GENES OF LARGE EFFECT AND THE SHAPE OF THE DISTRIBUTION OF A QUANTITATIVE CHARACTER

By K. HAMMOND* and J. W. JAMES†

[*Manuscript received February* 23, 1970]

Summary

The shape of the frequency distribution of a quantitative character will be non-normal if genes of large effect make an appreciable contribution to the variance of that character. The skewness coefficient will be dependent upon the relative frequencies of high and low alleles, while the kurtosis coefficient is negative for intermediate and positive for extreme frequencies of these genes. Two special models are considered in illustration.

The frequency distributions of two characters (abdominal chaetae number and progeny number) in large panmictic populations of *Drosophila* were considered, using the technique developed here and one of Mérat (1968).

The results suggest that genes of large effect do not make a significant contribution to the variance of these characters in such populations. The sensitivity of the procedure is considered.

I. INTRODUCTION

The techniques of direct examination of loci affecting quantitative characters (Thoday 1961) and of the use of observable pleiotropy as suggested by Robertson (1967) appear most promising to the better understanding of genetic variation. But the problem is complex and these techniques are tedious and limited in their use to certain species (particularly *Drosophila*). Hence any simple alternative, be it statistical or experimental, and particularly if it has wide application, would be most useful.

The statistical technique of examining the shape of a frequency distribution of a quantitative character was suggested by Fisher, Immer, and Tedin (1932). This idea was extended by Mérat (1968) in an attempt to detect genes of large effect segregating in populations. He utilized measures of skewness and kurtosis, and discussed their advantages and limitations relative to present techniques for examining genetic variation at the gene level.

The present study examines the influence of segregating genes of large effect and their frequencies upon the expected distribution of family means for a quantitative character. This theory was used in an attempt to detect such genes in large randomly breeding populations of *Drosophila* by examining the skewness (g_1) and kurtosis (g_2) statistics (Fisher 1948, p. 75) of the distributions of two characters, family mean abdominal chaeta number (one sternite) and total number of progeny per family.

* Department of Animal Husbandry, University of Sydney, N.S.W. 2006.

† School of Wool and Pastoral Science, University of New South Wales, Kensington, N.S.W. 2033.

These characters were chosen for their ease of scoring and the presumed different kinds of gene action operating for each. At the statistical level only, the genetic variation of abdominal chaeta number in the populations studied is primarily additive (Sheridan *et al.* 1968; Hammond, unpublished data); while Barnes (1968) has demonstrated considerable directional dominance and duplicate interaction for yield of progeny.

II. Theory

Since we are concerned with genetic effects it is desirable that the methods be little affected by non-normality of non-genetic effects. For this reason family means were used, since the contribution of within-family variation to the means would be very nearly normally distributed by virtue of the central limit theorem, unless the family size is very small. We must then calculate the cumulants of the distribution of family means.

Consider a character controlled by n additive loci, the proportionate effect of the ith locus being a_i, and the frequency of the favoured allele at this locus being p_i in the base population. An individual whose gene frequencies are q_i ($= 0, \frac{1}{2}, 1$) will then have a genetic value $\sum_i q_i a_i$, the origin being taken as the genetic value of individuals homozygous for the "low" allele at all n loci. If a pair of individuals are chosen and mated, the mean frequencies of the "high" alleles among their progeny may be denoted x_i ($=0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$) and the mean genetic value of the progeny will be $\sum_i x_i a_i$.

Thus the family means are linear functions of the random variables x_i, so that cumulants of the family mean distribution may be calculated from those of x_i. Now for families produced by random mating of randomly chosen individuals the random variates x_i have binomial distributions, whose cumulants are known (Fisher 1948). The first four cumulants of the family means are readily shown to be:

(Mean) $$K_1 = \sum_i p_i a_i,$$

(Variance) $$K_2 = \tfrac{1}{4} \sum_i p_i (1-p_i) a_i^2,$$

(Skewness) $$K_3 = \tfrac{1}{16} \sum_i p_i (1-p_i)(1-2p_i) a_i^3,$$

(Kurtosis) $$K_4 = \tfrac{1}{64} \sum_i p_i (1-p_i)[1-6p_i(1-p_i)] a_i^4.$$

Now let

$$v_i = \tfrac{1}{2} p_i (1-p_i) a_i^2$$

be the genetic variance due to the ith locus. Then

$$K_2 = \tfrac{1}{2} \sum_i v_i,$$

$$K_3 = \tfrac{1}{8} \sum_i v_i a_i (1-2p_i),$$

$$K_4 = \tfrac{1}{32} \sum_i v_i (a_i^2 - 12 v_i).$$

It is clear that skewness is determined by the relative frequencies of high and low alleles—"directional gene frequencies".

