RECURRENT SELECTION

I. THE DETECTION OF OVERDOMINANCE

J. C. BOWMAN *
Institute of Animal Genetics, West Mains Road, Edinburgh, 9

1. INTRODUCTION

THE results of regression analyses have been widely used on many occasions to draw conclusions about the genetic basis of quantitative characters. In recent years much theoretical work has been published dealing with the application of this form of analysis to the determination of the average dominance of genes controlling such characters. Most of this theory refers to the analysis of diallel crosses (see Hayman, 1954) or to the relationship between inbred line performance and the performance of the crosses derived from such lines (Hull, 1952; Comstock and Robinson, 1952). Some attempt has been made to apply this technique to cases where both parents and progeny are heterozygous (Dickinson and Jinks, 1956).

A modified version of recurrent selection as suggested by Hull (1945) has been tried out experimentally on laboratory animals by Bell, Moore and Warren (1955), Rasmuson (1956) and Bowman (1958). This method of selection is designed primarily to exploit overdominance. At present there is no method of estimating overdominance in outbred populations. Consequently any analysis which can estimate the average dominance, or detect overdominance of genes controlling the character under selection would be valuable. An attempt has been made here by means of regression analyses to show one way in which overdominance can be detected in recurrent selection programmes as applied to animals. The analyses are essentially an extension to random bred individuals of the constant parent regression technique as applied to inbred lines by Hull (1952).

2. THEORETICAL ANALYSIS

The plan of recurrent selection as applied to animals has been outlined by Bell, Moore and Warren (1955), and the system of mating under consideration here is shown in fig. 1. In such a scheme two relationships are of particular interest. (i) The covariance between, and regression of, a sire's testcross progeny mean performance and his full sister mean performance. This is important in cases where the character under selection can only be measured in females, or where selection of heterogeneous strain progeny is carried out before their performance for the character can be measured. (ii) The covariance

* Present address : Breeding Research Department, Thornber Bros. Ltd., Mytholmroyd, Halifax, Yorkshire.

between, and regression of, a sire's testcross progeny mean performance and his pure strain progeny mean performance.
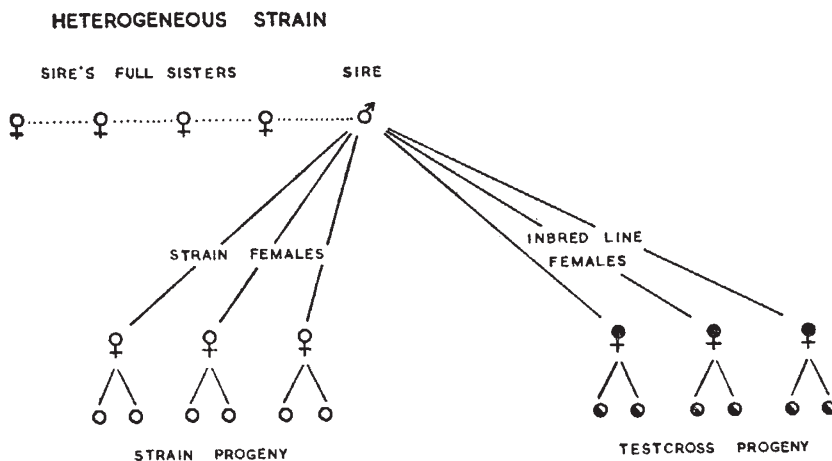
HETEROGENEOUS STRAIN

FIG. 1.—System of mating in a recurrent selection programme.

In the theory which follows a number of assumptions have been made. They are:

(a) Diploid segregation.

(b) No multiple allelism.

(c) Independent action of non-allelic genes.

Consider a single locus with two alleles, A_1 and A_2, such that the gene frequencies are p and q respectively, and such that $p+q=1$.

