

The Detection of a Sex Difference in Recombination Values Using Double Heterozygotes

R. A. FISHER

*Division of Mathematical Statistics, C.S.I.R.O.,
University of Adelaide, South Australia*

(Received 28 May 1962)

The method of detecting or measuring a difference between the recombination fractions in female and male gametogenesis is found on inserting typical numerical values to be quite as insensitive as its critics have suggested. If, however, double heterozygotes in Coupling and Repulsion are both available, reciprocal intercrosses will determine both recombination fractions without ambiguity. The use of double heterozygotes for studying linkage has thus fewer disadvantages, when both kinds are available, than has been supposed.

1. Linkage Values obtained by Self-fertilization

It has been observed by many authors (Fisher & Balmukand, 1928) that when a recombination fraction is determined by selfing double heterozygotes, what is calculable from such data is not the actual recombination value in male or female gametogenesis, unless these are equal, but the geometric mean of these two values. More rarely the comment has been added that if such data are available, not only from heterozygotes in Coupling but also from those in Repulsion, there will be a discrepancy between the linkage values obtained from these two sources, so that, in theory at least, a difference between the female and the male values could be detected in this way.

It might be guessed that it would require enormous numbers to detect, and still more to evaluate with any precision, a sex difference in this way. It was early observed also that even if a difference were recognizable, such data would supply no indication whether the recombination fractions were higher in the female or in the male.

Since I believe no case in the literature has been exhibited in detail, and since a general algebraic treatment would be complex and barely profitable, it may be worth the while to set out the numerical calculations for a sufficiently typical example, so that the provisional conclusions above may be seen to be verifiable and the steps of such a verification made clear.

The genetic situation may be expressed in a short table of gametic frequencies:

	♀ gametogenesis		♂ gametogenesis	
	Coupling	Repulsion	Coupling	Repulsion
<i>AB</i>	$q/2$	$p/2$	$q'/2$	$p'/2$
<i>Ab</i>	$p/2$	$q/2$	$p'/2$	$q'/2$
<i>aB</i>	$p/2$	$q/2$	$p'/2$	$q'/2$
<i>ab</i>	$q/2$	$p/2$	$q'/2$	$p'/2$.

If dominance is complete in both factors, the phenotypic frequencies arising from self-fertilization are:

	Coupling	Repulsion
<i>AB</i>	$(2 + qq')/4$	$(2 + pp')/4$
<i>Ab</i>	$(1 - qq')/4$	$(1 - pp')/4$
<i>aB</i>	$(1 - qq')/4$	$(1 - pp')/4$
<i>ab</i>	$qq'/4$	$pp'/4$.

In coupling, the frequencies expected in all observable classes are expressible in terms of qq' , which may, therefore, be estimated from such data: in repulsion all we can estimate is pp' . Now

$$p + q = 1 = p' + q',$$

so that if male and female gametogenesis yield the same gametic ratios, we shall find

$$\sqrt{qq'} + \sqrt{pp'} = 1.$$

In general, however, there will be, quite apart from errors of random sampling, a discrepancy in the expectations measured by

$$1 - \sqrt{qq'} - \sqrt{pp'}.$$

We may calculate the conditions in which such a discrepancy between the appropriate estimates will be significant.

If, for example, the recombination fraction were 30% in females, but 20% in males we should have

$$\sqrt{pp'} = \sqrt{\cdot 06} = \cdot 24494,897$$

$$\sqrt{qq'} = \sqrt{\cdot 56} = \cdot 74833,148$$

$$\text{Discrepancy} = \cdot 00671,955$$

$$\text{Total} \quad \underline{\quad \quad \quad} \quad 1 \cdot 00000,000.$$

At the 5% level of significance the discrepancy exceeds its standard error in the ratio

$$1 \cdot 95996.$$

The genetic situation may be expressed in a short table of gametic frequencies:

	♀ gametogenesis		♂ gametogenesis	
	Coupling	Repulsion	Coupling	Repulsion
<i>AB</i>	$q/2$	$p/2$	$q'/2$	$p'/2$
<i>Ab</i>	$p/2$	$q/2$	$p'/2$	$q'/2$
<i>aB</i>	$p/2$	$q/2$	$p'/2$	$q'/2$
<i>ab</i>	$q/2$	$p/2$	$q'/2$	$p'/2$

If dominance is complete in both factors, the phenotypic frequencies arising from self-fertilization are:

	Coupling	Repulsion
<i>AB</i>	$(2 + qq')/4$	$(2 + pp')/4$
<i>Ab</i>	$(1 - qq')/4$	$(1 - pp')/4$
<i>aB</i>	$(1 - qq')/4$	$(1 - pp')/4$
<i>ab</i>	$qq'/4$	$pp'/4$

In coupling, the frequencies expected in all observable classes are expressible in terms of qq' , which may, therefore, be estimated from such data: in repulsion all we can estimate is pp' . Now

$$p + q = 1 = p' + q',$$

so that if male and female gametogenesis yield the same gametic ratios, we shall find

$$\sqrt{qq'} + \sqrt{pp'} = 1.$$

In general, however, there will be, quite apart from errors of random sampling, a discrepancy in the expectations measured by

$$1 - \sqrt{qq'} - \sqrt{pp'}.$$

We may calculate the conditions in which such a discrepancy between the appropriate estimates will be significant.

