DISTURBED LINKAGE. I. PARTIAL MANIFESTATION AND DIFFERENTIAL VIABILITY AT THE SAME LOCUS IN BACKCROSS DATA

E. LAING

Department of Botany, University of Ghana, Legon, Accra, Ghana

1. INTRODUCTION

PREVIOUS papers examining the effects of partial manifestation or differential viability on the estimation of linkage values have dealt with these two disturbances separately (Fisher and Balmukand, 1928; Fisher, 1939, 1949; Bailey, 1949, 1950, 1961; Sanchez-Monge, 1952; Parsons, 1957). Bailey (1950), however, points out that linkage is estimable from the manifesting classes only when both disturbances operate at the same locus. Work with the skeletal factor, *Lp*, *Loop-tail*, in the house mouse suggested the possibility of the two disturbances affecting segregation at *Lp*/+ and hence recombination with the linked factors *ln*, *leaden*, and *py*, *polydactyly*, at least in a few families (Parsons, 1957; Laing, 1958). This study derives estimates of the recombination fraction between two loci, the manifestation or penetrance and the viability at the appropriate locus, and the variances of these estimates, when differential viability and partial manifestation affect one of the loci.

2. SYMBOLISM

Let A/a and B/b, at which there is complete dominance, be the two linked loci; let the recombination fraction between them be y. Whether the double heterozygote is in coupling (AB/ab) or in repulsion (Ab/aB), we will let a, b, c and d represent the observed numbers of progeny of phenotypes AB, Ab, aB and ab, respectively, the total being n.

μ represents the viability of the A phenotype as compared with that of the a, which is taken to be unity. If μ is less than one, we say A is imperfectly viable, if it is greater than one, then a is imperfectly viable. Occasionally the mutant phenotype is more viable than the wild type; usually, however, it is less viable. Throughout this study, μ will refer to A.

λ represents the proportion of one phenotype that fails to manifest, and so is misclassified as being of the other phenotype. We will use the symbolism of Bailey (1950), who denotes the direction of misclassification by an arrow. Thus $A \rightarrow a$ signifies the misclassification of A as a. When necessary, we can distinguish the two types of misclassification by indicating the direction: thus $\lambda(A \rightarrow a)$ may stand for the misclassification of A as a, $\lambda(a \rightarrow A)$, the reverse; λ here symbolically represents misclassification, as compared with μ, which symbolically represents viability disturbance. These symbols are useful in showing the sequence in time when the disturbances operate. We distinguish:

A   1

(μ, λ), the case where the viability effect operates before the misclassification,

and (λ, μ), where the misclassification takes place first. The estimates of some parameters differ with the sequence of action of the disturbances.

3. Coupling backcross data, viability disturbance preceding misclassification

We consider here the estimation of recombination between A/a and B/b from coupling data, when viability and manifestation at A/a are disturbed in the stated order, and A is misclassified as a with frequency λ: we symbolise this case as

$$(\mu, \lambda),\ A \to a.$$

3.1. *Derivation of estimates*

Table 1 summarises the observed numbers and the proportion of the total progeny expected in each phenotypic class.

Table 1

Expected proportions and observed numbers of individuals of various phenotypes from a coupling backcross in which viability disturbance precedes misclassification

Phenotype	Observed number	Expected proportion
AB	a	$\mu(1-\lambda)(1-y)$
Ab	b	$\mu(1-\lambda)y$
aB	c	$y+\lambda\mu(1-y)$
ab	d	$1-y+\lambda\mu y$
Total	n	$1+\mu$

The estimation is done by the maximum likelihood method (Fisher, 1922, 1954). We notice that there are three degrees of freedom available for estimating the three parameters; in such a case, where the number of degrees of freedom equals that of parameters to be estimated, the solutions to the maximum likelihood estimation equations may simply be obtained by equating observed to expected in each class (Fisher, 1922; Bailey, 1950). Thus we have:

$$\frac{a}{n} = \frac{\mu(1-\lambda)(1-y)}{1+\mu} \tag{1}$$

$$\frac{b}{n} = \frac{\mu(1-\lambda)y}{1+\mu} \tag{2}$$

$$\frac{c}{n} = \frac{y+\lambda\mu(1-y)}{1+\mu} \tag{3}$$

$$\frac{d}{n} = \frac{1-y+\lambda\mu y}{1+\mu}. \tag{4}$$

(1) and (2) give:

$$\frac{b}{a} = \frac{y}{1-y}$$

$$y = \frac{b}{a+b} \tag{5}$$

that is, the recombination fraction is simply estimated as the proportion of recombinants among the manifesting progeny.

