Theoretical and Applied Genetics 45, 157–162

Arguments Against Intermating Before Selection in a Self-fertilising Species

D. G. PEDERSON

Faculty of Agriculture, University of Sydney, Sydney, New South Wales (Australia)

Summary. The expected benefit from intermating the F_2 population from a cross of two homozygous parents is investigated. It is noted that intermating will only be of benefit for two loci if the alleles are initially in the repulsion phase, and this result is extended to more than two loci by population simulation. If V_n is the population variance after n cycles of intermating followed by repeated selfing, the merit of intermating is measured by the magnitude of V_1/V_0. For a character controlled by loci on a short chromosome segment the ratio is invariably greater than 1.0, but as the average recombination value between loci is increased, either by increasing the chromosome length or by situating the loci on different chromosomes, individual values less than 1.0 become possible. If loci are spread over three or more chromosomes then the expected gain from intermating is consistently small. A simulation study shows that truncation selection is usually preferable to intermating as a procedure for increasing the proportion of desirable homozygotes in a population.

Introduction

The most usual breeding procedure for a self-fertilising species is to cross two homozygous genotypes and then select, in later generations, those progeny which have the desirable characteristics of both parents. Although a complete reassortment of the parental genes is preferred, with self-fertilisation there is only limited opportunity for this to occur. Recombination may therefore be promoted by an initial round of intermating commencing at the F_2 generation, and Hanson (1959) has concluded from a study of the maintenance of linkage groups that at least four cycles of intermating are necessary in most situations. Miller and Rawlings (1967) did observe a limited occurrence of non-parental types in the F_3 progenies of a cotton cross for which one parent was of interspecific origin, and found that six cycles of 50% outcrossing produced a better source of material for selection, perhaps due to a partial breakup of linkage blocks in the original material.

Intermating can be a costly and time-consuming procedure in a normally self-fertilising species. It is therefore encouraging that as few as 30 F_2 pairs are usually sufficient to achieve the beneficial effects of intermating without the loss of genetic material through drift (Baker, 1968).

However, Hanson considered only linked loci spread over a map length of 2.0 morgans or less, and similarly Baker considered the effects of a single chromosome. Further, a predominance of repulsion linkages was either assumed or implied in each case. The present study assesses the worth of intermating when these two restrictions, on the distribution of loci over chromosomes and the nature of parental linkages, are removed. It is assumed that there are two alleles, designated 0 and 1, at each segregating locus, and for a character controlled by $2k$ loci it is assumed that each parent is homozygous for the 1-allele at k of the loci. If V_n is defined as the genotypic variance of a population produced by n cycles of intermating of an F_2 population followed by repeated self-fertilisation, then such ratios as V_1/V_0 and V_2/V_0 are used to assess the merit of intermating.

Two Linked Loci

Repeated self-fertilisation of a heterozygote in the repulsion phase (genotype 10/01) will give the four homozygous genotypes with relative frequencies (Nelder, 1952)

$$10/10 \text{ and } 01/01: \tfrac{1}{2}/(1+2r)$$
$$11/11 \text{ and } 00/00: r/(1+2r)\ ,$$

where r is the recombination value for these two loci. We therefore have

$$V_0 = 8r/(1+2r)$$

if the 0- and 1-alleles take their indicated values.

If individuals of the F_2 generation, obtained by selfing the heterozygote 10/01, are intermated once prior to repeated self-fertilisation then the relative frequencies are (Baker, 1968)

$$10/10 \text{ and } 01/01: (2-r+2r^2)/(4+8r)$$
$$11/11 \text{ and } 00/00: (5r-2r^2)/(4+8r)\ .$$

It follows that

$$V_1 = 2(5r-2r^2)/(1+2r)$$

so that

$$V_1/V_0 = (5-2r)/4\ , \qquad 0 < r \leq \tfrac{1}{2}\ .$$

This ratio tends towards a maximum of 1.25 as r tends towards zero.

Two cycles of intermating give the expected relative frequencies

$$10/10 \text{ and } 01/01: (2 - 2r + 5r^2 - 2r^3)/(4 + 8r)$$
$$11/11 \text{ and } 00/00: (6r - 5r^2 + 2r^3)/(4 + 8r),$$

so that

$$V_2 = 2(6r - 5r^2 + 2r^3)/(1 + 2r)$$

and

$$V_2/V_0 = (6 - 5r + 2r^2)/4\ ,\ 0 < r \leqq \tfrac{1}{2}\ .$$

The increase in genotypic variance from a single cycle of intermating will therefore be less than 25% of V_0 for two loci of equal effect, and further cycles of intermating will produce progressively smaller absolute increases.