TABLE 1

Genotypes and frequency of parents and progeny

Let $P = p^2$; $H = 2pq$; $Q = q^2$

Family type		Frequency of parent family	Genotype and frequency of progeny		
			A_1A_1	A_1A_2	A_2A_2
1.	$A_1A_1 \times A_1A_1$	P^2	P^2	...	...
2.	$A_1A_1 \times A_1A_2$	$2PH$	PH	PH	...
3.	$A_1A_1 \times A_2A_2$	$2PQ$	...	$2PQ$	...
4.	$A_1A_2 \times A_1A_2$	H^2	$\frac{1}{4}H^2$	$\frac{1}{2}H^2$	$\frac{1}{4}H^2$
5.	$A_1A_2 \times A_2A_2$	$2HQ$	...	HQ	HQ
6.	$A_2A_2 \times A_2A_2$	Q^2	...	...	Q^2
		Sums	$(P+\frac{1}{2}H)^2 = p^2$	$2(P+\frac{1}{2}H)(Q+\frac{1}{2}H)$ $= 2pq$	$(Q+\frac{1}{2}H)^2 = q^2$

In a random mating population corresponding to the strain population in the scheme above there will be six family types. The frequencies of these types together with the genotype and frequency of their progeny are shown in table 1.

Using the notation of Mather (1949) let the phenotypic values of the three genotypes be:

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ +d & h & -d \end{array}$$

and then the mean value of the strain progeny

$$= d(p-q)+2pqh$$

and the variance

$$= 2pq[d-h(p-q)]^2+4h^2p^2q^2.$$

In a random mating population the mean and variance of a sire's full sisters will be the same as the mean and variance of his strain

TABLE 2

Mean value of a sire's sisters

Family type	1	2	3	4	5	6
Mean value of sire's sisters	d	$\frac{1}{2}d+\frac{1}{2}h$	h	$\frac{1}{2}h$	$-\frac{1}{2}d+\frac{1}{2}h$	$-d$

TABLE 3

Mean value of a sire's testcross progeny

	Sire genotype		
	A_1A_1	A_1A_2	A_2A_2
Mean value of testcross progeny			
Tester A_1A_1	d	$\frac{1}{2}d+\frac{1}{2}h$	h
Tester A_2A_2	h	$\frac{1}{2}h-\frac{1}{2}d$	$-d$

TABLE 4

Products of values of sisters × testcross progeny
Tester A_1A_1

Family type	Sire genotype		
	A_1A_1	A_1A_2	A_2A_2
1	d^2	...	...
2	$\frac{1}{2}d^2+\frac{1}{2}dh$	$\frac{1}{4}d^2+\frac{1}{2}dh+\frac{1}{4}h^2$	...
3	...	$\frac{1}{2}dh+\frac{1}{2}h^2$	...
4	$\frac{1}{2}dh$	$\frac{1}{4}dh+\frac{1}{4}h^2$	$\frac{1}{2}h^2$
5	...	$-\frac{1}{4}d^2+\frac{1}{4}h^2$	$-\frac{1}{2}dh+\frac{1}{2}h^2$
6	...	...	$-dh$

progeny. In a recurrent selection programme the genotype of the inbred tester will be either A_1A_1 or A_2A_2. Suppose the tester is A_1A_1. Then the mean values of the sire's sisters and the sire's testcross progeny are as shown in table 2 and table 3 respectively. The products of values of sisters × testcross progeny are shown in table 4.

The uncorrected mean products = products × frequency

$$= (P^2 + \tfrac{3}{4}PH - \tfrac{1}{4}HQ)d^2 + (PH + \tfrac{1}{4}H^2 + PQ - \tfrac{1}{2}HQ - Q^2)dh$$
$$+ (\tfrac{1}{4}PH + PQ + \tfrac{1}{4}H^2 + \tfrac{3}{4}HQ)h^2.$$

The correction factor to obtain the covariance from the uncorrected mean products is the product of the mean value of sisters × mean value of testcross progeny. Converting p and q to P and Q, the mean value of sisters

$$= (p-q)d + 2hpq.$$

Now $P - Q = p - q$.

