# STUDIES IN ARTIFICIAL SELECTION OF QUANTITATIVE CHARACTERS

## I. SELECTION FOR ABDOMINAL BRISTLES IN DROSOPHILA MELANOGASTER

By B. L. SHELDON*

*Summary*

Genetic and phenotypic parameters for number of abdominal bristles were estimated from diallel crosses in a wild-type laboratory stock, Oregon-RC. Two replicate mass selection lines in both high and low directions were developed and run for over 30 generations. The same parameters were estimated from diallel crosses at intervals during the experiment.

In the base population, heritability was 20–25% and non-additive genetic variability was negligible. In general, neither the base population nor later parameters predicted adequately the subsequent selection responses. Errors of estimate of the later genetic parameters were large. Additive genetic variability was reduced to low or zero levels and non-additive genetic variability increased in all lines. Responses to selection occurred in apparent absence of additive genetic variability. Three out of the four lines reached a plateau at about generation 30–33, when a residue of non-additive variance remained.

Differences between replicate lines were small in the early generations, but relatively large in the long term. Many other aspects of the replicate lines also became different in the long term.

The results are discussed in relation to quantitative genetics theory and in comparison with previous selection experiments on this character.

## I. INTRODUCTION

In theory accurate predictions of the changes that will result from selective breeding can be calculated from the heritabilities of the selected characters and the genetic correlations between them (viz. Lush 1945, 1948; Lerner 1950, 1958). To date the most complete series of selection experiments designed to find out how closely selection lines follow theoretical predictions has been that reported by Clayton, Morris, and A. Robertson (1957), Clayton and A. Robertson (1957), and Clayton *et al.* (1957). Heritability of abdominal bristle number in *Drosophila melanogaster* was estimated in their base population before selection began by the methods of parent–offspring regression, half-sib correlation, and full-sib correlation. Different intensities of selection and different systems of selection and mating were used, and the repeatability of response to selection in replicate lines from the same base population was studied.

The *average* response of high selection lines to individual or family selection at different intensities agreed well with prediction over the first five or so generations. The low lines, however, responded at a considerably lower rate than expected from the beginning of the experiment. Departure from expectation seemed greatest where intensity of selection was least.

* Division of Animal Genetics, C.S.I.R.O., Zoology Department, University of Sydney.

The good early agreement of average observed with expected response in the high lines of the long-term experiment was interpreted by the authors and others (F. W. Robertson 1956; Falconer 1960) as reason for confidence in predictions from heritability, at least for the first few generations of selection. The poor repeatability between lines, however, may equally well be taken as a reason for distrusting predictions. The great variety of different situations that arose in the long term led the authors to conclude that predictions from heritability broke down completely after many generations of continued selection. Heritabilities were not estimated again in their lines, while the response to selection continued, so that it was the predictive value of heritability estimated at the beginning of the experiment only which was tested. At the end of the experiment estimated heritabilities were usually high in spite of cessation of response.

Falconer's (1953, 1955) selection experiment with body weight in mice also showed considerable asymmetry of response between high and low lines. The response in either direction was different from that predicted from heritability, but the rate at which high and low lines were separated by selection agreed well with estimated heritability. Falconer (1954) also tested the predictive value of genetic correlations, using body weight and tail length in mice as the correlated characters. In this short-term experiment the responses followed expectations quite well.

In rats, Kyle and Chapman (1953) selected for ovarian weight and obtained less response than predicted. Again, there was asymmetry of response, the deviation from expectation being more marked in the low direction.

A considerable literature exists on the results of selection in domesticated animals. Very few reports allow a significant comparison between actual responses and previously estimated heritabilities and genetic correlations. Some reports suggested reasonable agreement (e.g. Lerner and Hazel 1947; Lerner and Dempster 1951; Dempster, Lerner, and Lowry 1952; see also Lerner 1958 for a general review), but the results were at least inconclusive, and perhaps equally open to the opposite interpretation (see Kyle and Chapman 1953). Dickerson (1955) reported considerable departure from prediction. Here again interpretation is open to question, particularly where early selection is considered.

It is clear both from the small number of experiments reported in this field, and from the conflicting interpretations possible, that further experimental tests of the theory of quantitative genetics are in order. This view is reinforced by the recurring emphasis on heritability in the applied field (cf. Lush 1945; Lerner 1950, 1958; Rendel and A. Robertson 1950; A. Robertson and Rendel 1950; Morley 1951, 1955*a*, 1955*b*; Rendel, A. Robertson, and Alim 1951; Turner 1956, 1958, 1959; Doney 1958; Dun 1959; and Falconer 1960, to mention but a few authors).

The selection experiments to be reported in this and following papers are similar to the series of Clayton and colleagues. The common conception of the additive genetic model is that the additive genetic variance becomes reduced under continued selection, and that genetic interactions of one form or another increase in importance, usually as the selection response slows to a plateau (Lerner 1950, 1958; Reeve and F. W. Robertson 1953; F. W. Robertson and Reeve 1953; A. Robertson 1955; Clayton and A. Robertson 1957; Dun 1959; Falconer 1960). It was decided therefore,

to study trends in the non-additive genetic variance as well as changes in heritability. Heritability and genetic interaction (specific combining ability) are estimated in the base population and at intervals during selection by a form of diallel analysis appropriate to a population under random mating.

Since the relationship between a particular character and fitness has been postulated to be of prime importance in determining the character's response to selection (A. Robertson 1955), it was decided to study two different characters, abdominal bristle number and body weight. The former has been considered as a peripheral character with no direct effect on fitness, while body weight can be expected to have at least a moderate relationship to fitness. The genetic determination of the two characters should be correspondingly different. The abdominal bristle character might be expected to have a relatively high heritability with little non-additive genetic variance in an unselected population. Body weight should have lower heritability and considerable non-additive genetic variance (A. Robertson 1955).

The present paper deals with selection for abdominal bristles over the first 33 generations. Body weight selection over about 40 generations is reported in Part II of this series (Sheldon 1963).

## II. Methods

This series of selection lines came from an Oregon-RC wild-type stock, maintained at $25 \pm 0 \cdot 5$°C in 5-oz bottle cultures containing a standard semolina–treacle–agar medium seeded with live yeast suspension. These culture methods were also used for the selection lines.

Four selection lines were carried, two high ($H_1$ and $H_2$) and two low ($L_1$ and $L_2$). The sum of the bristles on the 4th and 5th abdominal sternites was the criterion of selection. For the greater part of the experiment, five bottles per line were set up in each generation. The parents were left in the bottles for 5 days. Progeny emerging from the 10th to the 12th days inclusive after cultures were set up were collected twice daily to ensure virginity of females, and then scored on the 13th and 14th days. 20 females and 20 males were scored per bottle and the extreme 4 out of the 20 scored were saved. The selected individuals from all five bottles in each line were then pooled and set up at random, 4 pairs to a bottle, at the end of the 14th day.

To produce the first selection generation ($S1$), the 20 pairs of selected parents from the foundation population were used in one bottle and transferred to fresh bottles on succeeding days, the first three of these bottles per line being used for scoring. This technique provided cultures that were far too crowded and it was discarded in favour of the one outlined above. In the 19th generation ($S19$) the technique of seeding each bottle with equal numbers of eggs (about 70), from a collection laid on a suitable surface by the 20 pairs of selected parents, was tried, but this method gave such variable and low hatchability that it too was discarded. It was, however, used with more uniform results in a parallel series of body weight selection lines (Sheldon 1963).

Diallel crosses were made in the foundation population and from each selection line at $S13$, $S23$, and $S33$ to estimate heritability and to see if there were any trends in the non-additive portion of the genetic variance. Only the fifth segment was scored

in these diallels. The form of the diallel was a two sires–two dams block of matings, replicated as often as possible, as described by Lerner (1950). Jerome, Henderson, and King (1956) utilized a more general polyallel in a poultry experiment. Lerner (1958) summarized the constitution of the variance components from the diallel form of analysis in terms of additive and non-additive portions of the genetic variance. This is repeated in Table 1 for ease of reference.

Thus heritability estimated from $\sigma_s^2$ or $\sigma_d^2$ will include one-quarter of the additive × additive variance, and, if this source of variability is important, will be an incorrect assessment of the level of additive genetic variance or heritability in the narrow sense. The interaction term ($\sigma_I^2$) includes one-quarter of the dominance deviations, between $\frac{1}{8}$ and $\frac{1}{16}$ of the first-order epistatic deviations, and smaller fractions of the higher-order epistatic deviations. Trends in $\sigma_I^2$ in the course of selection should provide an approximate guide to changes involving the main sources of non-additive genetic variation.