The kurtosis coefficient K_4 will be negative if on average $a_i^2 < 12v_i$ or $p_i(1-p_i) > \frac{1}{6}$. The condition for this is that the loci which contribute most to variability have intermediate gene frequencies, about 0·2–0·8. For positive kurtosis it is required that at the loci making large contributions to variance $a_i^2 > 12v_i$, implying low or high frequencies for favoured alleles. Since such genes are supposed to produce large variances this implies that the gene effects must be appreciable. It is helpful to consider two special models in illustration.

In the model discussed by Latter (1960) $v_i = 2\mu_i/I$, where I is Haldane's measure of selection intensity and μ_i is the mutation rate at the locus. (This model concerns natural selection for an intermediate phenotype.) Assuming the same mutation rate at each locus, each has an equal contribution to the variance V and then the kurtosis coefficient is

$$\begin{aligned} g_2 &= K_4/K_2^2 \\ &= \tfrac{1}{2}[(\bar{a}^2+V_a)/V-(3/n)], \end{aligned}$$

where $\bar{a}$ and V_a are the mean and variance of the proportionate effects over n loci. Unless n is quite small the kurtosis should be negligible.

A second simple model is one in which there are n_1 loci with effect a_1 and n_2 $(=n-n_1)$ with effects a_2. The loci with effects a_1 have frequencies p_1 or $1-p_1$ and those with effects a_2 have frequencies p_2 or $1-p_2$. In the absence of directional gene frequencies $K_3 = 0$ and we have

$$K_2 = \tfrac{1}{4}[n_1p_1(1-p_1)a_1^2+n_2p_2(1-p_2)a_2^2],$$

$$K_4 = \tfrac{1}{64}\{n_1p_1(1-p_1)[1-6p_1(1-p_1)]a_1^4+n_2p_2(1-p_2)[1-6p_2(1-p_2)]a_2^4\}.$$

Now let $a_1 = ta_2$ and let

$$n_2p_2(1-p_2)a_2^2 = 4K_2/(1+C),$$

$$n_1p_1(1-p_1)a_1^2 = 4CK_2/(1+C).$$

On substitution we find

$$K_4 = \frac{K_2}{16(1+C)}\left[(Ct^2+1)a_2^2 - \frac{24K_2}{1+C}\left(\frac{C^2}{n_1}+\frac{1}{n_2}\right)\right].$$

The kurtosis coefficient g_2 is then

$$g_2 = \frac{1}{4n_2\,(1+C)^2}\left[\frac{1+Ct^2}{p_2\,(1-p_2)} - \frac{6(1+n_2\,C^2)}{n_1}\right].$$

If we regard the second group of loci as "common gene" loci with intermediate frequency, $p_2(1-p_2) \simeq \frac{1}{4}$ and

$$g_2 = [1/(1+C)^2][(Ct^2-\tfrac{1}{2})/n_2-(3C^2/2n_1)].$$

Here C is the relative contribution of "rare genes" to variance and t is the relative magnitude of effect of "rare genes". It is apparent from this formula that if the greater part of the genetic variance is due to genes of large effect at extreme frequencies, then positive kurtosis is to be expected, as pointed out in the general case.

The above theory has been developed in terms of the true family means. In practice we must work with the average of r observed individuals. This may be regarded as addition of an error, which will add another component to the variance but through the central limit theorem the higher cumulants of the error will be effectively zero. The second cumulant or variance will have an additional component $[(2-h^2)/rh^2]K_2$, where h^2 is the heritability.

III. Methods

(a) *Cultural*

The experiment comprised five parts—A, B, C, D, and E—of which part A formed the principal analysis.

(i) *Part A.*—Samples of single females were taken from a cage population of the Canberra strain of *D. melanogaster* (Latter 1964) that had been maintained in our laboratory at a population size of 3000–4000 adults for some 3 years. Each female was allowed to lay eggs in a separate vial. One offspring was collected from each female, and single-pair matings set up. The pairs of flies were transferred to a fresh vial after 48 hr, again after a further 24 hr, and then discarded 24 hr later. After discarding the progeny emerging in the 12-hr period including first emergence, 10 randomly chosen pairs of flies per vial were scored for abdominal chaeta number on the fourth sternite, making a total of 30 pairs per family. The total number of progeny per family was also scored. Part A was spread over three consecutive periods of time (Table 1).