If the loci are in the coupling phase then the relative frequencies of the four homozygotes are given by interchanging the symbols 0 and 1 at either locus, and it is found that

$$V_0 = 4/(1 + 2r)\ ,$$
$$V_1 = 2(2 - r + 2r^2)/(1 + 2r)\ ,$$

and

$$V_2 = 2(2 - 2r + 5r^2 - 2r^3)/(1 + 2r)\ .$$

The ratio

$$V_1/V_0 = (2 - r + 2r^2)/2$$

is now less than or equal to unity, with a minimum value of 15/16 (0.94) when $r = 0.25$. The ratio

$$V_2/V_0 = (2 - 2r + 5r^2 - 2r^3)/2$$

has a minimum value of 0.89 when $r = 0.23$.

Before the two-locus theory can be extended, it is necessary to prove a result concerning the production of homozygotes from a multi-locus heterozygote

Determining Homozygote Frequencies

The aim is the develop a method for determining the relative frequencies of the homozygous genotypes given by repeated selfing of a multiple heterozygote.

Let P_{ij} be the event that a derived homozygote is in the parental condition for loci i and j. For example, if the parental genotypes were 01/01 and 10/10 then P_{12} is the event that a homozygote has either of the parental genotypes at loci 1 and 2. The event of the non-parental condition, namely genotypes 11/11 and 00/00 in the example, will be denoted by N_{ij}, and the recombination value for loci i and j will be written as r_{ij}.

For three loci, let

$$\varnothing = p(P_{12} \text{ and } P_{23}) - p(P_{12})\ p(P_{23})$$
$$= p(N_{12} \text{ and } N_{23}) - p(N_{12})\ p(N_{23})\ ,$$

where p denotes "probability of". The equivalence of the above expressions follows from the table of probabilities:

	Loci 1 and 2: parental	Loci 1 and 2: non-parental	
Loci 2 and 3: parental	$p(P_{12} \text{ and } P_{23})$	$p(N_{12} \text{ and } P_{23})$	$p(P_{23})$
Loci 2 and 3: non-parental	$p(P_{12} \text{ and } N_{23})$	$p(N_{12} \text{ and } N_{23})$	$p(N_{23})$
	$p(P_{12})$	$p(N_{12})$	

We therefore have

$$2\varnothing = [p(P_{12} \text{ and } P_{23}) + p(N_{12} \text{ and } N_{23})] - [p(P_{12})\ p(P_{23}) + p(N_{12})\ p(N_{23})]\ .$$

But $[p(P_{12} \text{ and } P_{23}) + p(N_{12} \text{ and } N_{23})]$ is the total probability that a homozygote is in the parental condition for loci 1 and 3, which Nelder (1952) has shown is equal to $1/(1 + 2r_{13})$. The above identity therefore becomes

$$2\varnothing = \frac{1}{1 + 2r_{13}} - \left[\frac{1}{1 + 2r_{12}}\frac{1}{1 + 2r_{23}} + \frac{2r_{12}}{1 + 2r_{12}}\frac{2r_{23}}{1 + 2r_{23}}\right]$$
$$= \frac{1}{1 + 2r_{13}} - \frac{1 + 4r_{12}r_{23}}{(1 + 2r_{12})\ (1 + 2r_{23})}\ .$$

But if map distance (x) and recombination value (r) are related according to the Kosambi (1944) mapping function

$$x = \frac{1}{4}\log_e\left(\frac{1 + 2r}{1 - 2r}\right)$$

then

$$r_{13} = (r_{12} + r_{23})/(1 + 4r_{12}r_{23})\ ,$$

so that

$$1 + 2r_{13} = (1 + 2r_{12})\ (1 + 2r_{23})/(1 + 4r_{12}r_{23})$$

and $\varnothing = 0$.

Provided that the Kosambi mapping function holds, we therefore have

$$p(P_{12} \text{ and } P_{23}) = p(P_{12})\ p(P_{23})\ ,$$

so that the condition of a chromosome segment is independent of the condition of any other segment once homozygosity has been attained. Interference of the type indicated by the Kosambi mapping function has been found to occur in practice (e.g. Wallace, 1957), and this function is therefore a reasonable model on which to build further theory.

