuals of the former to match the precision of an estimate from 100 individuals of the latter type of family. In general, efficiencies are highest with tight linkage, either in the coupling or repulsion phase, and lowest with loose linkage, which is, of course, a characteristic which the present family shares with Mather's incompletely classified F_2 family. Lastly, it is worth pointing out that, at worst, an incompletely classified intercross is as efficient as a backcross ($e = 50$ per cent) or, to put this the other way round, the estimation of male and female recombination frequencies are more economically obtained from an intercross family, even though in the case of incomplete classification, these estimates are not, of course, independent.

There is little doubt, therefore, that where three or more alleles are available at two or more linked loci, the gain in the precision of the experiment in respect of the estimation of male and female recombination frequency compared with the conventional case of only two alleles per locus can be considerable.

Acknowledgments.—We are greatly indebted to Dr J. S. Gale and especially to Professor Sir Kenneth Mather for much useful discussion and critical comment on the problems of the detection and estimation of linkage. We wish to acknowledge the receipt of a Science Research Council CASE award which enabled the work described in this paper to be carried out.

5. References

AYALA, F. A. ed. 1976. *Molecular Evolution.* Sinauer, Sunderland, U.S.A.

BAILEY, N. T. J. 1961. *Introduction to the Mathematical Theory of Genetic Linkage.* Clarendon Press, Oxford.

CALLAN, H. G., AND PERRY, P. E. 1977. Recombination in male and female meiocytes contrasted. *Phil. Trans. R. Soc. Lond. B. 277*, 227-233.

CORNISH, M. A., HAYWARD, M. D., AND LAWRENCE, M. J. 1979. Self-incompatibility in rye-grass. I. Genetic control in diploid *Lolium perenne* L. *Heredity 43*, 95-106.

HAYWARD, M. D., AND McADAM, N. J. 1977. Isozyme polymorphism as a measure of distinctiveness and stability in cultivars of *Lolium perenne*. *Z. Pflanzenzuchtg., 79*, 59-68.

MATHER, K. 1936. Types of linkage data and their value. *Ann. of Eugenics, 8*, 251-264.

MATHER, K. 1938, 1951. *The Measurement of Linkage in Heredity.* Methuen, London.

DE NETTANCOURT, D. 1977. *Incompatibility in Angiosperms.* Springer-Verlag, Berlin.