TABLE 1

CONSTITUTION OF THE VARIANCE COMPONENTS IN A DIALLEL ANALYSIS IN TERMS OF ADDITIVE AND NON-ADDITIVE PORTIONS OF THE GENETIC VARIANCE

| Kind of Genetic Variance | Variance Components | | | |
|---|---|---|---|---|
| | Within Full Sibs | Between Maternal Half-sibs ($\sigma_d^2$) | Between Paternal Half-sibs ($\sigma_s^2$) | Interaction ($\sigma_I^2$) |
| Additive | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | — |
| Dominance | $\frac{3}{4}$ | — | — | $\frac{1}{4}$ |
| Epistasis | | | | |
| Additive × additive | $\frac{3}{4}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{8}$ |
| Additive × dominance | $\frac{7}{8}$ | — | — | $\frac{1}{8}$ |
| Dominance × dominance | $\frac{15}{16}$ | — | — | $\frac{1}{16}$ |

Since in the diallels a second male had to be used on each of the females, a small preliminary test was made with Oregon females and males from the same stock, as well as from three different dominant mutant stocks, to see how soon after the first mating a second mating could be tested. The results indicated that a period of at least 1 week was needed between the first and second matings to be reasonably sure of obtaining a sample of progeny sired only by the second male on each female. In practice the first male was removed from the female after 2 days. The second male was introduced to the female in a storage vial about 5 or 6 days afterwards. This second mating pair was then transferred to a fresh culture 2 days later to obtain a brood for scoring. The diallel crosses in the base population were done in bottle cultures, while those during the selection programme were done in 3 in. by 1 in. vials.

Phenotypic variability between flies within bottles, and internal variability between the fourth and fifth segments and the phenotypic correlation between segments, were examined each generation.

## III. Results

### (a) *Character in the Base Population*

The *means and variances* for the character in the base population are given in Table 2. There is close agreement between the means obtained from different sources. The variance estimates show more variation according to the source of the estimate than do the means. It is of some interest that the largest estimates of the phenotypic variance were those derived from the uniformly uncrowded C cultures.

TABLE 2

MEANS AND WITHIN-CULTURE VARIANCES OF THE STERNITAL BRISTLE CHARACTER IN THE BASE POPULATION

| | A* | | B† | | C‡ | | D§ | |
|---|---|---|---|---|---|---|---|---|
| | Females | Males | Females | Males | Females | Males | Females | Males |
| 4th and 5th sternite | | | | | | | | |
| Means | 41·32 | 32·28 | 40·64 | 31·94 | 40·71 | 33·05 | | |
| Variances | 8·05 | 5·55 | 8·78 | 6·60 | 9·39 | 7·42 | | |
| 4th sternite | | | | | | | | |
| Means | 20·65 | 15·69 | 20·18 | 15·68 | 20·30 | 16·22 | | |
| Variances | 3·06 | 2·33 | 3·29 | 2·62 | 4·27 | 2·96 | | |
| 5th sternite | | | | | | | | |
| Means | 20·66 | 16·60 | 20·45 | 16·26 | 20·41 | 16·83 | 19·87 | 16·37 |
| Variances | 3·47 | 2·63 | 3·35 | 2·33 | 3·87 | 2·83 | 3·00 | 2·85 |
| Variance of difference between sternites ($\sigma_D^2$)‖ | 2·12 | 1·72 | 1·97 | 1·38 | 3·34 | 1·97 | | |

* Based on 20 bottle cultures, each with one pair of parents; 10 females and 10 males scored from each bottle; selection lines derived from this source.

† Based on 20 bottle cultures, each with four pairs of parents; 10 females and 10 males scored per bottle.

‡ Based on 16 bottle cultures, each seeded with about 70 random eggs; 10 females and 10 males scored per bottle.

§ Base population diallel crosses; 276 bottles, 5 females and 5 males scored per bottle.

‖ 4th sternite minus 5th sternite.

The phenotypic variances are of the same order of magnitude as observed by previous workers on other populations, rather lower than reported by Clayton, Morris, and A. Robertson (1957), and slightly higher than some inbreds and $F_1$'s given by Reeve and F. W. Robertson (1954) and by Rasmuson (1952).

The variance of the difference between sternites, $\sigma_D^2$, is considerably smaller in this population than in the populations observed by Clayton, Morris, and A. Robertson, and by Reeve and F. W. Robertson. $\sigma_D^2$ also varies appreciably with different culture conditions. As with the phenotypic variances, the uniformly uncrowded cultures (C) gave the highest $\sigma_D^2$ values.

The *phenotypic correlation* between the 4th and 5th sternites is, of course, closely related to $\sigma_D^2$. The correlations calculated from the C cultures of Table 2 were therefore the lowest, 0·44 for females and 0·45 for males. From the A cultures the estimates were 0·58 and 0·53, and from the B cultures 0·63 and 0·65 for females and males respectively.

TABLE 3

ANALYSES OF VARIANCE OF THE DIALLEL CROSSES IN THE BASE POPULATION

| Source of Variation | Degrees of Freedom | Mean Squares | |
|---|---|---|---|
| | | Females | Males |
| Sires | 69 | 4·29 | 4·16 |
| Dams | 69 | 4·52 | 6·16 |
| Interaction | 69 | 3·24 | 2·46 |
| Residual | 1104 | 3·00 | 2·85 |
| $h_s^2$ (from sires term) | | 0·13±0·09 | 0·16±0·09 |
| $h_d^2$ (from dams term) | | 0·16±0·09 | 0·40±0·12 |

The full-sib families of the A cultures also provided a very rough estimate of the *genetic correlation* between the two sternites. The values obtained were 1·13 for females and 1·52 for males. Since only 20 full-sib families were used, the statistical

TABLE 4

FULL-SIB ANALYSES OF VARIANCE OF CULTURES FROM WHICH SELECTION BEGAN

| Source of Variation | Degrees of Freedom | Mean Squares (females) | | Mean Squares (males) | |
|---|---|---|---|---|---|
| | | 4th Sternite | 5th Sternite | 4th Sternite | 5th Sternite |
| Between families | 19 | 5·23 | 6·35 | 5·55 | 5·09 |
| Within families | 180 | 3·06 | 3·47 | 2·33 | 2·63 |
| Heritability ($h^2$) | | 0·13 | 0·15 | 0·24 | 0·17 |

sampling error of these estimates is large. They merely indicate that this population does not differ from those of Reeve and F. W. Robertson, and Clayton, Morris, and A. Robertson, where the genetic correlations were practically 100%.

Analyses of variance of data from the diallel crosses in the base population, and the *heritabilities* derived from them, are given in Table 3. Table 4 gives the analyses

of variance and heritability estimates for the full-sib family structure of the cultures from which selection actually began, and from which the genetic correlation between sternites was estimated (i.e. the A cultures of Table 2).

The statistical sampling error of all the heritability estimates is large. However, except for the $h_d^2$ terms of Table 3, they are in close agreement with each other. It seems reasonable to accept a value of 0·15–0·20 as the heritability, in the narrow or additive sense, of the bristle score of a single sternite in this stock.

Since the genetic correlation between sternites seems to be unity, and the variance and heritability for each sternite are the same, the bristle numbers on each sternite can be regarded as repeated measurements of the same character with error

TABLE 5

FULL-SIB ANALYSES OF VARIANCE OF CULTURES FROM WHICH SELECTION BEGAN, USING SUM OF 4TH AND 5TH STERNITES

| Source of Variation | Degrees of Freedom | Mean Squares | |
|---|---|---|---|
| | | Females | Males |
| Between families | 19 | 18·76 | 19·78 |
| Within families | 180 | 8·05 | 5·61 |
| Heritability ($h^2$) | | 0·23 | 0·40 |

peculiar to each sternite causing the difference between them. The heritability of the total score of two or more sternites can then be calculated as $n/\{1+(n-1)r\}$ times the heritability of a single sternite, where $n$ = number of sternites scored, and $r$ = the repeatability, or phenotypic correlation, between sternites (see Lerner 1958, p. 186).

With $r = 0{\cdot}60$ approximately, and $n = 2$, in this instance, the heritability of the total score is therefore 1·25 times that of a single sternite, that is 0·19–0·25 compared with 0·15–0·20 for a single sternite. This is supported by the full-sib analysis of female progeny shown in Table 5. The higher value of heritability obtained from male progeny in this table, a similar but smaller trend in Table 4, and the higher $h_d^2$ term in Table 3 may be due to sex-linked effects. Heritability in this population is much smaller than in the populations of Reeve and F. W. Robertson, or Clayton, Morris, and A. Robertson.