Therefore mean value of sisters $= d(P-Q) + hH$.

The mean value of testcross progeny $= (p^2 + pq)d + (pq + q^2)h$.
Now $(P + \tfrac{1}{2}H) = p$.
Therefore mean value of testcross progeny

$$= dp + hq = d(P + \tfrac{1}{2}H) + h(Q + \tfrac{1}{2}H).$$

The product of the means

$$= (P^2 + \tfrac{1}{2}PH - \tfrac{1}{2}HQ - PQ)d^2 + (1\tfrac{1}{2}PH + \tfrac{1}{2}H^2 - Q_2 + PQ - \tfrac{1}{2}HQ)dh$$
$$+ (HQ + \tfrac{1}{2}H^2)h^2.$$

Then the subtraction of the correction factor from the sum of cross-products leaves the covariance, which can be shown to equal

$$d^2(\tfrac{1}{2}pq) + dh(-p^2q) + h^2(\tfrac{1}{2}pq[p-q])$$
$$= \tfrac{1}{2}pq[d^2 - 2pdh + (p-q)h^2]. \qquad (1)$$

Hence the covariance of a sire's sister mean performance with his testcross progeny mean performance is:

when tester is $A_1A_1 = (1)$
when tester is $A_2A_2 = \tfrac{1}{2}pq[d^2 + 2qdh + (q-p)h^2]. \qquad (2)$

By a similar method of calculation it can be shown that the covariance of a sire's testcross progeny mean performance with his strain progeny mean performance is the same as (1) and (2) above when the tester is A_1A_1 and A_2A_2 respectively. The variance of a sire's strain progeny is the same as the variance of his sisters in the absence of selection. The conclusions about the covariance and regression which follow can therefore be applied to both relationships (i) and (ii).

The regression values for different values of p and q and for different values of h (various dominance levels) when $d = 1$ have been calculated and are given in tables 5 and 6. In cases where h has a value other than zero, dominance or overdominance is assumed to be in the direction of higher values, and p is the frequency of A_1 the allele that has the higher homozygote.

I am grateful to Dr Alan Robertson and Dr D. S. Falconer for showing me and allowing me to reproduce here an alternative method of deriving formulæ (1) and (2).

This method depends on knowing how breeding values and average effects of genes can be expressed in terms of the d and h notation of Mather (1949). The necessary interrelationships are given by Falconer (1959). Again let p and q be the gene frequencies of A_1 and A_2 in the

TABLE 5

Regression of a sire's testcross daughters on his sisters
Tester A_1A_1 (" high ")

Values of p	$d=1$ $h=0$	$d=1$ $h=1$	$d=1$ $h=0.5$	$d=1$ $h=2$	$d=1$ $h=1.5$	$d=1$ $h=1.25$	$d=0$ $h=1$
1·0	0	0	0	0	0	0	0
0·9	0·25	0	0·12	0·14	0·06	0	0·27.
0·8	0·25	0	0·15	0·04	—0·02	—0·03	0·23
0·7	0·25	0	0·13	—0·03	—0·05	—0·04	0·16
0·6	0·25	0	0·12	—0·09	—0·06	—0·03	0·08
0·5	0·25	0	0·11	—0·08	—0·06	—0·03	0
0·4	0·25	0	0·10	—0·09	—0·06	—0·03	—0·08
0·3	0·25	0	0·10	—0·09	—0·06	—0·03	—0·16
0·2	0·25	0	0·09	—0·09	—0·05	—0·03	—0·23
0·1	0·25	0	0·10	—0·09	—0·05	—0·03	—0·27
0·0	0	0	0	0	0	0	0

p is frequency of A_1 in the strain. A_1 increases value.

strain from which the sires are taken; and let p' and q' be the corresponding gene frequencies in the tester strain. Let X be the average