Table 1

Experimental designs for parts A–E

Canberra stock used in parts A and B, Canberra *yellow scute* for parts C, D, and E

Part	Relative Time of Sampling (yr)	Sternite Scored for Number of Chaeta	No. of Each Sex per Family Scored	No. of Families
A	0	Fourth	30	196*
B	−2	Fifth	10	62
C	−0·5	Fourth and fifth†	25	50
D	0·2	Fourth and fifth†	10	61
E	0·9	Fourth and fifth†	20	68

* Composed of 57, 43, and 96 families at each time. All of each sex scored for progeny number for 192 of these 196 families.

† Different characters (sternites) in each sex—sternite scored in females placed before that scored in males.

All matings of all parts of the experiment were set up in 3 by 1 in. vials containing a dead yeast fortified medium (medium F of Claringbold and Barker 1961) and were maintained at 25±0·5°C, 65–70% relative humidity, and 12 hr light per day.

(ii) *Parts B, C, D, and E.*—To help detect any systematic scorer error and to add further to the results, data collected from several other experiments using abdominal chaeta number were

also analysed. All data of any one of these parts were collected at one time, but each part was sampled from the cage at different times. Further, the population, character scored, number scored per family, and number of families varied between parts (Table 1). No transferring of parents was used.

The data used in part B was kindly supplied by the authors of Sheridan *et al.* (1968), that used in part C by Mr. K. A. Rathie, and that of parts D and E were collected by the senior author. The Canberra *yellow scute* population used in parts C, D, and E was derived by repeated backcrosses from the same Canberra cage as that used for part A by Rathie (1969), and had been maintained at approximately 2000 adult flies for 3 months prior to taking the sample for part C. Table 1 also gives the relative time periods that cage sampling occurred for each part of the experiment.

(*b*) *Analytical*

The following analyses were carried out using the chaeta number data.

(i) *The Overall Analysis.*—All chaeta number data of any one part were pooled over families to obtain the shape of the distribution of individual scores.

(ii) *The Family Means Analysis.*—The distribution of family means was then examined, the g statistics being computed using two different variance values—(1) the unpartitioned family means analysis, which utilized the variance of the family means; and (2) the partitioned family means analysis. This latter analysis utilized the between family variance obtained by partitioning the variance into time, vial, and family components. The variance components between times and between vials were negligible. The errors of the normality statistics in the latter analysis will now contain an additional component and standard tests of significance will now be invalid. Instead t-tests were applied firstly to test the sample means for sex differences and, secondly, where the test of sample means was not significant, to test the pooled mean for difference from zero.

(iii) *Mérat's Method.*—Mérat (1968) suggested an alternative method for detecting genes of large effect in a population, which also utilized measures of skewness and kurtosis. After testing for homogeneity of family variances, the deviations from the family mean for those families having extreme high or low variances are pooled to form high and low variance groups, and the normality statistics of these two groups are then compared. The high variance groups are expected to depart from normality, exhibiting asymmetry and platykurtosis, if genes of large effect are segregating in the population. This analysis also was applied to the chaeta number data.

(iv) *The Pooled Analysis.*—Following each analysis a χ^2 test of equality was applied to the results over all parts of the experiment and where this was not significant ($P = 0{\cdot}05$) the weighted means of the g statistics were used as estimates of departure from normality. In the unpartitioned family means analysis the standard test of significance on these weighted means remains valid for they are still distributed about zero. However, their interpretation is not direct as family size varied between parts, and the g statistics for this analysis are expected to change with change in family size. The value obtained by subtraction of any two of these means and then division by a weighted estimate of the standard error is normally distributed; thus they can be tested for difference. This method is not entirely valid as we are using the same set of data but the means would probably be quite independent due to the different variance components utilized in their derivation.

The analysis of progeny number was performed on the original data, then on these data less the zero observations which may have arisen from technical or inherited faults, and then on this latter group of data after transforming to remove skewness. Progeny number, though a family characteristic, is not a family average, and the central limit theorem could not be expected to remove non-normality due to non-genetic effects. The theory of Section II is thus not strictly applicable, but it seemed worth while to apply the method to a trait with a genetic control so different from that of chaeta number. It is hoped that transformation may help to remove non-genetic causes of departure from normality. A number of transformations were tried. The power function of $2{\cdot}5$ proved the most suitable, for the three groups of data could still be legitimately pooled and the weighted mean g_1 value did not differ significantly from zero. Only the variance of the family total progeny number could be used to derive the g values for the distribution of this character which were then pooled over the three times after a non-significant ($P = 0{\cdot}05$) χ^2 test of equality was obtained.