Finally, the genetic interaction term in Table 3 is negligible, indicating that non-additive genetic effects do not occur or are at least insignificant in the base population. This agrees with the findings of the above authors. The following

tabulation shows the relative contributions of the different sources of variation in the present population compared with Clayton, Morris, and A. Robertson:

| Source of Variation | Clayton, Morris, and Robertson (1957) | Present Population |
|---|---|---|
| Additive genetic | 0·52 | 0·20–0·25 |
| Non-additive genetic | 0·09 | 0 –0·05 |
| Developmental error (difference between sternites) | 0·35 | 0·25 |
| Environmental error (common to both sternites) | 0·04 | 0·45–0·55 |

### (*b*) *Selection Results*

(i) *Response of Mean Bristle Number*

Figures 1 and 2, for females and males respectively, show the changes in mean number of bristles as selection proceeded. The aberrant results in the first generation (*S*1) need to be considered first, since all lines, high and low, declined by 3–4 bristles from the base population mean. This was undoubtedly due to extreme crowding of the cultures, as indicated in Section II. Adults began to emerge only on the 12–13th day from these cultures at 25°C instead of the usual 9–10th day. There is no obvious explanation why the mean score in all lines was still below the base population level at generation *S*2. High and low lines continued to diverge in *S*2, as in *S*1. Since the crowding presumed responsible for depression of the *S*1 mean did not occur in *S*2, some other unknown environmental factor could have been responsible, perhaps a rise in temperature in the constant-temperature room in that generation. The available laboratory records give no further evidence for this period of time. The depressed scores of *S*1 and *S*2 can be neglected in the following consideration of average responses during different periods of the experiment.

Agreement between replicate lines in the short term is excellent. The low lines do not deviate from each other consistently until *S*16. Agreement between high lines is almost as good to about *S*8, not as close between *S*8 and *S*12–*S*13, and poor after this. Repeatability between replicates seems to be higher and concordance lasts over a longer period than in the population used by Clayton and colleagues, but this may be partly a chance effect due to the minimum replication in the present instance. The predicted response can be calculated from the heritability (approx. 20%), the selection differential (1·4 standard deviations), and the known standard deviations in the base population (3·0 bristles for females and 2·6 bristles for males). The expected rate of progress at the beginning of the experiment was therefore $0{\cdot}2\times1{\cdot}4\times3{\cdot}0 = 0{\cdot}84$ bristles per generation for females, and $0{\cdot}2\times1{\cdot}4\times2{\cdot}6 = 0{\cdot}73$ bristles per generation for males. Table 6 shows the average observed responses compared with the expected response in two periods, the first five generations and the first 10 generations. The divergence between $H_1$ and $H_2$ at *S*10 is ignored in the present comparison. These data show in specific terms the trend that was already apparent in Figures 1 and 2, namely that the low lines responded at a faster rate than the high lines in this period. The high lines achieved only about half the

predicted response, although generations *S5–S10* showed a slight increase in rate of response over *S1–S5*. The low lines responded at up to $1\frac{1}{2}$ times the predicted rate

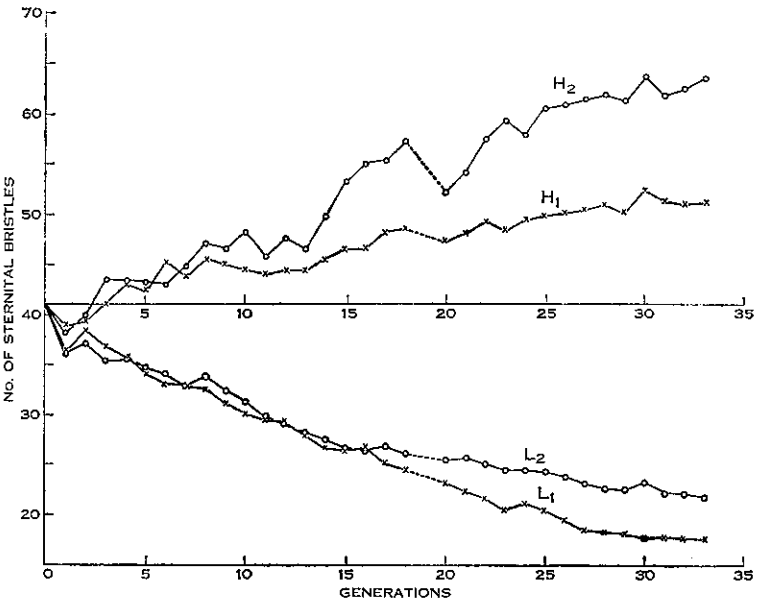


Fig. 1.—Response to selection for sternital bristles—females.

for five generations before slowing down. The asymmetry of response is in the opposite direction to, and of greater magnitude than, that reported by Clayton, Morris, and A. Robertson (1957).

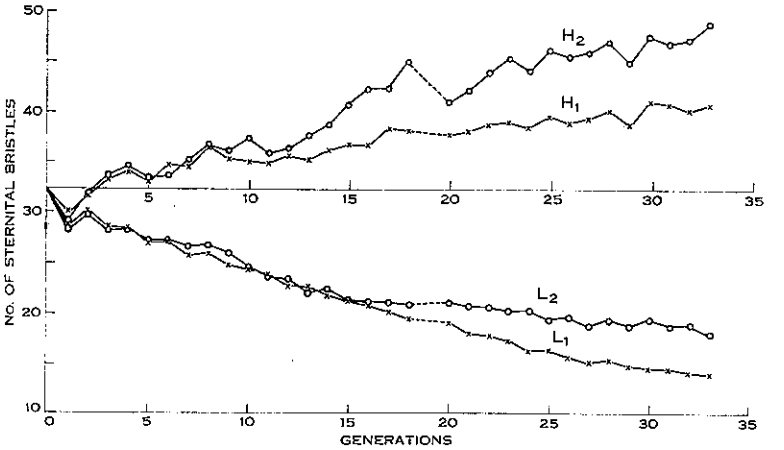


Fig. 2.—Response to selection for sternital bristles—males.

In the long-term view each line pursues an individual course so that long-term repeatability of selection response is extremely low, which agrees with the findings

of the above authors. The high lines diverge earlier and to a greater extent than the low lines. $H_1$ proceeds at about the same rate for the whole period but $H_2$ has an accelerated response after *S*13 and finishes with about twice the absolute response of $H_1$. $L_2$ makes less selection gain than $L_1$ after *S*16 and continues at this slow rate till *S*33.

The effects of *relaxing selection* were not studied directly during the period covered by this paper. However, the generation of enforced relaxation at *S*19 was accompanied by considerable regression towards the mean of the base population in line $H_2$. The effect in $H_1$ was mild and in the low lines negligible. It is probably significant that $H_2$ was the line with an increased variance and the most rapid rate of response at the time relaxation occurred.

Before a more detailed assessment of the individual lines is attempted, data will be presented on changes in the different components of the phenotypic variance.

TABLE 6

COMPARISON OF OBSERVED WITH PREDICTED RESPONSE IN EARLY STAGES OF EXPERIMENT

| | Predicted Response per Generation | | Observed Response per Generation | | | |
|---|---|---|---|---|---|---|
| | | | *S*1–*S*5 | | *S*1–*S*10 | |
| | Females | Males | Females | Males | Females | Males |
| High lines | 0·84 | 0·73 | 0·4 | 0·3 | 0·5 | 0·4 |
| Low lines | 0·84 | 0·73 | 1·3 | 1·1 | 1·0 | 0·8 |

(ii) *Phenotypic Variance*

Within-culture estimates of the variance of the sum of the two sternites ($\sigma_P^2$) and the variance of the difference between sternites ($\sigma_D^2$) and the ratio $\sigma_D^2/\sigma_P^2$ are plotted in Figure 3. They are given in arithmetic units. Cavalli (1952) and Reeve and F. W. Robertson (1954) concluded that a log transformation is not justified. The rather independent behaviour of means and variances in the results of Clayton and A. Robertson (1957), and in the data presented in this report, also cast doubts on the usefulness of a different scale.

There is no evidence here of the highly exaggerated variance effects obtained by Clayton and A. Robertson, though $H_2$ is different from the other lines and has some of the features described by these authors. $H_1$, $L_1$, and $L_2$ show a slightly lower $\sigma_P^2$ than the base population even before *S*10, and this continues to *S*33 with relatively small fluctuations from generation to generation. Occasional larger deviations occur, such as the increases in $L_2$ females at *S*13, $H_1$ females at *S*18, and