As an example of the application of the above result, suppose that the genotype 1011/0100 were repeatedly selfed. Then the relative frequency of the homozygote 1111/1111 would be given by

$$p(11 \text{ for segment } 1, 2)$$
$$\times\ p(\text{non-parental for segment } 2, 3)$$
$$\times\ p(\text{parental for segment } 3, 4)$$
$$= \frac{r_{12}}{1 + 2r_{12}} \cdot \frac{2r_{23}}{1 + 2r_{23}} \cdot \frac{1}{1 + 2r_{34}}\ .$$

For a particular set of loci and recombination values it would only be necessary to determine the homozygote frequency distribution for a single source genotype, say 111 ... 11/000 ... 00. The homozygote distribution for any other source genotype could then

be obtained by the appropriate transformation of the symbols 0 and 1.

More than Two Linked Loci

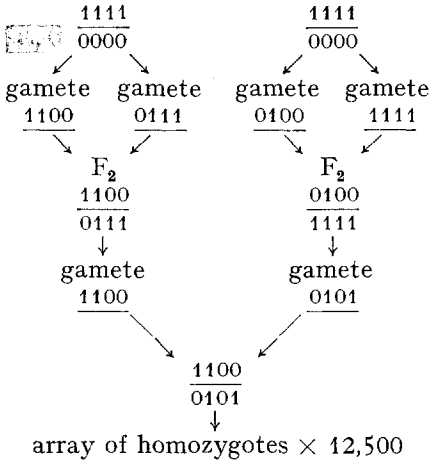
One Chromosome

For four loci on a single chromosome there are three unique heterozygous arrangements, assuming that there are two 1-alleles on each chromosome, namely

$$\frac{1100}{0011}, \quad \frac{1010}{0101}, \quad \text{and} \quad \frac{1001}{0110}.$$

The loci were assumed to be equally spaced over chromosome segments of length 0.1 and 0.5 morgans, and the array of homozygotes produced by repeated self-fertilisation was determined in each case. The variance (V_0) was calculated for each array.

A simulation procedure was also carried out, commencing with the genotype 1111/0000. An F_2 individual was produced by the random generation of two gametes, using a random walk procedure based on recombination frequencies given by the Kosambi function. A second F_2 individual was then produced, and a random gamete from each F_2 gave an individual which was the product of one cycle of random mating. The array of homozygous genotypes from repeated selfing of this individual was determined. The whole process was repeated 12,500 times for the 4-locus case to give an "average" homozygous array. Once this had been done for the genotype 1111/0000 it was possible to determine the array for each of the three 4-locus heterozygotes by appropriate transformations of the gene symbols, and in each case the value of V_1 was calculated. Diagrammatically, the procedure was as follows:



For a chromosome segment of length 0.1 morgans the values of V_1/V_0 were 1.175 (parent 1100/0011), 1.224 (parent 1010/0101), and 1.270 (parent 1001/0110). The average of 1.223 is close to the maximum value of 1.25 for two loci in the repulsion phase. The corresponding values of V_1/V_0 for a segment of length 0.5 morgans were 1.021, 1.189, and 1.144, with an average of 1.118. Both sets of values are plotted in Fig. 1.

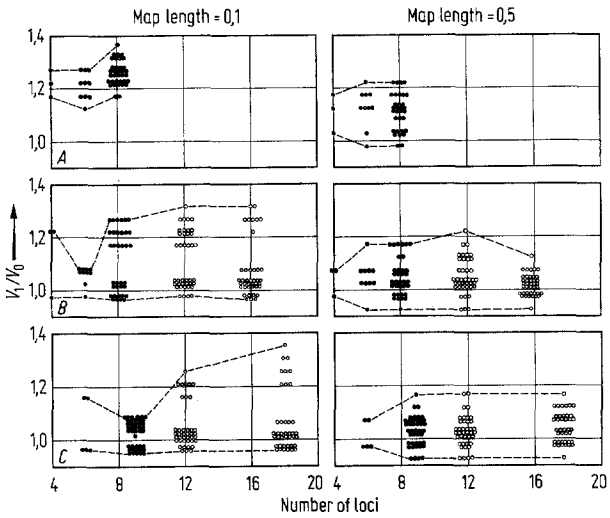


Fig. 1. Values of V_1/V_0 for either all possible allele arrangements (●) or a random sample of 40 arrangements (○); A, B, and C indicate 1, 2, and 3 chromosomes respectively. Extreme values have been joined with a dotted line to give a clearer indication of the range

Thus, if a character is controlled by four closely linked loci then one cycle of intermating may be expected to produce a 20% increase in genetical variance. If the loci are evenly spread over a larger chromosome segment then the expected increase is smaller and individual increases may be close to zero.