If, for example, the recombination fraction were 30% in females, but 20% in males we should have

$$\sqrt{pp'} = \sqrt{.06} = .24494,897$$

$$\sqrt{qq'} = \sqrt{.56} = .74833,148$$

$$\text{Discrepancy} = .00671,955$$

$$\text{Total} \quad 1.00000,000.$$

At the 5% level of significance the discrepancy exceeds its standard error in the ratio

$$1.95996.$$

For a reasonable chance of significance it is necessary that the standard error of the estimated value of

$$\sqrt{pp'} + \sqrt{qq'}$$

shall be brought down to approximately

$$\frac{.00671,955}{1.95996} = .00342,841.$$

Now if θ stand for pp' or qq' , the sampling variance of $\sqrt{\theta}$ as estimated is known (Fisher, 1928-1958) to be

$$V(\sqrt{\theta}) = \frac{(1 - \theta)(2 + \theta)}{2n(1 + 2\theta)},$$

where the estimate is based on n observations.

For values of θ , .06 and .56, the coefficients of $1/n$ are

θ	$nV(\sqrt{\theta})$
.06	.86446,429
.56	.26566,038.

The variance of the sum is, therefore,

$$\frac{1}{n_1} (.86446,429) + \frac{1}{n_2} (.26566,038),$$

where n_1 offspring are bred from parents in Repulsion, and n_2 from parents in Coupling. To minimize this for a given total N of the two groups, take

$$\frac{n_1}{\sqrt{.86446,429}} = \frac{n_2}{\sqrt{.26566,038}}$$

or

$$\frac{n_1}{.92976,572} = \frac{n_2}{.51542,253} = \frac{N}{1.44518,825},$$

and when the optimal allocation has been made, the Standard Error will be

$$\frac{1.44518,825}{\sqrt{N}},$$

and the necessary value of N

$$\left(\frac{1.44518,825}{.00342841} \right)^2 = (421.533)^2 = 177,690,$$

confirming that the method is insensitive to the point of impracticability.

2. The Use of Intercrosses

It seems not to have been noticed that with no more experimental material than was postulated above—both types of double heterozygote, but no possibility of back-crossing, for the double recessives may be sterile—it is still possible to detect and to measure a sex difference in the recombination fraction, by making reciprocal intercrossoes between the two types of double heterozygotes. Moreover by this path, when a difference exists there is no ambiguity of interpretation: it is clear which sex shows closer linkage. The reciprocal crosses will yield the phenotypic frequencies shown below.

	Coupling ovules × repulsion pollen	Repulsion ovules × coupling pollen
<i>AB</i>	$(2 + p'q)/4$	$(2 + pq')/4$
<i>Ab</i>	$(1 - p'q)/4$	$(1 - pq')/4$
<i>aB</i>	$(1 - p'q)/4$	$(1 - pq')/4$
<i>ab</i>	$p'q/4$	$pq'/4$

Using the recombinations chosen above,

$$p'q = \cdot 24, pq' = \cdot 14,$$

the difference is

$$\cdot 10,$$

and for significance this needs a standard error not greater than

$$\cdot 0510214.$$

The sampling variance of either of the estimates is

$$2\theta(1 - \theta)(2 + \theta) + n(1 + 2\theta),$$

which gives, for the values $\cdot 24$ and $\cdot 14$, the coefficients of $1/n_1$ and $1/n_2$

$$\cdot 55212,973; \cdot 40258,750,$$

of which the square roots are

$$\cdot 74305,433; \cdot 63449,783,$$

giving the relative numbers to be bred from the two crosses to obtain optimal precision.

The sum is

$$1\cdot 37755,216,$$

and this divided by

$$0\cdot 0510214$$

gives

$$\sqrt{N} \ 26\cdot 9995$$

$$N \ 728\cdot 973,$$

so that only a few hundred from each cross would serve to detect the sex differences, and a few thousand plants would give separate estimates of the recombination fraction of female and male gametogenesis.

In practice the scoring coefficients given in "Statistical Tables" (Fisher & Yates, 1957), Table XIII could be used to evaluate $\sqrt{p'q}$ and $\sqrt{pq'}$.

REFERENCES

- FISHER, R. A. (1928-1958). "Statistical Methods for Research Workers," 2nd to 13th editions. Oliver & Boyd, Edinburgh.
- FISHER, R. A. & BALMUKAND, B. (1928). *J. Genetics* **20**, 79-91.
- FISHER, R. A. & YATES, F. (1957). "Statistical Tables," Section 55. Oliver & Boyd, Edinburgh.