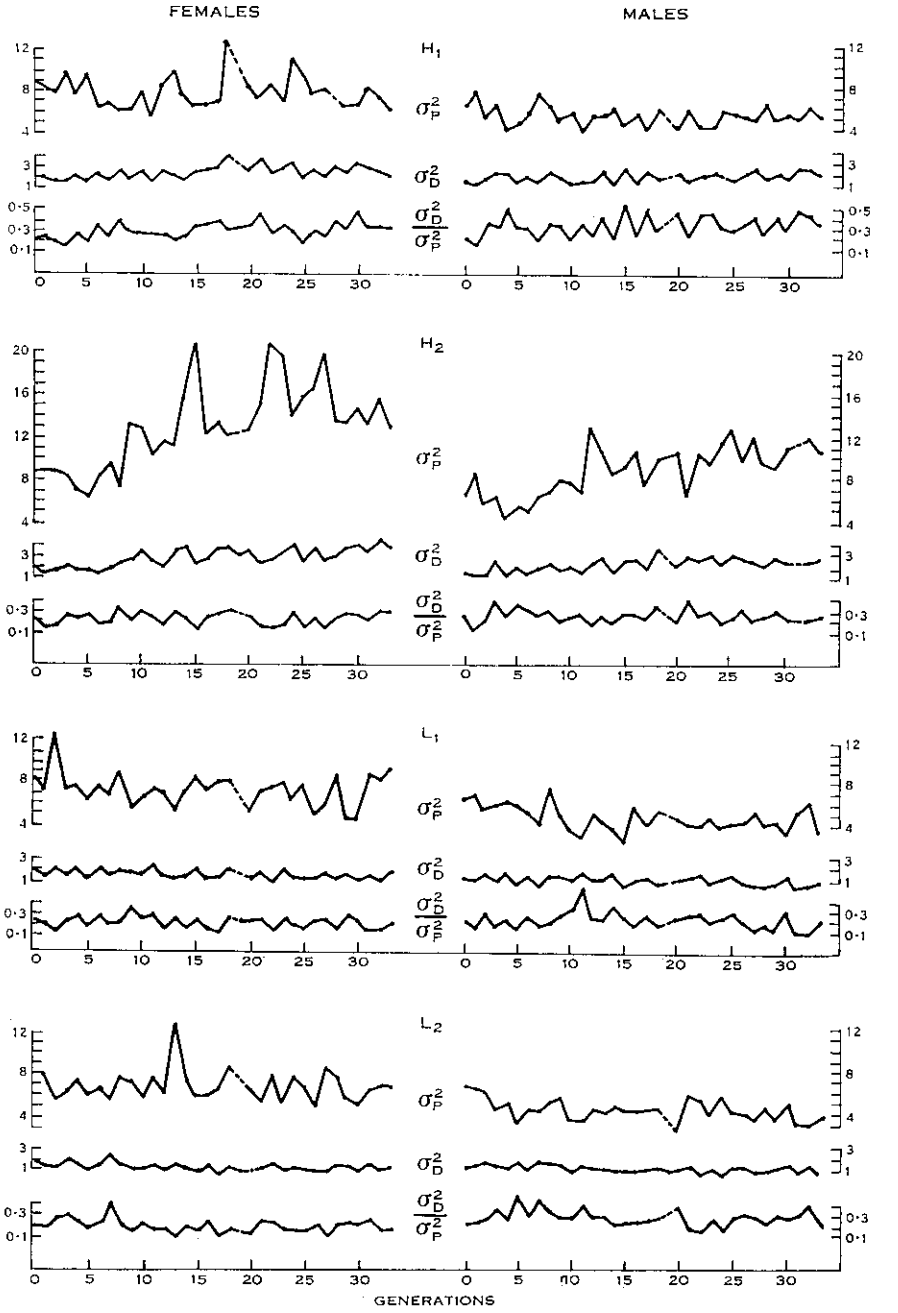


Fig. 3.—Phenotypic variances of four selection lines through 33 generations of selection.

$L_1$ females at $S2$. These instances do not fit into any general pattern. Females show more extreme fluctuations than males and the low lines rather more reduction in variance than $H_1$.

$H_2$ is quite different. There is some fluctuation, perhaps even a decline in $\sigma_P^2$ in the first seven or eight generations. At $S9$ a sudden marked increase occurs in females, which is maintained, maybe enhanced, up till $S33$, highest variances being recorded at $S15$, $S22$, $S23$, and $S27$. Male variance increases more gradually, lagging a couple of generations behind the females; fluctuations in a high direction are not as extreme as in females and they occur in different generations.

TABLE 7

HERITABILITIES ($h_s^2$ and $h_d^2$), AND GENETIC INTERACTION COMPONENTS ($\sigma_I^2$) EXPRESSED AS A FRACTION OF THE PHENOTYPIC VARIANCE ($\sigma_P^2$), WHICH WERE DERIVED FROM THE DIALLEL ANALYSES OF $S13$, $S23$, AND $S33$

$n$ = degrees of freedom for sires, dams, and interaction terms in analyses of variance from which these statistics derived

| Generation | | Line $H_1$ | | Line $H_2$ | | Line $L_1$ | | Line $L_2$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Females | Males | Females | Males | Females | Males | Females | Males |
| $S13$ | $h_s^2$ | 0·23 | 0·24 | — | 0·07 | 0·33 | 0·02 | 0·11 | — |
| | $h_d^2$ | 0·17 | 0·35 | 0·32 | 0·52 | 0·06 | — | 0·09 | 0·13 |
| | $\sigma_I^2/\sigma_P^2$ | 0·12 | 0·02 | 0·04 | — | 0·09 | 0·12 | 0·06 | 0·08 |
| | $n$ | 35 | 35 | 31 | 31 | 30 | 30 | 35 | 35 |
| $S23$ | $h_s^2$ | 0·09 | — | 0·10 | 0·20 | 0·05 | 0·33 | — | 0·03 |
| | $h_d^2$ | 0·32 | — | 0·19 | 0·16 | — | 0·16 | 0·11 | 0·14 |
| | $\sigma_I^2/\sigma_P^2$ | 0·07 | 0·10 | 0·06 | — | 0·06 | — | 0·09 | — |
| | $n$ | 43 | 42 | 29 | 29 | 21 | 21 | 33 | 33 |
| $S33$ | $h_s^2$ | 0·33 | 0·10 | — | — | — | 0·15 | 0·04 | 0·12 |
| | $h_d^2$ | 0·29 | — | 0·17 | — | — | 0·51 | 0·04 | 0·11 |
| | $\sigma_I^2/\sigma_P^2$ | 0·06 | 0·01 | 0·02 | — | 0·13 | 0·10 | 0·03 | 0·02 |
| | $n$ | 32 | 32 | 18 | 19 | 8 | 9 | 43 | 43 |

In general $\sigma_D^2$ tends to increase in the high lines but decrease in the low lines during the period of selection. The trends are far from uniform, however, and each line shows individual characteristics, with some differences between the sexes and considerable fluctuation from generation to generation. Some caution is needed in interpreting these results in view of the results for the base population under different culture conditions (Section III($a$)). Here again the two low lines parallel each other more closely than do the high replicates.

The lines $H_2$, $L_1$, and $L_2$ do not show any consistent trends or large fluctuations in $\sigma_D^2/\sigma_P^2$. In line $H_1$ the ratio increases from a base population level of 0·21 to an

average level of approximately 0·3 for females and approaching 0·4 for males. Clayton, Morris, and A. Robertson (1957) used the ratio $\sigma_D^2/\sigma_P^2$ as an indicator of trends in the amount of genetic variance in a population. The validity of this procedure will be discussed in Section IV.

(iii) *Genetic Variance*

The diallel analyses at *S*13, *S*23, and *S*33 were planned to give information on trends in the genetic variance, particularly the non-additive portion. Table 7 shows the results of the diallel analyses in the form of heritability estimates from both the sire ($h_s^2$) and dam ($h_d^2$) components and the fraction $\sigma_I^2/\sigma_P^2$, where $\sigma_I^2$ is the genetic interaction component. It is $\sigma_I^2$ which indicates the importance of the non-additive portion of the genetic variance.

It should be pointed out again that the analyses in Table 7 are based on the number of bristles on only one sternite, the fifth, for which heritability in the base population was 0·15–0·20 and $\sigma_I^2/\sigma_P^2$ was negligible. Interpretation of the data in Table 7 is difficult in view of the small number of degrees of freedom for most estimates. Inability to handle larger numbers, and low fertility and viability in some lines, were the major factors responsible for this. However, some possible trends can be considered.

At *S*13, $h^2$ is perhaps slightly higher in the high lines but slightly lower in the low lines than the base population level. At *S*23, the low lines are not much different from *S*13, but $h^2$ in the high lines has now dropped below base population level, at least in $H_1$. By *S*33, $H_1$ and $L_1$ (subject, of course, to the small numbers involved) seem to be back more or less to base population level, $H_2$ has declined sharply, and $L_2$ is about the same as at *S*13 and *S*23.

The $h^2$ values cannot be considered apart from the $\sigma_I^2/\sigma_P^2$ terms because, as mentioned earlier (Section II), $h^2$ estimated by the diallel method does include $\frac{1}{4}$ of the additive × additive epistatic variance. Table 7 shows (i) that the non-additive genetic variance has increased in all four selection lines, to a level between 2 and 8 times its value in the base population; (ii) that the maximum level of $\sigma_I^2$ in each line was already attained by *S*13, when the first estimates were made; and (iii) that, in $H_1$, $H_2$, and $L_2$, $\sigma_I^2/\sigma_P^2$ decreases again between *S*13 and *S*33 but in $L_1$ a decrease at *S*23 is followed by a rise at *S*33.

Since $\sigma_I^2$ includes $\frac{1}{4}$ of the dominance variance, between $\frac{1}{8}$ and $\frac{1}{16}$ of the first-order epistatic deviations, and smaller fractions of the higher-order epistatic deviations, and since one does not know what types of non-additive effects are operating, one can estimate only very crudely how much of the phenotypic variance is due to non-additive genetic effects by multiplying $\sigma_I^2$ by some factor between 4 and 16, or even larger if higher-order interactions are suspected. Thus in the most extreme case of $L_1$ at *S*13 and *S*33, at least 40–50% of the phenotypic variance is calculated as being due to non-additive effects. In the least extreme case, $H_2$ females at *S*33, the minimum estimate is 8%, and the maximum could be 30% or even higher, again depending on the type of interaction involved.