TABLE 6

Regression of a sire's testcross daughters on his sisters
Tester A_2A_2 (" low ")

Values of p	$d=0$ $h=0$	$d=1$ $h=1$	$d=1$ $h=0.5$	$d=1$ $h=2$	$d=1$ $h=1.5$	$d=1$ $h=1.25$	$d=0$ $h=1$
1·0	0	0	0	0	0	0	0
0·9	0·25	0·45	0·35	—0·42	—0·29	0	—0·27
0·8	0·25	0·42	0·46	—0·11	0·09	0·25	—0·23
0·7	0·25	0·38	0·40	0·09	0·23	0·31	—0·16
0·6	0·25	0·36	0·36	0·20	0·28	0·32	—0·08
0·5	0·25	0·33	0·33	0·25	0·29	0·32	0
0·4	0·25	0·31	0·31	0·27	0·29	0·30	0·08
0·3	0·25	0·29	0·29	0·27	0·29	0·29	0·16
0·2	0·25	0·28	0·28	0·27	0·27	0·28	0·23
0·1	0·25	0·26	0·30	0·26	0·26	0·26	0·27
0·0	0	0	0	0	0	0	0

p is frequency of A_1 in the strain. A_1 increases value.

effect of the gene substitution in the pure strain, and X' the average effect of the gene substitution in the tester strain.

Now the breeding values of the three possible genotypes (*a*) when the sires are mated in their own strain, and (*b*) when they are mated in the tester strain, can be expressed as deviations from the population

means. These are shown in table 7. The expected values of daughters are half the appropriate breeding values given in columns (a) and (b). The covariance of sires' strain daughters with testcross daughters is therefore obtained by multiplying the frequency by $\frac{1}{2}$ the breeding value in column (a) by $\frac{1}{2}$ the breeding value in column (b) and summing. This gives

$$XX'[p^2qq'+\tfrac{1}{2}pq(q-p)(q'-p')+q^2pp']$$

which can be reduced to

$$\tfrac{1}{2}pqXX'$$

The average effect of a gene substitution expressed in the d and h notation is $X = d+h(q-p)$; so the covariance of the two sets of daughters becomes

$$\tfrac{1}{2}pq[d+h(q-p)][d+h(q'-p')] = \tfrac{1}{2}pq[d+h(q-p)][(d-h)]$$

if the tester is A_1A_1. This formula is equal to (1) above.

TABLE 7

Sires' breeding values in their own strain and in a tester strain

Genotype of sire	Frequency	Breeding values	
		(a) In own strain	(b) In tester strain
A_1A_1	p^2	$2qX$	$2q'X'$
A_1A_2	$2pq$	$(q-p)X$	$(q'-p')X'$
A_2A_2	q^2	$-2pX$	$-2p'X'$

Alternatively the covariance $= \tfrac{1}{2}pq[d+h(q-p)][(d+h)]$ if the tester is A_2A_2. This formula is equal to (2) above.

Now consider the sire's sisters. The expected value of any sister (or the mean of a number of sisters) is $\frac{1}{2}$ the breeding value $+\frac{1}{4}$ of the dominance deviation. The dominance deviation does not contribute to the covariance of sisters with daughters. So the covariance of sisters with testcross daughters is the same as that between strain and testcross daughters. This may be demonstrated as follows. The dominance of the sire's three genotypes are:

Sire genotype	Dominance deviation
A_1A_1	$-2q^2h$
A_1A_2	$2pqh$
A_2A_2	$-2p^2h$

These give the following additional terms to the covariance of strain sisters with testcross daughters (under (b) in table 7).

$$\tfrac{1}{4}hX'[-4p^2q^2q'+4p^2q^2(q'-p')+4p^2q^2p']$$
$$= hX'p^2q^2(-q'+q'-p'+p') = 0.$$