The simulation procedure was also carried out for six loci (7,500 F_2 pair-matings) and eight loci (5000 F_2 pair matings) for the two map lengths. There are ten heterozygous arrangements for six loci and 35 possible arrangements for eight loci, and the value of V_1/V_0 for each is plotted in Fig. 1. For the greater map length there are now genotypes for which intermating would actually decrease the homozygous genotypic variance. In general, genotypes with gene complexes linked in the coupling phase gave a high V_0, in some cases above the equilibrium value ($V_\infty = 8.0$), and a low value of V_1/V_0, and the reverse was true of genotypes for which repulsion phase linkages predominated.

As an illustration, the results for the 8-locus genotypes giving the highest and lowest values of V_1/V_0 were as follows, for the 0.5 morgan map length:

Genotype:	$\frac{11110000}{00001111}$	$\frac{10011001}{01100110}$
V_0:	11.96	2.21
V_1:	11.79	2.70
V_1/V_0:	0.99	1.22

More than one Chromosome

If the array of homozygotes and their frequencies is known for a heterozygote of n linked loci then it is

a matter of simple multiplication to find the array for two unlinked heterozygous segments, each with n loci. The simulation results for 2, 3, 4, 6, and 8 loci were therefore used to obtain the values of V_1/V_0 for 4, 6, 8, 12, and 16 loci equally divided between two chromosomes. The values are plotted in Fig. 1, with a solid circle indicating that the plotted point corresponds to one of all possible arrangements for that number of loci, and an open circle indicating that the point is one of a random sample of 40 chosen from the much larger set of possible values.

The new feature for the shorter map length is that there are individual values of V_1/V_0 less than 1.0. The set of points for six loci has low variability because it is not possible to have all parental chromosomes of equal effect when there are three loci on each.

The study was then extended to the case of three chromosome segments with a total of either 6, 9, 12, or 18 equally spaced loci, and the values of V_1/V_0 are plotted in Fig. 1. In general, the spread of points is less than for the two-chromosome case, and a greater proportion of the values are close to or less than 1.0.

Two Cycles of Intermating

The simulation procedure was repeated for two cycles of intermating preceding selfing to homozygosity. The average values of V_2/V_0 are presented in Table 1, together with the average values of V_1/V_0.

Table 1. *Average values of V_1/V_0 and V_2/V_0, over all possible arrangements or a random sample of 40 arrangements, for various numbers of chromosome segments and numbers of loci*

Number of chromosomes	Map distance	Number of loci	$\frac{V_1}{V_0}$	$\frac{V_2}{V_0}$
1	0.1	4	1.223	1.434
		6	1.216	1.449
		8	1.270	1.496
	0.5	4	1.118	1.242
		6	1.128	1.248
		8	1.107	1.248
2	0.1	4	1.121	1.229
		6	1.056	1.106
		8	1.111	1.216
		12	1.113	1.222
		16	1.082	1.148
	0.5	4	1.025	1.040
		6	1.047	1.085
		8	1.054	1.112
		12	1.056	1.114
		16	1.014	1.061
3	0.1	6	1.069	1.132
		9	1.036	1.070
		12	1.058	1.114
		18	1.082	1.138
	0.5	6	1.014	1.023
		9	1.027	1.050
		12	1.033	1.070
		18	1.041	1.086

In general, the absolute gain from two cycles of intermating is approximately twice the gain from a single cycle, and this pattern was found to be consistent across genotypes.

Selection of F_3 Families

The aim of intermating is to increase the relative frequency of desirable recombinants, an aim which may also be accomplished by artificial selection. For many commercial species family selection is the most appropriate method, and accordingly a study was made of the directional selection of F_3 families.

Eight parental heterozygotes were chosen, including two with eight loci on a single chromosome

$$\frac{01010101}{10101010} \quad \text{and} \quad \frac{00111100}{11000011},$$

three with eight loci on two chromosomes

$$\frac{0111}{1000}\,\frac{1000}{0111}, \quad \frac{0110}{1001}\,\frac{0110}{1001}, \quad \text{and} \quad \frac{0101}{1010}\,\frac{0101}{1010},$$

and three with eight loci on four chromosomes

$$\frac{01}{10}\,\frac{01}{10}\,\frac{11}{00}\,\frac{00}{11}, \quad \frac{00}{11}\,\frac{00}{11}\,\frac{11}{00}\,\frac{11}{00}, \quad \text{and} \quad \frac{01}{10}\,\frac{01}{10}\,\frac{01}{10}\,\frac{01}{10}.$$