Some additional information on these aspects is provided by some full-sib correlations obtained from all the initial matings set up for the diallel crosses. At

*S*23 and *S*33, and especially in lines $H_2$ and $L_1$, some of these matings produced progeny which were scored but not included in the statistical analysis of the diallels, because of failure of one or more of the later matings needed to complete each $2 \times 2$ set. All the initial productive matings in each line were analysed to give full-sib correlations and the resulting heritabilities are given in Table 8. Each full-sib family had five females and five males scored, just as in the diallel analyses. These $h^2$ estimates are rather more reliable statistically than those in Table 7. They are also considerably higher. The difference is doubtless due to the increase in the non-additive portion of the variance, since the variance component for between full-sib families (used in Table 8) includes about $\frac{1}{4}$ of the dominance variance and rather more of the epistatic variance than the sires and dams components of a diallel analysis (as in Table 7).

TABLE 8

HERITABILITY ESTIMATED FROM FULL-SIB CORRELATIONS—DATA OBTAINED IN ASSOCIATION WITH DIALLEL CROSSES AT *S*13, *S*23 AND *S*33

$n$ = degrees of freedom for "between families" term in the analyses of variance

| Generation | | Line $H_1$ | | Line $H_2$ | | Line $L_1$ | | Line $L_2$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Females | Males | Females | Males | Females | Males | Females | Males |
| *S*13 | $h^2$ | 0·44 | 0·26 | 0·33 | 0·28 | 0·59 | 0·30 | 0·33 | 0·13 |
| | $n$ | 69 | 69 | 61 | 61 | 59 | 59 | 69 | 69 |
| *S*23 | $h^2$ | 0·31 | 0·08 | 0·26 | 0·22 | 0·23 | 0·24 | 0·22 | 0·15 |
| | $n$ | 89 | 89 | 63 | 61 | 67 | 67 | 79 | 81 |
| *S*33 | $h^2$ | 0·21 | 0·02 | 0·19 | 0·07 | 0·29 | 0·47 | 0·13 | 0·14 |
| | $n$ | 73 | 73 | 65 | 65 | 49 | 51 | 95 | 95 |

The $h^2$ values in Table 8 show a parallel trend to the $\sigma_I^2/\sigma_P^2$ values of Table 7. Maximum estimates occur at *S*13, followed by a decline in $H_1$, $H_2$, and $L_2$ up to *S*33 and in $L_1$ a decline at *S*23 but a rise to the previous level at *S*33. It is fairly clear, therefore, that the additive genetic variance declined in all lines by *S*13, accompanied by an increase in the non-additive genetic variance, which also declined subsequently, except in $L_1$. However, each line presents a more or less individual picture so the relationships in the different aspects of the response will now be summarized separately for each line.

(iv) *Individual Lines in Detail*

(1) *Line* $H_1$.—This line showed quite an even response rate in mean bristle number through the 30-odd generations under study, as shown previously in Figures 1 and 2. If the indications of a plateau in the last few generations were real, and some later observations have confirmed this, then the rate during the period of active response was 0·32 bristles per generation for females and 0·27 for males, not

much less than the response in the first 5–10 generations. This represents a fairly constant realized heritability through 30 generations of about 10%, somewhat less than half that predicted from the heritability in the base population. The ratio of mean score of females to mean score of males did not change.

It is difficult to relate this even response picture, and eventual termination of response, with the phenotypic and genetic variance statistics given previously in Tables 2, 7, 8, and in Figure 3. In summary, these showed a mild reduction in phenotypic variance ($\sigma_P^2$), an increase in the developmental error peculiar to each sternite ($\sigma_D^2$), a reduction in the additive genetic variance, and an increase in the non-additive genetic variance up to *S*23, but some decrease in the latter between *S*23 and *S*33. However, the non-additive genetic variance was still probably at least 25% of the phenotypic variance at *S*33. The *S*33 $h^2$ estimates also indicated a small residual amount of additive genetic variance, though the small number of degrees of freedom and uncertainty as to the type of non-additive genetic effects makes this doubtful.

The expectation from the variance statistics of this line would have been a gradually diminishing rate of response, starting from a heritability of approximately 20%, and the maintenance of the non-additive, presumably non-utilizable, genetic variance at its high *S*23 level. Instead, the response occurred at a constant rate (10% realized $h^2$) for 30 generations before ceasing, and the non-additive variance declined from its peak level towards the end of the period. It seems that a considerable part of variability analysed as non-additive genetic has in fact been utilized by mass selection, though a significant residue still remains when the line has reached a definite plateau.

(2) *Line* $H_2$.—$H_2$ and $H_1$ began to diverge at about *S*8. The period *S*8–*S*13 was either one of little real genetic response or one in which the genetic response was masked by some unknown environmental trend. The latter may be more likely because $H_1$ also showed apparently no response in this period, and the low lines gave a hint of a slightly enhanced response at the same time. Whatever the reason, this period was followed by a rapid rate of response to *S*18 and again to *S*23, following the decline at *S*19 after one generation of relaxation of selection pressure.

Although the period of accelerated response was preceded by a large increase in the phenotypic variance, the estimated heritability at *S*13 was not very different from the base population. Nevertheless, the realized heritability in females for the period *S*13–*S*18 was about 40%. This time actual response has been twice as great as what one would have predicted from *S*13 parameters, the reverse of what happened in the early generations. Females responded at a faster rate than males between *S*13 and *S*23, so the ratio of female score to male score at *S*33 was 1·31 compared with 1·26 in the base population.

The realized heritability after *S*23 was again less than 10%, even though estimated heritability at *S*23 was only slightly smaller than at *S*13. Finally at *S*33, when response had ceased (confirmed by later observations not reported here), estimated additive genetic variability was negligible and the residue of non-additive variability was less than in $H_1$.

In this line then, as in $H_1$, the additive genetic variance was gradually exhausted, the non-additive variance increased, some of it was utilized by selection, and a small residue remained after response had ceased. At no stage did the realized $h^2$ seem to follow prediction based on these variances, except perhaps at the plateau ($S33$), when estimated $h^2$ was also close to zero.

(3) *Line* $L_1$.—Initially, this line responded at a rate up to 50% higher than that predicted from the base population heritability. By $S10$ it had slowed to a realized heritability of 20% (initially = 30%). The rate of response continued to decline, but much more gradually, up to $S30$, as seen in Table 9. A plateau was then fairly evident, confirmed, as with $H_1$ and $H_2$, by later observations. No difference occurred between male and female rates of response.

TABLE 9

RESPONSE IN MEAN NUMBER OF BRISTLES PER GENERATION AT DIFFERENT STAGES OF SELECTION IN LINE $L_1$

| | $S1$–$S5$ | $S6$–$S10$ | $S11$–$S15$ | $S16$–$S20$ | $S21$–$S25$ | $S26$–$S30$ |
|---|---|---|---|---|---|---|
| Females | 1·3 | 0·80 | 0·72 | 0·64 | 0·58 | 0·57 |
| Males | 1·1 | 0·54 | 0·62 | 0·40 | 0·56 | 0·35 |

Contrasted with this smooth response curve, indicating gradual diminution of utilizable genetic variance, is the variance picture, which indicated that little additive genetic variance remained by $S13$, that at least 40–50% of the phenotypic variance was by then non-additive genetic, and that the latter declined by $S23$ but recovered its previous level by $S33$. Again it seems that much of the variability which has been analysed as non-additive has been utilized by mass selection. At the plateau, however, this line has a much larger residue of this type of variability than any of the other selection lines.

(4) *Line* $L_2$.—Up to $S16$, $L_2$ followed $L_1$ very closely both in response in mean bristle number and in the reduction of $\sigma_P^2$ and $\sigma_D^2$. As in $L_1$, the diallel data at $S13$ indicated that the additive genetic variability was now very small and that the non-additive portion was a large part of the phenotypic variance, at least 30% and possibly larger. From $S16$, however, $L_2$ diverged sharply from $L_1$ and proceeded at half the $L_1$ rate of response through to $S33$, in females at least. $L_2$ males had a slower and more discontinuous response than females, so the ratio of female to male mean bristle score finished at 1·22 compared with 1·26 in the base population.

In spite of the different rate of response after $S16$, $L_2$ and $L_1$ remained similar in the size of $\sigma_P^2$, $\sigma_D^2$, and the diallel statistics at $S23$. Again, despite the apparent exhaustion of additive genetic variability in $L_2$ and a much smaller residue of non-additive variability in $L_2$ than in $L_1$ at $S33$, $L_2$ continued to respond slowly for many more generations while $L_1$ was at a definite plateau. The detailed data on which the latter observations are based will not be reported here.