3. DISCUSSION AND PRACTICAL OBSERVATIONS

The exact relevance of applying conclusions based on calculations from single loci to characters of multigenic inheritance is unknown. For quantitative characters the overall regression observed is a summation of values for all loci controlling the character concerned. If the loci have unequal effects the observed regressions will be automatically weighted in favour of those with the largest effect. In practice it is likely that the inbred tester would be homozygous for the superior (or dominant to use the terminology of this paper) allele at the majority of loci. Superior here is defined as advantageous under conditions of natural selection. It has been suggested by Bowman (1958) that populations already at plateau level under selection for additive genetic variance would be suitable as the random mating populations in recurrent selection programmes. Plateau populations are theoretically expected to be homozygous for the superior allele at loci where gene action is additive. Therefore it seems reasonable to conclude that when plateau populations are used only loci where gene action is other than additive will contribute to the regressions being discussed. It must be stressed that no effort has been made to take account of epistasis and multiple allelism which could bias the value of the regressions.

With these considerations in mind it will be seen from examination of tables 5 and 6 that a negative regression will only be found in characters predominantly controlled by genes which are overdominant. This, however, does not imply the truth of the converse that if the regression is zero or positive, overdominance is absent. These conclusions regarding negative regressions are in agreement with those based on theoretical work on inbred lines by Hull (1952).

The value of the regression is dependent on and indicative of gene frequency. In practice, for characters which show overdominance, maximum performance would be achieved by obtaining 100 per cent. heterozygotes in the crossbred progeny. If the inbred tester is homozygous A_2A_2 then to obtain a negative regression the frequency of A_1 in the strain must be 0·7 or greater, except for very high degrees of overdominance. Improvement of hybrid mean performance will be obtained by changing the frequency of A_1 in the strain from $p = 0·7$ to $p = 1·0$. In that situation there will probably be less scope for improvement than if the tester is A_1A_1. A negative regression then means that the frequency of A_1 in the strain may be anywhere from $p = 0·9$ to $p = 0·1$. Maximum hybrid performance will be realised when $p = 0$ in the strain.

The performance of the strain itself will be affected by selection for hybrid performance. When the tester is homozygous recessive, strain performance will rise as hybrid performance increases, but when the tester is dominant the inverse relationship will hold. It may be that in the case where the tester is homozygous dominant natural selection would oppose any fall in strain performance caused by artificial selection for increased hybrid performance.

Two recurrent selection programmes have been carried out by Bowman (1958). In one case selection was for increased litter size in mice, and in the other for decreased number of abdominal chaetæ in Drosophila. Full details of these experiments will be described elsewhere. The mouse experiment had a design similar to that outlined in fig. 1, and was carried on for four cycles of selection. The regression of sires' crossbred progeny (X) litter size on their full sib (S) litter size was calculated for each cycle as well as a pooled estimate for the four cycles (see table 8). It can be seen that there is marked variation between cycles and from the pooled estimate in the value of the regressions. Since the mean litter size of S females varied

TABLE 8

Regression values of sires' crossbred progeny litter size on their full sisters' litter size, recurrent selection for litter size in mice

Cycle	Regression
1	-0.22 ± 0.11
2	0.00 ± 0.17
3	-0.22 ± 0.24
4	$+0.55 \pm 0.15$
Pooled estimate cycles 1-4	-0.02 ± 0.08

slightly between cycles it was decided to test the heterogeneity of the regressions with one another and at the same time adjust for the variation in mean. The method of Snedecor (1956) was used. There was no significant difference between adjusted means, but the separate cycle regressions were significantly heterogeneous.

The mean crossbred progeny litter size was increased by selection, and it seems probable that selection significantly altered the value of the regression at the same time. The fact that two of the regressions are negative, though the values are not significant, is slight evidence of overdominance. Further, if the regression has really changed from negative in the first three cycles to positive in the fourth, this indicates that dominant genes are probably at high frequency ($p = 0.7$-0.9) in the S strain. Such a situation offers ample scope for further improvement of hybrid performance.