Each chromosome segment was of length 0.1 morgan and the loci were equally spaced. For each heterozygote the simulation routine was as follows:

(1) generate 100 F_2 genotypes, using the Kosambi function to determine recombination frequencies

(2) for each F_2 genotype generate 20 F_3 individuals by simulated self-fertilisation and determine the family mean (m_j; $j = 1, 2, \ldots, 100$) and the overall F_3 mean ($\bar{m}$) by giving the 0- and 1-alleles their indicated values

(3) calculate the variance of the $m_j(\sigma_G^2)$

(4) nominate a value (h^2) for the heritability of F_3 family means and calculate the phenotypic variance as

$$\sigma_P^2 = \sigma_G^2/h^2$$

(5) assign a weight (w_j) to the jth F_3 family, and hence the jth F_2 genotype,

$$w_j = 1 + \bar{i}(m_j - \bar{m})/\sigma_P,$$

(Kimura, 1958; Griffing, 1960) where $\bar{i}$ is the standardised selection differential corresponding to the proportion of F_3 families selected on the basis of their mean value

(6) determine the genotypic array from repeated self-fertilisation of each original F_2 genotype, and hence determine the pooled array for all genotypes following selection by assigning the weight w_j to the array of the jth genotype.

The heritability was set at 0.2 and 0.5, and selection was of the top 80% and top 50% of F_3 families. Three replicate runs were carried out for each combination of parameters.

Table 2. *The relative frequency of desirable homozygotes in a repeatedly selfed, initially heterozygous population, when either intermating or selection is first carried out; p = proportion of F_3 families selected on the basis of phenotypic mean*

Parent genotype	Cycles of intermating			Heritability			
				0.2		0.5	
	0	1	2	$p = 0.8$	$p = 0.5$	$p = 0.8$	$p = 0.5$
01010101 / 10101010	0.0001	0.0003	0.0005	0.0001 ±0.0000	0.0001 ±0.0000	0.0001 ±0.0000	0.0002 ±0.0001
00111100 / 11000011	0.0273	0.0337	0.0420	0.0389 ±0.0131	0.0501 ±0.0185	0.0477 ±0.0173	0.0700 ±0.0282
0111 1000 / 1000 0111	0.2233	0.2195	0.2185	0.2640 ±0.0119	0.3142 ±0.0150	0.3032 ±0.0143	0.4035 ±0.0207
0110 0110 / 1001 1001	0.0068	0.0110	0.0145	0.0104 ±0.0030	0.0136 ±0.0042	0.0129 ±0.0039	0.0192 ±0.0063
0101 0101 / 1010 1010	0.0036	0.0056	0.0075	0.0071 ±0.0004	0.0099 ±0.0005	0.0093 ±0.0005	0.0147 ±0.0008
01 01 11 00 / 10 10 00 11	0.1716	0.1640	0.1587	0.1833 ±0.0122	0.2222 ±0.0146	0.2137 ±0.0140	0.2913 ±0.0189
00 00 11 11 / 11 11 00 00	0.2285	0.2182	0.2100	0.2653 ±0.0264	0.3173 ±0.0317	0.3060 ±0.0306	0.4101 ±0.0413
01 01 01 01 / 10 10 10 10	0.0305	0.0414	0.0518	0.0355 ±0.0076	0.0477 ±0.0103	0.0427 ±0.0097	0.0609 ±0.0152

The array of homozygotes for zero, one and two cycles of intermating had previously been determined in the initial simulation study.

The relative frequency of homozygotes with a value of 12 or greater was measured in each case (Table 2). A single value is given for the intermated populations since there was sufficient replication to give reliable estimates, but for each selected population the mean and standard error of the three replicate values is presented.

The equilibrium frequency of 12-plus homozygotes with repeated intermating and selfing is $37/256 = 0.145$, and the relative frequency with intermating tends towards this value. Selection always increases the frequency of desirable homozygotes, and for the five cases where intermating is also beneficial it is observed that selection of the top 80% of F_3 families is roughly equivalent to a single cycle of intermating. However, the advantage of selection is clear for those genotypes for which intermating reduces the probability of obtaining desirable homozygotes.

Discussion

Two-locus theory suggests that if close linkages are predominantly in the repulsion phase for a multilocus character then intermating will promote their breakup, leading to an increase in the relative frequency of desirable homozygotes after repeated self-fertilisation. However, if all possible arrangements of alleles are equally likely then close coupling linkages will also occur and it would be an advantage to maintain these in the parental condition. These principles govern the results presented in this paper.