It is, therefore, no easier to demonstrate agreement between prediction and actual response in this line than in the other lines. One further phenomenon was observed in $L_2$, however, in the second half of the experiment. This was the occurrence, segregation, and unconscious selection of a gene or gene complex, which had extreme infertility and a very large reduction in bristle number in homozygotes, but rather uncertain effects in heterozygotes. Numerous data were collected on this situation but proved quite inconclusive as regards the exact role of this "gene" in the response picture of the line. They will be reported in detail elsewhere.

## IV. Discussion

At the beginning of a discussion of the interrelationships in these results, and comparison with previous studies in artificial selection, one might consider briefly the question of general applicability of *Drosophilia* experiments to an understanding of the inheritance of quantitative characters. Clayton and A. Robertson (1957) considered the possibility that the extreme individuality of their replicate selection lines in the long term might be related largely to peculiarities of *Drosophila* physiology and genetics, such as very high reproductive ability, small chromosome number, and lack of crossing-over in the male. Lerner (1958) has, on similar grounds, questioned the applicability to other organisms of the unpredictable nature of long-term selection response in *Drosophila*. Not unnaturally, this doubt is relatively prevalent in circles concerned with the application of quantitative genetics theory to animal breeding. The possibility certainly cannot be ignored. Nevertheless, *Drosophila* results provide some of the main experimental evidence so far in this field, and probably ought to be taken at their face value until sufficient experimental evidence on other organisms is available.

F. W. Robertson and Reeve (1952) and Reeve and F. W. Robertson (1953, 1954) have also argued specifically against the general usefulness of the abdominal bristle character for selection experiments of this type, largely on the grounds that all the non-genetic variability in the character is due to chance developmental effects peculiar to each sternite. This was certainly so in the populations studied by them (Reeve and F. W. Robertson 1954) and by Clayton, Morris, and A. Robertson (1957). However, the base population in the present experiment had approximately two-thirds of its non-genetic variability due to environmental or chance effects *common* to both sternites, and only one-third due to *independent* effects on each sternite. Whatever the merits of the above criticism, it does not readily apply to the present population. The Oregon-RC stock used here, therefore, can be regarded as providing a situation more analogous to that of a quantitative character of low to medium heritability in applied breeding work. From considerations of rate of response relative to the higher selection intensities involved, it seems likely that the base populations of Mather and Harrison (1949) and Rasmuson (1955) were similar to the present one in this respect.

The question of first importance in this study has been the *predictive value of heritability estimates*. The response in the early generations was decidedly not in close agreement with the heritability in the base population and later responses were

not generally consistent with the genetic variance picture presented by the diallels at *S*13, *S*23, and *S*33. The large errors of estimate in the latter three sets of diallels should be stressed again. However, it is the totality of inconsistency in the long-term responses which is the basis of the conclusion against heritability being a valid predictor of selection response, rather than isolated instances. The view developed in the review of previous studies (Section I), that the weight of evidence was already suggesting this conclusion, is, therefore, supported well by the present results.

This does not mean that estimated heritability bears no relation at all to the selection response. Indeed the early responses showed that the base population heritability did in fact predict the rate at which high and low lines diverged. A similar comparison was not possible with the later heritability estimates. Falconer's (1953) experiment with mouse body weight gave a very similar result, but those of Clayton, Morris, and A. Robertson (1957) departed somewhat from this view. Lerner (1958) mentioned the possibility that estimated heritability might predict accurately only the rate of separation of high and low lines. However, the main tenor of his and other generally accepted ideas of heritability (Lush 1945; Falconer 1960) estimated by a variety of methods, has been that it can be used to predict progress in either direction. If this is not generally true, as now seems likely, then the utility of heritability theory is much restricted. Perhaps the only way to obtain a heritability estimate which would allow accurate predictions for a few generations would be by the method of parent–offspring regression, using only parents with a phenotype extreme in the direction towards which selection was going to be practised—in other words a one-generation selection experiment.

The particular aspect on which the present experiment was designed to provide more information than previous studies was the non-additive portion of the genetic variance. The results certainly followed the expectation that additive genetic variance would decline and the non-additive genetic variance would increase as selection proceeded. Of greater interest, however, was the apparent utilization of much of the non-additive variation by a mass selection programme. Again, one cannot deny that large sampling errors in the diallel analyses may have made this effect seem more widespread than it actually was. It is highly improbable, though, that it can all be explained away in this way. A further possibility is that genotype–environment interactions became important in the selection lines and were confounded statistically with the genetic interaction terms of the diallels. No further data are available on this. If genotype–environment interactions were not important, it follows that non-additive variation was selected for fairly extensively. It is difficult to fit this finding into the existing framework of theory (cf. Lush 1945, 1948; Lerner 1958; Falconer 1960), which places such variation largely outside the reach of mass selection.

Of course, the diallel method used here gives only a very general estimate of non-additive variability. If more were known on how much of it was in the form of dominance, overdominance, or the different types of epistasis, the results might begin to be more intelligible. Experimental work along these lines, taking more care to eliminate sources of error such as genotype–environment interactions in the estimates of parameters is certainly required to check empirically the trend suggested

here. Even so, perhaps the present data are enough to warrant some preliminary theoretical reappraisal of the problem of non-additive variability in a population subject to mass selection.

Several large selection experiments with abdominal bristles in *Drosophila* have now been reported, though not always conducted from the standpoint of heritability theory. Many useful comparisons with the present study are possible and are discussed below.

### (*a*) *Basic Differences in Variance*

The basic differences in variance between this base population and that of Clayton *et al.* have already been referred to. They agreed only in the absence of non-additive genetic variability. There is no immediately obvious reason for the large common environmental component in this population and its virtual absence in the other. The much lower heritability here might reflect a higher degree of previous inbreeding, hence loss of genetic variance, in the Oregon-RC stock, in which case it could be expected to be more susceptible to influences of the environment. However, this would also lead one to expect a higher degree of developmental instability, i.e. greater $\sigma^2_D$. In fact the reverse occurs, Clayton's population having a higher $\sigma^2_D$ component. Though more information on these and other populations is needed, it may well be that they were dealing with a situation which was essentially not a polygenic one, in the sense of the genetic variance being due to many genes each of small effect. If so, Reeve and F. W. Robertson's criticism (1953, 1954) that the character is not a typical quantitative character would be important for Clayton's population. For the present, the Oregon-RC population used here can be considered as relatively free of this criticism.

### (*b*) *Differences in Response between Replicate Lines*

Differences in response between replicate lines are an important part of any assessment of predictive value of heritability. This comparison involves only short-term selection responses. As discussed previously Clayton, Morris, and A. Robertson's (1957) high lines showed excellent average agreement with prediction, but there was a significant difference between replicates, which leads to the opposite conclusion from one based on the average response. Conversely, their low lines showed smaller differences between replicates, but the average was a much smaller response than predicted by heritability. The latter was the situation in the present selection lines in the short term, agreement between replications but deviation from the predicted response. The agreement between replicates was in fact better in this experiment. This too may be a function of the kind of genetic control in the particular population in terms of gene frequencies, number of genes, size of effects of individual genes, and extent of environmental effects. This point, however, cannot be argued too strongly, as repeatability in this experiment is based on only two replicates in each direction, while Clayton's experiments had five replicates. Differences between replicates in the long term (30 generations) was of similar magnitude in both experiments.

### (c) *Asymmetry of Response*

Asymmetry of response between high and low directions has occurred in all reported selection experiments on abdominal bristles. Mather and Harrison (1949) and Clayton, Morris, and A. Robertson (1957) obtained greater response in the high direction, while Rasmuson (1955) and the present study show the opposite. Lerner (1958) and Falconer (1960) have discussed possible causes of asymmetry arising in lines from a single population. Discussion ranges around differences in actual selection differentials in high and low lines, directional dominance or directional gene frequencies or both in the base population, selection for heterozygotes in one direction more than the other, and inbreeding depression. These factors can also be invoked in attempting to explain differences in amount and direction of asymmetry arising from different base populations. Inbreeding depression is not likely to be an important factor where abdominal bristles are concerned. As Falconer (1960) has pointed out, the other factors probably cannot account sufficiently for immediate, short-term asymmetry. As this is what occurred in the experiments mentioned, it seems that some other unknown features of the genetic control of abdominal bristles in the various base populations must be responsible for the asymmetry observed within each and for the differences between them. A combined theoretical and experimental attack is needed on this problem, for it is symptomatic of where heritability theory fails as a complete description of a population's selection potential. Experimentally, it should be possible to approach in *Drosophila* by intensive chromosomal and linkage analysis of replicate lines in both high and low directions after a few generations of selection. Here, as in most aspects of quantitative genetics, the experimental basis for discussion is deficient.

### (d) *Total Response and Duration of Response*

Total response and duration of response in several different populations were also discussed briefly by Falconer (1960). He detected a certain consistency in the results in that the total range, or the difference between the upper and lower selection limits, in each case was between 15 and 30 times the square root of the original additive variance, and between 10 and 20 times the original phenotypic standard deviation. The total range in this experiment was of this order, the values being 24 and 12 respectively. However, these values are averages of replicate lines and for prediction purposes would give no more than the merest approximation to the selection limit that would be attained by any single line. Their value and practical utility is, therefore, doubtful.