Addition of (1) and (3) gives:

$$\frac{a+c}{n} = \frac{y+\mu(1-y)}{1+\mu} \tag{6}$$

which, after substitution for y by means of (5), and rearrangement, yields:

$$\mu = \frac{a(a+c)-b(b+d)}{ad-bc}, \tag{7}$$

(2), (5) and (7) give:

$$1-\lambda = \frac{a^2-b^2}{a(a+c)-b(b+d)} \tag{8}$$

$$\lambda = \frac{ac-bd}{a(a+c)-b(b+d)} \tag{9}$$

3.2. *Precision of the estimates*

This is found by means of the general large sample variance formula (Fisher, 1954):

$$V(T) = \sum_i a_i \left(\frac{\partial T}{\partial a_i}\right)^2 - n\left(\frac{\partial T}{\partial n}\right)^2. \tag{10}$$

T is an expression in terms of the observed numbers a_i and their total n. The derivation of the variance of a complex expression may be simplified by finding, first, the variance of its logarithm, and then deducing that of the original expression from this (Fisher, 1953). If we know the variance $V(P)$ of P, with corresponding information $I(P)$ and logarithmic likelihood L, we can find $V(p)$, the variance of p, which is a function of P, by the relations:

$$\frac{1}{V(P)} = I(P) = -\frac{d^2L}{dP^2}$$

$$V(p) = I^{-1}(p) = -\left(\frac{d^2L}{dp^2}\right)^{-1} = -\left(\frac{d^2L}{dP^2}\cdot\frac{dP^2}{dp^2}\right)^{-1} = \left(\frac{dp}{dP}\right)^2 V(P). \tag{11}$$

$V(y)$ may be found thus:
from (5)

$$\frac{\partial y}{\partial a} = -\frac{b}{(a+b)^2}, \quad \frac{\partial y}{\partial b} = \frac{a}{(a+b)^2}, \quad \frac{\partial y}{\partial c} = \frac{\partial y}{\partial d} = \frac{\partial y}{\partial n} = 0. \tag{12}$$

The substitution in (10) of these differentials gives:

$$V(y) = \frac{ab}{(a+b)^3} = \frac{y(1-y)}{(a+b)}. \tag{13}$$

We may distinguish

$$V_r(y) = \frac{ab}{(a+b)^3} \tag{14}$$

the variance realised, and

$$V_e(y) = \frac{y(1-y)(1+\mu)}{n\mu(1-\lambda)} \tag{15}$$

the variance expected, which is derived by substituting the expected numbers for the observed ones in (14). We notice that for a given value of μ, $V_e(y)$ increases with increase in value of λ:

$$\frac{\partial V_e(y)}{\partial \lambda} = \frac{y(1-y)(1+\mu)}{n\mu(1-\lambda)^2}. \tag{16}$$

For a given value of λ, $V_e(y)$ decreases with increase in value of μ:

$$\frac{\partial V_e(y)}{\partial \mu} = -\frac{y(1-y)}{n\mu^2(1-\lambda)}. \tag{17}$$

The theoretically possible case where

$$\lambda\mu = -1$$

would give

$$V_e(y) = \frac{y(1-y)}{n}$$

but this is genetically trivial since it would demand a negative expectation of aB individuals. We also notice that $V_e(y)$ increases with y up to $y = \frac{1}{2}$, when it is a maximum: the closer the linkage the higher the precision of measurement of y.

Let

$$M = \log \quad \mu = \log\left(a(a+c) - b(b+d)\right) - \log\left(ad - bc\right). \tag{18}$$

Then,

$$
\left.
\begin{aligned}
\sqrt{a'} &= \frac{\partial M}{\partial a} = \frac{2a+c}{a(a+c) - b(b+d)} - \frac{d}{ad - bc} \\[2mm]
\sqrt{b'} &= \frac{\partial M}{\partial b} = \frac{-2b-d}{a(a+c) - b(b+d)} + \frac{c}{ad - bc} \\[2mm]
\sqrt{c'} &= \frac{\partial M}{\partial c} = \frac{a}{a(a+c) - b(b+d)} + \frac{b}{ad - bc} \\[2mm]
\sqrt{d'} &= \frac{\partial M}{\partial d} = \frac{-b}{a(a+c) - b(b+d)} - \frac{a}{ad - bc} \\[2mm]
\frac{\partial M}{\partial n} &= 0.
\end{aligned}
\right\} \tag{19}
$$