There are three likely situations in practice:

(1) intermating will always be of benefit for a character controlled by loci on a single short chromosome segment. The increase in the variance of homozygotes for each cycle of intermating would be of the order of 20%.

(2) on the average, intermating will be of benefit for a character controlled by loci spread over a long chromosome segment or situated on two or more short chromosome segments. However, individual cases will occur for which intermating decreases the genotypic variance of derived homozygotes. It may be better to search for crosses of this type, which display a high degree of transgressive segregation, than to attempt to improve the potential of other crosses by intermating.

(3) intermating will be of doubtful value for a character controlled by loci spread over three or more long chromosome segments, due to the over-riding effect of chromosome re-assortment. For a map length of 0.5 morgans the expected increase in homozygous genotypic variance per cycle of intermating is approximately 10%, 5%, and 3% for loci situated on 1, 2, and 3 chromosomes respectively (Table 2). Increases of this magnitude will usually be a poor return for the amount of work involved in the intermating process.

Directional selection is preferred as a method for increasing the frequency of desirable homozygotes, even when both the intensity of selection and the heritability of the character are of very low magnitude.

Of the three situations defined above, it is not possible to state which is more likely in practice. Studies involving aneuploidy in wheat (Law and Worland, 1972) have shown that the loci controlling such characters as plant height, grain size, and number of grains per ear form closely linked complexes, at least for chromosome 7B of the wheat complement. At the same time, for final plant height and 250-grain weight a number of chromosomes from particular varieties have been shown to have significant effects. It is tempting to conclude that intermating will therefore not be of benefit for such characters, but when two highly developed varieties are crossed in a breeding programme it may happen that case (1) applies since only a small proportion of loci are segregating. Thus the present state of our knowledge makes it difficult to generalise on the possible merits of intermating, even for a particular character in a particular species, but the odds are against it being a useful procedure.

The variance of a derived homozygous population has been used as a basic parameter since most breeding programmes delay selection for quantitative characters until the degree of heterozygosity has considerably diminished. When the number of controlling loci is small the distribution of homozygote values may be non-normal, but this was not thought to be an important factor since changes of variance were the quantities of interest. The directional selection study demonstrated that a change in variance is a good reflection of a change in the amount of transgressive segregation.

The results also apply to the case where each parent possesses the desirable qualities of one of two quantitative characters, and these qualities are to be recombined in the progeny. Each parent then possesses the "1-alleles" for one character and the aim is to select progeny with the complete set of 1-alleles, as is the case when a single character is undergoing improvement. In fact, if the loci controlling particular characters tend to be localised on the chromosomes then coupling linkages will predominate in the above situation, and intermating will be even less likely to be of advantage than if all possible arrangements of alleles are equally frequent.

Acknowledgements

The author acknowledges the support of the Australian Research Grants Committee in providing funds for computing, and is grateful to the Department of Genetics, University of Birmingham for providing facilities during the writing-up of this work. Members of the Department of Genetics took part in helpful discussion.

Literature

Baker, R. J.: Extent of intermating in self-pollinated species necessary to counteract the effects of genetic drift. Crop Sci. **8**, 547–550 (1968).

Griffing, B.: Theoretical consequences of truncation selection based on the individual phenotype. Aust. J. biol. Sci. **13**, 307–343 (1960).

Hanson, W. D.: The breakup of initial linkage blocks under selected mating systems. Genetics **44**, 857–868 (1959).

Kimura, M.: On the change of population fitness by natural selection. Heredity **12**, 145–167 (1958).

Kosambi, D. D.: The estimation of map distance from recombination values. Ann. Eugen., Lond., **12**, 172 to 175 (1944).

Law, C. N., Worland, A. J.: Aneuploidy in wheat and its use in genetic analysis. Plant Breeding Inst., Camb. Annual Report, 25–65 (1972).

Miller, P. A., Rawlings, J. O.: Breakup of initial linkage blocks through intermating in a cotton breeding population. Crop Sci. **7**, 199–204 (1967).

Nelder, J. A.: Some genotypic frequencies and variance components occurring in biometrical genetics. Heredity **6**, 387–394 (1952).

Wallace, M. E.: A balanced three-point experiment for linkage group V of the house mouse. Heredity **11**, 223–258 (1957).

Received March 8, 1974

Communicated by R. Riley

Dr. D. G. Pederson
Faculty of Agriculture
University of Sydney
Sydney, New South Wales
(Australia) 2006