The present selection lines also agreed with the general picture that response does not last longer than 20 or 30 generations. In this respect they were more uniform in behaviour than Clayton and A. Robertson's lines, which also tended to plateau earlier. Three out of the present four lines stopped responding within a generation or two of each other at about $S30$, while the other ($L_2$) kept going at a very slow rate for about 15 generations longer. The reality of these plateaux will be confirmed by subsequent data given in a later report, together with details of the further response of $L_2$.

A. Robertson (1960) has recently formulated a theory of limits in artificial selection which describes the expected limit as a function of $N\bar{i}$, where $N$ = the effective population size and $\bar{i}$ = the selection differential in standard units. The total advance expected is something like $2N$ times the advance in the first generation for additive genes and for low values of $N\bar{i}$. The value of $N$ aimed at in this experiment was 40 but the actual value, especially in later generations may have been nearer 30. The limit in $H_2$ approximated $2N$ times the average advance in the first generations where $2N$ is taken as 60, but one wonders about the meaning of this apparent agreement. The other lines were nowhere near such prediction. Also $N\bar{i}$ here is possibly not low, the condition required for the above relationship to hold, though the paper leaves some doubt as to the level of $N$ which is to be considered low or high. Furthermore, this prediction presumes a continuing additive model of gene action with correspondingly steady response to selection, which the capricious behaviour of line $H_2$ does not readily fit. The theory also allows for estimation of the half-life of effective selection advance as being up to $1 \cdot 4N$, which in these lines would be about 40 generations, if effective $N = 30$ instead of the 40 aimed at, and about 30 generations if effective $N$ approached as low as $0 \cdot 55$ of the actual size, as indicated by Crow (1954) and Morris (1954). $N$ was probably not as low as this latter estimate in the present populations, because of the split-up of potential parents into five bottles per line. In any case the calculated half-life is either equal to or greater than the total number of generations of actual response. As the half-life value is an upper limit, this may or may not represent a reasonable agreement. The agreement is probably closer for Clayton and A. Robertson's own selection lines. However, Robertson did not discuss those lines in relation to his theory, and again one wonders about its applicability to the complex situations that arose at one time or another in most of the selection lines that have been under discussion.

### (e) *Changes in Phenotypic Variance and* $\sigma_D^2$

Changes in the phenotypic variance and $\sigma_D^2$, the variance of the difference between sternites, and in the ratio $\sigma_D^2/\sigma_P^2$ were much less extreme in the present lines than in the individual lines of Clayton and A. Robertson (1957). Some of this is undoubtedly due to the various major-gene phenomena which developed in their lines, but otherwise the reasons are not obvious. Some discussion is necessary on their interpretation of the ratio $\sigma_D^2/\sigma_P^2$. Primarily they use this ratio as an indicator of trends in the amount of genetic variance, i.e. the ratio becomes larger if the genetic variance declines. Sometimes, however, they refer to it as an index of developmental stability in the same way as Mather (1953) and Thoday (1958) use bilateral asymmetry of the sternopleural bristle character. It can be readily argued that these two interpretations of $\sigma_D^2/\sigma_P^2$ are incompatible, and that the former is the one that is faulty. It implies (i) that $\sigma_D^2$ is not likely to change much with changes in genetic background or in the environment; and (ii) that the common environmental portion of the phenotypic variance will also be relatively constant. If (i) is correct, then $\sigma_D^2/\sigma_P^2$ cannot reflect changes in developmental stability, which must reflect changes in genetic background. However, (i) is not true, because differences in $\sigma_D^2$ do exist between different base populations, because $\sigma_D^2/\sigma_P^2$ does increase in some lines while

the *total* genetic variance does not decrease (e.g. $H_1$), and because the data in Table 2 and for generation $S1$ in Figure 3 showed that $\sigma^2_D$ is susceptible to environmental modification. Assumption (ii) also does not hold since the common environmental variance of the character can change under different environmental conditions (e.g. Reeve and F. W. Robertson 1954), and it is well-known that changes in genotype of a population can alter its susceptibility to the same environmental conditions. The ratio $\sigma^2_D/\sigma^2_P$ is, therefore, not a valid indicator of trends in the amount of genetic variance. It may be useful as an index of changes in developmental stability.

Interpretations of variance statistics depend a lot on choice of scale. In view of discussions on this question for this character by Cavalli (1952), Reeve and F. W. Robertson (1954), Clayton and A. Robertson (1957), and considering the rather independent behaviour of the variances and the means in this experiment, there seems to be no compelling reason for transforming the data from the arithmetic scale. With $\sigma^2_D/\sigma^2_P$ or merely $\sigma^2_D$ used as an index, the present selection lines do not show nearly as much developmental instability as Clayton's. This might simply reflect the same tendency as the initial difference between the two base populations.

It is not clear whether the absence of variability due to common environmental effects on both sternites in the Clayton base population persisted in the selection lines derived from it. In the present selection lines the proportion of the variance attributable to this source (50% in the base population) tended to decrease in the selection lines as the non-additive genetic variance became large. However, as the Clayton base population was so different, it would be interesting to know whether variance due to this source increased in his selection lines, as they became more developmentally unstable.

The levels of developmental instability, measured by $\sigma^2_D$, and of common environmental variance seem to be largely independent of a line's fitness. One might have expected higher levels of internal instability or environmental variability or both to be accompanied by reduced fitness. However, line $H_1$, the only one to show an increase in $\sigma^2_D/\sigma^2_P$, was the fittest of the four lines in this study, not obviously poorer than the base population. In the studies of Clayton and Robertson, one of the most infertile low lines had an extreme increase in $\sigma^2_D/\sigma^2_P$, but the weakest high line apparently had little change.

### (*f*) *Differential Response of the Two Sexes to Selection*

Differential response of the two sexes to selection is a common feature of selection experiments on abdominal bristles. Unfortunately the reports of Rasmuson (1955) and Mather and Harrison (1949) do not allow such a comparison as they gave only pooled means of the two sexes. The most extreme effects were found in Clayton and A. Robertson's low lines (1957) and in two of Harrison's lines (1953), in all of which the usual relationship was reversed, females finally having fewer bristles than males. In Clayton and Robertson's lines, this was associated with major-gene phenomena. In the present experiment the effects were much milder, but in both instances, $H_2$ and $L_2$, females responded at a slightly greater rate than males. Only one population, $H_5$ of Clayton and Robertson, has had the opposite trend, males having slightly more response than females. Most of the data supports the conclusion

that females are more amenable to change in this character than males, which might therefore be considered more buffered developmentally. This suggests uses to which such material might be put in canalization studies. It would be interesting to know for a start whether the sex difference found for bristles extends to development of the abdomen generally, or even further. As differences between the sexes occur generally in mean level of many quantitative characters in most organisms, a watch should be kept for similar trends in selection experiments with other characters. In large animals the phenomenon of alteration, perhaps reversal, of a secondary sex character by selection could have considerable practical and economic importance.

The foregoing comparisons between the selection experiment of Clayton and colleagues and the one reported here have been dealt with in some detail. Differences in response occurred on nearly all aspects observed. However, a certain uniformity was obvious in that changes under selection in the present lines were less extreme than in Clayton's lines for most traits. The more conservative responses here may have been partly a function of an initial lower absolute level of genetic variability in the base population. Alternatively, they could be viewed as reflecting a greater stability or degree of canalization in this population and its derivatives. Perhaps it is not realistic to look for a general connecting relationship for these phenomena. Whatever the reasons, these comparisons re-emphasize the things we did not know about the respective base populations. They must have differed in many ways other than variance and heritability. The results do not simplify the task of trying to obtain more complete and predictive descriptions of populations than the heritability concept has previously provided, yet this must be one of the main aims in this field for the immediate future.

Finally, the significance of these results to applied breeding work needs to be considered. It is assumed for the time being that results with this organism are relevant to other organisms, including domesticated animals. Heritability, estimated by most of the standard methods, would now seem to be deficient as an accurate predictor of subsequent selection response. As noted earlier, heritability estimated as the regression of offspring on selected parents may still be a measure that will predict response for some generations. This really means starting a selection programme and estimating the immediate prospects from the first generation of response. A. Robertson's (1960) use of the first generation response to predict total response and duration of response is a rather different proposition, and for the moment can be regarded as of relatively unknown utility for animal breeding purposes.

A large proportion of animal breeding effort and literature has been devoted to the estimation of heritability. To the extent that many of these estimates have been based on regression of offspring on selected parents, they may still be useful. Otherwise they are of doubtful value by reason of the results and discussion presented above. Accumulation of results from more selection experiments may qualify this conclusion.