IV. Results

The skewness and kurtosis values for all analyses of male and female chaeta number and for progeny number on two scales are given in Table 2.

Table 2

Measures of skewness (g_1)† and kurtosis (g_2)† for chaeta number (analyses 1–4) and progeny number

Analysis	Part	g_1			g_2		
		Male	Female	S.E.	Male	Female	S.E.
1. Overall	A	0·17*	0·11*	0·03	0·08	0·20*	0·07
	B	−0·15	−0·10	0·10	0·11	0·41*	0·20
	C	0·31*	0·13	0·07	0·22	0·22	0·14
	D	0·45*	0·14	0·10	0·25	0·02	0·20
	E	0·34*	0·17*	0·07	−0·03	0·04	0·13
2. Family means							
(i) Using unpartitioned variance	A	−0·05	−0·02	0·17	−0·42	−0·05	0·35
	B	−0·52	−0·15	0·30	0·23	−0·78	0·60
	C	−0·48	0·09	0·34	−0·90	−0·53	0·66
	D	0·36	0·18	0·31	−0·62	0·05	0·60
	E	0·36	0·30	0·29	0·42	−0·28	0·57
(ii) Using family variance component	A	−0·07	−0·03		−0·65	−0·06	
	B	−0·81	−0·22		0·41	−1·30	
	C	−0·67	0·16		−1·41	−1·06	
	D	0·70	0·33		−1·49	0·11	
	E	0·49	0·37		0·62	−0·38	
3. Mérat‡	A H	0·11	0·14	0·10	−0·19	0·24	0·20
	A L	0·15	0·09		−0·24	−0·23	
	B H	−0·22	−0·36	0·31	−0·68	0·77	0·61
	B L	−0·12	0·27		−0·45	−0·65	
	C H	0·39*	0·25	0·19	0·33	−0·04	0·39
	C L	0·08	−0·09		0·06	−0·03	
	D H	0·44	0·08	0·31	0·48	−0·25	0·61
	D L	−0·06	−0·63*		−0·74	0·44	
	E H	0·42	0·09	0·22	−0·09	−0·25	0·44
	E L	0·13	0·39		−0·56	0·24	
4. Pooled 1		—	0·11*	0·03	0·10*‖	0·18*	0·05
2 (i)		−0·04	0·05	0·12	−0·28	−0·23	0·23
(ii)§		0·03			−0·52		
3 H		0·19*	0·12	0·08	−0·08	0·14	0·15
3 L		0·11	0·06		−0·28‖	−0·13	
Progeny number							
Original		−1·25*		0·17	1·48*		0·34
Transformed		−0·19			−0·14		

* $P = 0·05$.

† Expected to be zero for a normal population.

‡ H and L represent the high and low variance group respectively.

§ Pooled over sexes after obtaining non-significant t-test ($P = 0·05$) of sample means.

‖ Indicates a significant t-test ($P = 0·05$).

All but one group (viz. the male g_1 values of the overall analyses) of the chaeta number results could be pooled and only the pooled estimates will be examined. Male kurtosis tends to be more negative than female in all analyses. The overall analysis did show small but significant departure from normality, all g values being positive. All results of the family means analyses did not differ significantly from zero nor did they show significant difference from those of the overall results, although the kurtosis of the former analysis tended to negative values (for example, $-0{\cdot}28 \pm 0{\cdot}23$ compared with $0{\cdot}10 \pm 0{\cdot}05$ respectively, for males).

A Bartlett's test of homogeneity of family variances using the theoretical distribution of $(1+\frac{1}{2}g_2)\chi^2_{k-1}$ suggested by Box (1953), was not significant for any one part. The method of Mérat (1968) was applied separately to the data of each part. Results did not follow the trend expected in the presence of genes of large effect. The g_1 values showed small change (e.g. $0{\cdot}19$ and $0{\cdot}11 \pm 0{\cdot}08$ for males high and low respectively) between the high and the low variance groups, while the g_2 values of the high variance groups were *not* more negative than those of the low variance groups (for example, $-0{\cdot}08$ and $-0{\cdot}28 \pm 0{\cdot}15$ for males high and low respectively). On the contrary the low group g_2 value for males differed significantly from that of the overall analysis.

In comparing the two methods of analysis, that of the family means produced results tending to less asymmetry and more platykurtosis than did that by Mérat's method. None of these differences were significant.