By (10),

$$V(M) = aa' + bb' + cc' + dd'. \tag{20}$$

From (18),

$$\frac{\partial M}{\partial \mu} = \frac{1}{\mu}. \tag{21}$$

By (11)
$$V(\mu) = \left(\frac{\partial\mu}{\partial M}\right)^2 V(M)$$
$$= \mu^2 V(M)$$
$$= \mu^2(aa' + bb' + cc' + dd'). \tag{22}$$

$$\lambda = \frac{ac - bd}{a(a+c) - b(b+d)} \tag{23}$$

Let
$$L = \log\lambda = \log(ac - bd) - \log(a(a+c) - b(b+d)). \tag{24}$$

Then

$$\sqrt{a''} = \frac{\partial L}{\partial a} = \frac{c}{ac - bd} - \frac{2a+c}{a(a+c) - b(b+d)}$$

$$\sqrt{b''} = \frac{\partial L}{\partial b} = \frac{-d}{ac - bd} + \frac{2b+d}{a(a+c) - b(b+d)}$$

$$\sqrt{c''} = \frac{\partial L}{\partial c} = \frac{a}{ac - bd} - \frac{a}{a(a+c) - b(b+d)} \tag{25}$$

$$\sqrt{d''} = \frac{\partial L}{\partial d} = \frac{-b}{ac - bd} + \frac{b}{a(a+c) - b(b+d)}$$

$$\frac{\partial L}{\partial n} = 0.$$

By (10),
$$V(L) = aa'' + bb'' + cc'' + dd''. \tag{26}$$

By (11), (24) and (26),
$$V(\lambda) = \left(\frac{\partial\lambda}{\partial L}\right)^2 V(L)$$
$$= \lambda^2(aa'' + bb'' + cc'' + dd''). \tag{27}$$

Alternatively:
Let
$$z = \frac{\lambda}{1-\lambda} = \frac{ac - bd}{a^2 - b^2}, \tag{23a}$$

$$Z = \log z = \log(ac - bd) - \log(a^2 - b^2); \tag{24a}$$

then

$$\sqrt{a''} = \frac{\partial Z}{\partial a} = \frac{c}{ac - bd} - \frac{2a}{a^2 - b^2}$$

$$\sqrt{b''} = \frac{\partial Z}{\partial b} = \frac{2b}{a^2 - b^2} - \frac{d}{ac - bd}$$

$$\sqrt{c''} = \frac{\partial Z}{\partial c} = \frac{a}{ac - bd} \tag{25a}$$

$$\sqrt{d''} = \frac{\partial Z}{\partial d} = \frac{-b}{ac - bd}$$

$$\frac{\partial Z}{\partial n} = 0.$$

A 2

By (10),

$$V(\mathcal{Z}) = aa'' + bb'' + cc'' + dd''. \tag{26a}$$

By (11), (23a) and (26a),

$$V(\lambda) = \frac{z^2}{(1+z)^4} V(\mathcal{Z})$$

$$= \lambda^2(1-\lambda)^2(aa'' + bb'' + cc'' + dd''). \tag{27a}$$

3.3 *Tests of significance*

A test for linkage is impossible since, for $y = \frac{1}{2}$, the expected numbers corresponding to

$$a \qquad\qquad b \qquad\qquad c \qquad\qquad d$$

are

$$\frac{n\mu(1-\lambda)}{2(1+\mu)}, \quad \frac{n\mu(1-\lambda)}{2(1+\mu)}, \quad \frac{n(1+\lambda\mu)}{2(1+\mu)}, \quad \frac{n(1+\lambda\mu)}{2(1+\mu)} \tag{28}$$

which do not yield separate estimates for λ and μ. As Bailey (1950) points out, we can estimate λ and μ only when there is linkage. Effectively there is only 1 d.f., that for estimating

$$\frac{\mu(1-\lambda)}{1+\lambda\mu}.$$

To test the significance of the departure of λ from the value 0 which it has when there is perfect manifestation admitting linkage and viability disturbance, we compare the observed numbers with their expectations

$$\frac{(a+b)(a+d)}{n}, \quad \frac{(a+b)(b+c)}{n}, \quad \frac{(b+c)(c+d)}{n}, \quad \frac{(a+d)(c+d)}{n}. \tag{29}$$

These are obtainable from the maximum likelihood estimates

$$\mu = \frac{a+b}{c+d}, \tag{30}$$

$$y = \frac{b+c}{n},$$

which are appropriate to linkage with a viability disturbance at the locus A/a only.