A further point of some importance to applied work is the finding here that response to mass selection can apparently continue when the parameters are showing negligible additive genetic variability but a medium to large proportion of non-

additive genetic variability. The theoretical implications of this are by no means clear to the present author, but the available data suggest caution in turning to an alternative method of selection (e.g. reciprocal recurrent selection) as soon as the genetic parameters or other aspects of the population indicate the genetic variability is largely non-additive.

In summary, the import of this study for animal breeding is to indicate that the approach to applied breeding must still remain for the time being, largely empirical. This may sound like turning the clock back but this is probably preferable to placing too much reliance on rather over-simplified, largely untested, theory. It cannot be stressed too much that a prime need for the development of this field of study is large-scale and adequately designed selection programmes, especially with laboratory mammals and larger animals. Failing this, the theory of quantitative genetics will continue to outstrip our knowledge of actual population behaviour, which the theory aims to predict.

## V. Acknowledgments

I wish to thank Dr. J. M. Rendel, my other colleagues in the Division of Animal Genetics, C.S.I.R.O., University of Sydney, and Dr. A. S. Fraser, now at the Department of Genetics, University of California, Davis, California, for helpful advice and criticism on most aspects of this work. My thanks are also due to Dr. P. J. Claringbold, also of the Division of Animal Genetics, C.S.I.R.O., for statistical advice, to Dr. J. S. F. Barker, Department of Animal Husbandry, University of Sydney, for his continuing interest and criticism, and to Dr. Sewall Wright and Dr. James F. Crow, University of Wisconsin, U.S.A., for much useful discussion while the paper was in preparation. The material in this paper formed part of a thesis presented to the University of Sydney as requirement towards a Ph.D. degree.

## VI. References

Cavalli, L. L. (1952).—An analysis of linkage in quantitative inheritance. In "Quantitative Inheritance". pp. 135–44. (H.M.S.O.: London.)

Clayton, G. A., Knight, G. R., Morris, J. A., and Robertson, A. (1957).—An experimental check on quantitative genetical theory. III. Correlated responses. *J. Genet.* **55**: 171–80.

Clayton, G. A., Morris, J. A., and Robertson, A. (1957).—An experimental check on quantitative genetical theory. I. Short-term responses to selection. *J. Genet.* **55**: 131–51.

Clayton, G. A., and Robertson, A. (1957).—An experimental check on quantitative genetical theory. II. The long-term effects of selection. *J. Genet.* **55**: 152–70.

Crow, J. F. (1954).—Breeding structure of populations. In "Statistics and Mathematics in Biology." (Eds. O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush.) (Iowa State College Press: Ames, Iowa.)

Dempster, E. R., Lerner, I. M., and Lowry, Dorothy C. (1952).—Continuous selection for egg production in poultry. *Genetics* **37**: 693–708.

Dickerson, G. E. (1955).—Genetics slippage in response to selection for multiple objectives. *Cold Spr. Harb. Symp. Quant. Biol.* **20**: 213–24.

Doney, J. M. (1958).—The role of selection in the improvement of Welsh mountain sheep. *Aust. J. Agric. Res.* **9**: 819–29.

Dun, R. B. (1958).—The influence of selection and plane of nutrition on the components of fleece weight in Merino sheep. *Aust. J. Agric. Res.* **9**: 802–18.

DUN, R. B. (1959).—The introduction of mutant genes into animal populations as an aid to the breakdown of canalised development. Ph. D. Thesis, University of Sydney.

FALCONER, D. S. (1953).—Selection for large and small size in mice. *J. Genet.* **51**: 470–501.

FALCONER, D. S. (1954).—Validity of the theory of genetic correlation. *J. Hered.* **45**: 42–4.

FALCONER, D. S. (1955).—Patterns of response in selection experiments with mice. *Cold Spr. Harb. Symp. Quant. Biol.* **20**: 178–96.

FALCONER, D. S. (1960).—"Introduction to Quantitative Genetics." (Ronald Press Co.:New York.)

HARRISON, B. J. (1953).—Reversal of a secondary sex character by selection. *Heredity* **7**: 153–64.

JEROME, F. N., HENDERSON, C. R., and KING, S. C. (1956).—Heritabilities, gene interactions and correlations associated with certain traits in the domestic fowl. *Poult. Sci.* **35**: 995–1013.

KYLE, W. H., and CHAPMAN, A. B. (1953).—Experimental test of the effectiveness of selection for a quantitative character. *Genetics* **38**: 421–33.

LERNER, I. M. (1950).—"Population Genetics and Animal Improvement." (Cambridge Univ. Press.)

LERNER, I. M. (1958).—"The Genetic Basis of Selection." (J. Wiley & Sons, Inc.: New York.)

LERNER, I. M., and DEMPSTER, E. R. (1951).—Attenuation of genetic progress under continued selection in poultry. *Heredity* **5**: 75–94.

LERNER, I. M., and HAZEL, L. N. (1947).—Population genetics of a poultry flock under artificial selection. *Genetics* **32**: 325–39.

LUSH, J. L. (1945).—"Animal Breeding Plans." (Iowa State College Press: Ames, Iowa.)

LUSH, J. L. (1948).—"The Genetics of Populations." (Mimeo.) (Iowa State College.)

MATHER, K. (1953).—Genetical control of stability in development. *Heredity* **7**: 297–336.

MATHER, K., and HARRISON, B. J. (1949).—The manifold effect of selection. I and II. *Heredity* **3**: 1–52; 131–62.

MORLEY, F. H. W. (1951).—Selection for economic characters in Australian Merino sheep. I. Estimates of phenotypic and genetic parameters. Sci. Bull. Dep. Agric. N.S.W. No. 73. 45 pp.

MORLEY, F. H. W. (1955*a*).—Selection for economic characters in Australian Merino sheep. V. Further estimates of phenotypic and genetic parameters. *Aust. J. Agric. Res.* **6**: 77–90.

MORLEY, F. H. W. (1955*b*).—Selection for economic characters in Australian Merino sheep. VI. Inheritance and inter-relationships of some subjectively graded characteristics. *Aust. J. Agric. Res.* **6**: 873–81.

MORRIS, J. A. (1954).—Ph. D. Thesis, University of Edinburgh.

RASMUSON, M. (1952).—Variation in bristle number of *Drosophila melanogaster*. *Acta Zool.* **33**: 277–307.

RASMUSON, M. (1955).—Selection for bristle numbers in some unrelated strains of *Drosophila melanogaster*. *Acta Zool.* **36**: 1–49.

REEVE, E. C. R., and ROBERTSON, F. W. (1953).—Studies in quantitative inheritance. II. Analysis of a strain of *Drosophila melanogaster* selected for long wings. *J. Genet.* **51**: 276–316.

REEVE, E. C. R., and ROBERTSON, F. W. (1954).—Studies in quantitative inheritance. VI. Sternite chaeta number in *Drosophila*, a metameric quantitative character. *Z. Vererb. Lehre* **86**: 269–88.

RENDEL, J. M., and ROBERTSON, A. (1950).—Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J. Genet.* **50**: 1–8.

RENDEL, J. M., ROBERTSON, A., and ALIM, K. A. (1951).—The extent of selection for milk yield in dairy cattle. *Emp. J. Exp. Agric.* **19**: 295–301.

ROBERTSON, A. (1955).—Selection in animals: Synthesis. *Cold. Spr. Harb. Symp. Quant. Biol.* **20**: 225–9.

ROBERTSON, A. (1960).—A theory of limits in artificial selection. *Proc. Roy. Soc.* B **153**: 234–49.

ROBERTSON, A., and RENDEL, J. M. (1950).—The use of progeny testing with artificial insemination in dairy cattle. *J. Genet.* **50**: 21–31.

ROBERTSON, F. W. (1956).—The use of *Drosophila* in the experimental study of animal breeding problems. *Anim. Breed. Abstr.* **24**: 218–24.

ROBERTSON, F. W., and REEVE, E. C. R. (1952).—Studies in quantitative inheritance. I. The effects of selection of wing and thorax length in *Drosophila melanogaster*. *J. Genet.* **50**: 414–48.

ROBERTSON, F. W., and REEVE, E. C. R. (1953).—Studies in quantitative inheritance. IV. The effects of substituting chromosomes from selected strains in different genetic backgrounds in *Drosophila melanogaster*. *J. Genet.* **51**: 586–610.

SHELDON, B. L. (1963).—Studies in artificial selection of quantitative characters. II. Selection for body weight in *Drosophila melanogaster*. *Aust. J. Biol. Sci.* **16**: 516–41.

THODAY, J. M. (1958).—Homeostasis in a selection experiment. *Heredity* **12**: 401–15.

TURNER, H. N. (1956).—Measurement as an aid to selection in breeding sheep for wool production. *Anim. Breed. Abstr.* **24**: 87–118.

TURNER, H. N. (1958).—Relationships among clean wool weight and its components. I. Changes in clean wool weight related to changes in the components. *Aust. J. Agric. Res.* **9**: 519–52.

TURNER, H. N. (1959).—Ratios as criteria for selection in animal or plant breeding, with particular reference to efficiency of food conversion in sheep. *Aust. J. Agric. Res.* **10**: 563–80.