King (1963) observed leptokurtosis in the frequency distributions of egg hatching time, a fitness component, in heterogeneous randomly breeding populations of *D. melanogaster*. In the present case the progeny number distribution on the original scale exhibited significant negative skewness and positive kurtosis. Removal from the data of those observations of zero progeny did not appreciably change the results ($-1{\cdot}32$ to $-1{\cdot}25$ for g_1 and $1{\cdot}51$ to $1{\cdot}48$ for g_2). However, the kurtosis was removed ($-0{\cdot}14 \pm 0{\cdot}34$) by transforming to eliminate the skewness.

The graphic form of the distributions of data of part A is given in Figure 1. Those of the remaining parts of the experiment were not very informative, each having a relatively small number of observations, and are not presented. The progeny number data less the zero observations are given using both untransformed and transformed scales.

V. Discussion

Over both methods of analysis homogeneity of family variances and the g statistics for chaeta number suggest that we failed to detect genes of large effect segregating for this trait in the base population. This occurred despite indirect evidence from selection experiments that such genes apparently influence abdominal chaeta number (Jones 1967; Robertson 1967; Frankham 1969; Rathie, personal communication).

Assuming the method is of adequate sensitivity to detect these genes, the above discrepancy may be accounted for by postulating a change in the relative magnitude of effect of the genes concerned as the background genotype changes. Thus, potentially large genes would have relatively small effects in the base population and would contribute little to the variance of this population. The genetic variance

would be determined by a large number of genes of relatively small effects and the kurtosis would tend to zero.

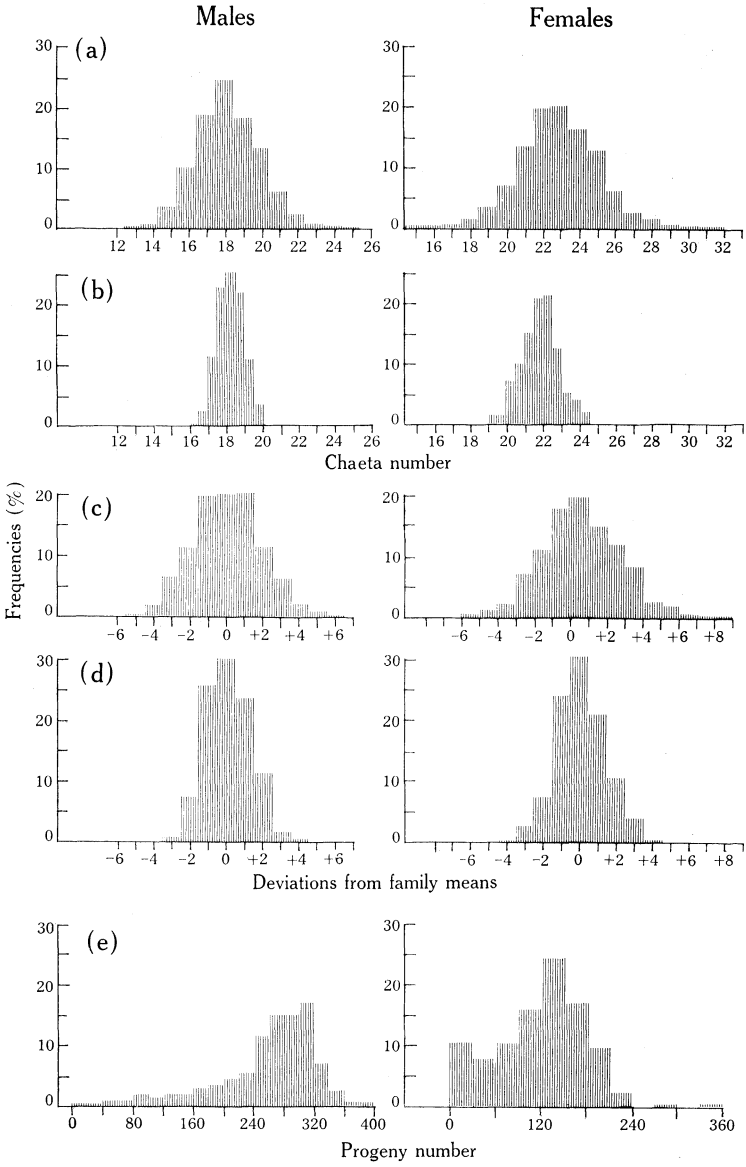

Fig. 1.—(*a*), (*b*) Distributions of male and female chaeta number as calculated by the overall (*a*) and family means (*b*) analyses. (*c*), (*d*) Mérat's analysis for high (*c*) and low (*d*) family variances. (*e*) Distributions of progeny number less zero observations. Untransformed values (left); transformed values (right). Transformed values = (untransformed values)$^{2\cdot5} \times 10^{-4}$.