The test of significance is

$$\chi^2 = \frac{(bd-ac)^2 n}{(a+b)(c+d)(a+d)(b+c)}, \tag{31}$$

with one degree of freedom, two degrees of freedom out of the available three being taken up by the estimation of y and μ. This χ_1^2 obviously tests the dependence between classification as single dominant *versus* double dominant and manifesting *versus* non-manifesting phenotype.

The test of significance for the presence of a viability disturbance, $\mu \neq 1$, admitting linkage and partial manifestation is similarly conducted. The expectations corresponding to a, b, c, and d are, respectively,

$$\frac{an}{2(a+c)}, \quad \frac{bn}{2(b+d)}, \quad \frac{cn}{2(a+c)}, \quad \frac{dn}{2(b+d)} \tag{32}$$

which are derivable from the maximum likelihood estimates:

$$y = \frac{b(a+c)}{a(b+d)+b(a+c)},$$

$$\lambda = \frac{cd-ab}{(a+c)(b+d)}, \tag{33}$$

appropriate to linkage with a partially manifesting factor. The comparison of expected with observed yields:

$$\chi_1^2 = \frac{1}{n}\{(a+c)-(b+d)\}^2.$$

4. COUPLING BACKCROSS DATA, MISCLASSIFICATION PRECEDING VIABILITY DISTURBANCE

We symbolise this case as (λ, μ), $A \to a$.

This case differs from what we have just considered only when the viability of the non-manifesting individual is that characteristic of its phenotype and not of its genotype. If the A/a individual of phenotype a has the viability μ of the A phenotype, then the problem is the same as that treated above.

Table 2 summarises the expected proportions and observed numbers for the various classes.

TABLE 2

Expected proportions and observed numbers of individuals of various phenotypes from coupling backcross in which misclassification precedes viability disturbance

Phenotype	Observed number	Expected proportion
AB	a	$\mu(1-\lambda)(1-y)$
Ab	b	$\mu(1-\lambda)y$
aB	c	$y+\lambda(1-y)$
ab	d	$1-y+\lambda y$
Total	n	$1+\lambda+\mu(1-\lambda)$

As before, the number of degrees of freedom available for estimation is that of parameters to be estimated. Equating observed to expected yields the following estimates:

$$y = \frac{b}{a+b}, \tag{1}$$

$$\lambda = \frac{ac-bd}{ad-bc}, \tag{2}$$

$$\mu = \frac{a-b}{d-c}. \tag{3}$$

We notice that whereas the estimates of λ and μ are now different, y is estimable by the same expression as before (3 : 5).

The precisions of the estimates of λ and μ are easily determined by the help of the large sample variance formula.

The variances are:

$$V(\lambda) \;=\; \frac{(a+b)(c+d)}{(ad-bc)^4}\,\{cd(a+b)(a-b)^2+ab(c+d)(c-d)^2\}, \tag{4}$$

$$V(\mu) \;=\; \frac{(a-b)^2(c+d)+(a+b)(c-d)^2}{(c-d)^4}. \tag{5}$$

Tests of significance are the same as those given above, since the expectations when $\mu = 1$ and when $\lambda = 0$ are the same in this case as in the other.

5. Results for backcross data

Table 3 summarises the estimates of y, λ and μ derived from coupling backcross data for the possible cases of operation of viability disturbance and partial manifestation at the same linked locus. The description of each case gives the order of action of the disturbances and the direction of the misclassification. a, b, c and d always refer to the phenotypes AB, Ab, aB and ab respectively. For repulsion data corresponding to the coupling ones, the formulae are the same except that the expressions that estimate y for coupling now estimate $1-y$.

Table 3

Estimates of y, λ *and* μ *for backcross data*

| | | Estimate of | |
Case	y	λ	μ
1. (μ, λ), $A{\to}a$	$\dfrac{b}{a+b}$	$\dfrac{ac-bd}{a(a+c)-b(b+d)}$	$\dfrac{a(a+c)-b(b+d)}{ad-bc}$
2. (μ, λ), $a{\to}A$	$\dfrac{c}{c+d}$	$\dfrac{ac-bd}{c(a+c)-d(b+d)}$	$\dfrac{ad-bc}{d(b+d)-c(a+c)}$
3. (λ, μ), $A{\to}a$	$\dfrac{b}{a+b}$	$\dfrac{ac-bd}{ad-bc}$	$\dfrac{a-b}{d-c}$
4. (λ, μ), $a{\to}A$	$\dfrac{c}{c+d}$	$\dfrac{bd-ac}{ad-bc}$	$\dfrac{a-b}{d-c}$