The Drosophila experiment was carried on for fourteen cycles. Owing to the small numbers of individuals used, only three pooled estimates of the regression of sires' crossbred progeny bristle number on their pure strain progeny bristle number were calculated (see table 9). Though the regressions are not significant a marked decline in the observed value concomitant with a fall in the mean bristle number of both the crossbred and pure strain progeny was found. The inbred tester had previously been selected for low bristle number and was

consequently expected to be homozygous recessive at many loci. This considered in conjunction with the regression values suggests that overdominance for bristle number is negligible. Once again selection has probably altered the value of the regression, but in this case it is not so easy to make predictions about the changes in gene frequency in the pure strain.

Additive genetic variance can be more speedily exploited by mass or family selection methods within strains than by recurrent selection. It is therefore, useful to know from the first cycle of a recurrent selection programme whether there is any overdominance to exploit in the strain to be used for crossing to the inbred tester. It is seen that it is possible

TABLE 9

Regression values of sires' crossbred progeny bristle number on their pure strain progeny bristle number—recurrent selection for bristle number in Drosophila

Cycles	Regression
Pooled 1-5 . . .	$+0 \cdot 81 \pm 0 \cdot 60$
Pooled 6-10 . . .	$+0 \cdot 38 \pm 0 \cdot 89$
Pooled 11-14 . . .	$-0 \cdot 04 \pm 2 \cdot 03$

to detect overdominance, and that some predictions about expected response to recurrent selection can be made by the above approach.

4. SUMMARY

1. The values of two regressions—the relationship between the performance of a sire's testcross progeny and the performance of (i) his full sisters, or (ii) his pure strain progeny—are calculated by two methods in terms of gene frequencies and genotypic values for two alleles at a single locus.

2. It is pointed out that a negative regression can only be obtained when overdominance is present, though a positive or zero regression does not necessarily imply the absence of overdominance.

3. The possible use of these regressions in detecting overdominance and predicting progress in recurrent selection programmes is discussed.

4. Data on regressions calculated on two recurrent selection programmes, one for litter size in mice, and the other for bristle number in Drosophila are presented. With the theoretical findings in mind, a possible interpretation is placed on the regression values observed.

5. REFERENCES

BELL, A. E., MOORE, C. H., AND WARREN, D. C. 1955. The evaluation of new methods for the improvement of quantitative characteristics. *Cold Spring Harbor Symp. Quant. Biol., 20,* 197-212.

BOWMAN, J. C. 1958. Recurrent selection : an experimental study with mice and Drosophila. Ph.D. Thesis, University of Edinburgh.

COMSTOCK, R. E., AND ROBINSON, H. F. 1952. Estimation of average dominance of genes. Chapter 30 of *Heterosis,* ed. by J. W. Gowen. Ames, Iowa State College Press. 552 pp.

DICKINSON, A. G., AND JINKS, J. L. 1956. A generalised analysis of diallel crosses. *Genetics, 41,* 65-78.

FALCONER, D. S. 1959. *Introduction to Quantitative Genetics.* Oliver and Boyd, Edinburgh.

HAYMAN, B. I. 1954. The theory and analysis of diallel crosses. *Genetics, 39,* 789-809.

HULL, F. M. 1945. Recurrent selection for specific combining ability in corn. *J. Amer. Soc. Agron., 37,* 134-145.

HULL, F. H. 1952. Recurrent selection and overdominance. Chapter 28 of *Heterosis,* ed. by J. W. Gowen. Ames, Iowa State College Press. 552 pp.

MATHER, K. 1949. *Biometrical Genetics.* Methuen and Co. Ltd., London.

RASMUSON, M. 1956. Recurrent reciprocal selection. Results of three model experiments on Drosophila for improvement of quantitative characters. *Hereditas* (Lund), *42,* 397-414.

SNEDECOR, G. W. 1956. *Statistical Methods.* Fifth Edition. Ames, Iowa State College Press. 534 pp.