The sensitivity of the methods used is somewhat uncertain for the traits studied. Heritability estimates for abdominal chaeta number in the Canberra stock are quite

low—16% (Sheridan *et al.* 1968) on wild type, to 20% (Hammond, unpublished data) on *y sc* Canberra. This, together with the effect of any non-random environmental component, could influence the analyses considerably.

Non-additive variation, linkage, and natural selection would also affect the g statistics but the direction of influence cannot easily be predicted. Fisher, Immer, and Tedin (1932) demonstrated the usefulness of third-degree statistics in detecting directional dominance. Substantial directional dominance and duplicate interaction in progeny number has been demonstrated by Barnes (1968). Thus the observed significant g statistics for this character in the present study are probably not the result of large gene influences. Scale effects for abdominal chaeta number are negligible in the Canberra stock and any that were present would be reduced by dealing with family means. The effect of scale on the interpretation of the progeny number results cannot be determined but removal of the skewness also removed the observed leptokurtosis.

In the chaeta number analysis there is no consistent skewness, which is to be expected in the absence of directional gene frequencies. However, directional gene frequencies cause easily detectable skewness only if the genes are of large effect. In the absence of evidence for genes of large effect the skewness data are not readily interpreted.

In conclusion, it would appear that the use of higher-degree statistics for detection of segregating genes of large effect on traits of low heritability is questionable. However, the method may prove more useful on traits which demonstrate a greater proportion of genic variation.

VI. Acknowledgments

We wish to thank Associate Professor J. S. F. Barker for constructive criticism of the manuscript and are grateful to Mr D. H. Miller for providing the necessary computer programs. The work was carried out whilst K. H. was the holder of a Commonwealth Postgraduate Research Studentship.

VII. References

Barnes, B. W. (1968).—Maternal control of heterosis for yield in *Drosophila melanogaster*. *Heredity* **23**, 563–72.

Box, G. E. P. (1953).—Non-normality and tests on variances. *Biometrika* **40**, 318–35.

Claringbold, P. J., and Barker, J. S. F. (1961).—The estimation of relative fitness of *Drosophila* populations. *J. theor. Biol.* **1**, 190–203.

Fisher, R. A. (1948).—"Statistical Methods for Research Workers." 10th Ed. (Oliver and Boyd: Edinburgh.)

Fisher, R. A., Immer, F. R., and Tedin, O. (1932).—The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, Princeton* **17**, 107–24.

Frankham, R. (1969).—Genetical analyses of two abdominal bristle selection lines. *Aust. J. biol. Sci.* **22**, 1485–95.

Jones, L. P. (1967).—Effects of X-rays on response to selection for a quantitative character of *Drosophila melanogaster*. *Genet. Res.* **9**, 221–31.

King, J. C. (1963).—The fourth moment of a character distribution as an index of the regulative efficiency of the genetic code. In "Biological Organization". pp. 129–46. (Academic Press, Inc.: New York.)

LATTER, B. D. H. (1960).—Natural selection for an intermediate optimum. *Aust. J. biol. Sci.* **13**, 30–5.

LATTER, B. D. H. (1964).—Selection for a threshold character in *Drosophila*. I. An analysis of the phenotypic variance on the underlying scale. *Genet. Res.* **5**, 198–210.

MÉRAT, P. (1968).—Distributions de fréquences, interprétation du déterminisme génétique des caractères quantitatifs et recherche de 'gènes majeurs'. *Biometrics* **24**, 277–93.

RATHIE, K. A. (1969).—Faster scoring of a quantitative trait of *Drosophila melanogaster*. *Drosophila Inf. Serv.* **44**, 104.

ROBERTSON, A. (1967).—The nature of quantitative genetic variation. In "Heritage from Mendel". (Ed. R. A. Brink.) pp. 265–80. (University of Wisconsin Press: Madison, Wisconsin.)

SHERIDAN, A. K., FRANKHAM, R., JONES, L. P., RATHIE, K. A., and BARKER, J. S. F. (1968).—Partitioning of variance and estimation of genetic parameters for various bristle number characters of *Drosophila melanogaster*. *Theor. appl. Genetics* **38**, 179–87.

THODAY, J. M. (1961).—Location of polygenes. *Nature, Lond.* **191**, 368–70.