The estimates of y are of the form

$$\frac{u}{u+v}$$

and their variances are of the form:

$$V(y) \;=\; \frac{uv}{(u+v)^3} = \frac{y(1-y)}{u+v}. \tag{1}$$

$V(\mu)$ for case 1 is given by formulae (3 : 18-22). For case 2,

$$\mu \;=\; \frac{ad-bc}{d(b+d)-c(a+c)}, \tag{2}$$

$$M = \log \mu = \log (ad - bc) - \log (d(b+d) - c(a+c)). \tag{3}$$

$$\left.\begin{aligned}
\sqrt{a'} &= \frac{\partial M}{\partial a} = \frac{d}{ad-bc} + \frac{c}{d(b+d)-c(a+c)} \\[2mm]
\sqrt{b'} &= \frac{\partial M}{\partial b} = \frac{-c}{ad-bc} - \frac{d}{d(b+d)-c(a+c)} \\[2mm]
\sqrt{c'} &= \frac{\partial M}{\partial c} = \frac{-b}{ad-bc} + \frac{2c}{d(b+d)-c(a+c)} \\[2mm]
\sqrt{d'} &= \frac{\partial M}{\partial d} = \frac{a}{ad-bc} - \frac{2d}{d(b+d)-c(a+c)} \\[2mm]
&\quad \frac{\partial M}{\partial n} = 0.
\end{aligned}\right\} \tag{4}$$

By (3 : 10)

$$V(M) = aa' + bb' + cc' + dd', \tag{5}$$

and

$$V(\mu) = \mu^2(aa' + bb' + cc' + dd'). \tag{6}$$

$V(\mu)$ for cases 3 and 4 is given by formula (4 : 5). $V(\lambda)$ for these cases is given by formula (4 : 4).

$V(\lambda)$, for case 1 is given by formulae (3 : 24-27). For case 2, we use formula (3 : 27), with the following values of $\sqrt{a''}$ etc.

$$\left.\begin{aligned}
\sqrt{a''} &= \frac{c}{ac-bd} - \frac{c}{c(a+c)-d(b+d)} \\[2mm]
\sqrt{b''} &= \frac{-d}{ac-bd} + \frac{d}{c(a+c)-d(b+d)} \\[2mm]
\sqrt{c''} &= \frac{a}{ac-bd} - \frac{a+2c}{c(a+c)-d(b+d)} \\[2mm]
\sqrt{d''} &= \frac{-b}{ac-bd} + \frac{b+2d}{c(a+c)-d(b+d)}.
\end{aligned}\right\} \tag{7}$$

6. Summary

1. Formulae are presented for estimating recombination, viability and manifestation from backcross data, when one of two linked loci suffers from the two disturbances simultaneously.

2. Recombination is estimable from the manifesting classes only: the estimate is obtainable by equating the ratio of the observed numbers in the manifesting classes to the ratio of their expectation when linkage is not disturbed.

7. References

BAILEY, N. T. J. 1949. A method of allowing for differential viability in estimating linkage from backcross matings in coupling only or repulsion only. *Heredity*, *3*, 225-228.

BAILEY, N. T. J. 1950. The influence of partial manifestation on the detection of linkage. *Heredity*, *4*, 327-336.

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

FISHER, R. A. 1922. On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. Lond.,* A., *222*, 309-368.

FISHER, R. A. 1939. The precision of the product formula for the estimation of linkage. *Ann. Eugen., 9*, 50-54.

FISHER, R. A. 1949. Note on the test of significance for differential viability in frequency data from a complete three-point test. *Heredity, 3*, 215-219.

FISHER, R. A. 1953. *The Design of Experiments,* 8th Ed. Oliver and Boyd, Edinburgh.

FISHER, R. A. 1954. *Statistical Methods for Research Workers,* 12th Ed. Oliver and Boyd, Edinburgh.

FISHER, R. A., AND BALMUKAND, B. 1928. The estimation of linkage from the offspring of selfed heretozygotes. *J. Genet., 20*, 79-92.

LAING, E. 1958. Disturbed linkage and linkage group XIII of the mouse. Ph.D. Dissertation. Cambridge: University Library.

PARSONS, P. A. 1957. Partial manifestation of a gene in complete three and higher point backcross data. *Heredity, 11*, 217-222.

SÁNCHEZ-MONGE, E. 1952. The estimation of linkage with incomplete penetrance. *Heredity, 6*, 121-125.