Replicated selection for body weight in mice

By D. S. FALCONER*

Institute of Animal Genetics, Edinburgh, EH9 3JN

SUMMARY

The variation in the response to selection was studied by replication of selected lines. A random-bred strain of mice was divided into six replicates. Two-way selection for 6-week weight was applied in an identical manner to each replicate, and each had an unselected control. Each line (6 large, 6 control, 6 small) was maintained by minimal inbreeding with 8 single-pair matings. The overall mean responses, both up and down, were linear and very regular for ten generations, with realized heritabilities of 40 % upwards, 33 % downwards and 37 % for the divergence. The separate replicates, however, differed greatly in their realized heritabilities, with upward selection ranging from 25 to 46 %, and downward selection from 16 to 50 %. The theoretical prediction that, because of genetic drift, the standard error of a realized heritability is underestimated by the standard error of the regression of response on cumulated selection differential was borne out in this experiment. The empirical standard error, calculated from the observed variance between replicates, was more than twice as great as that of the regression. The empirical standard errors showed that the asymmetry between upward and downward responses was not significant. The variation between the replicates was ascribed mainly to random drift, which may seriously influence the conclusions about the realized heritability and the asymmetry of response that would be drawn from a single experiment with the population size of one of these replicates. After 23 generations of selection the large lines were approaching limits, and the limit appeared to be at the same level in all. The small lines showed an undiminished realized heritability after 23 generations, but the selection differentials were then so small that little progress was made. There was evidence of counter-acting natural selection. All aspects of productivity – proportion of fertile matings, litter size and weaning rate – declined in the control lines. The overall productivity of the large lines was a little below the controls, and that of the small lines was reduced to about half the level of the controls. The separate replicates differed from each other significantly in all the components of productivity.

1. INTRODUCTION

Selection experiments without replication of the selected lines are difficult to interpret in several respects because the random effects due to the small population size cannot be readily assessed. How repeatable are the rates of response, the

* Agricultural Research Council, Unit of Animal Genetics.

asymmetry of response, and correlated changes? The experiment described here was done with the object of answering these questions. The character selected was the weight of mice at 6 weeks of age, and there were six replicates of upward and of downward selection and of unselected controls. When the experiment was started the main objects were, more specifically, as follows. (1) To find out how much the realized heritabilities would differ, (2) To identify the causes of variation of generation means about the regression line. The recent work of Hill (1971, 1972 *a*, *b*) on the error structure of selection responses has provided a theoretical basis for analysing the 'repeatability' and answering these first two questions. (3) To compare the observed response with the prediction derived from the parameters of the base population, as a check on the validity of selection theory. (4) To find out if there was an asymmetry of response from the beginning. There is no known reason for immediate asymmetry, yet several previous experiments seemed to show it. (5) To provide material for studying the physiological and cellular correlates of genetic differences in growth-rate. With just a single pair of divergent lines one can never be sure that any difference in another character is causally related to the character selected. This is part of the problem of correlated responses which have been shown on theoretical grounds to be extremely variable and unpredictable (Bohren, Hill & Robertson, 1966). But if some correlated character was found to differ in the same way in all replicates, this would be firm evidence of a physiological connexion with the character selected.

The replication was not itself directly relevant to all of these questions, because the 'errors' due to random genetic effects are reduced by the larger population size rather than by the replication. For the estimation of responses to selection it is theoretically no better to have a population of fixed total size divided into replicates (Hill, 1971); the replication does, however, provide the means of estimating the amount of random genetic drift.

This paper describes the origin of the Q-strain, used as the base population, the parameters of the base population, the selection procedure, and the 'repeatability' of the selection response. The analysis of the differences between replicates and of the causes of variation between generations, in the light of Hill's work, will be given in a later paper.

2. MATERIALS AND METHODS

(i) *Origin of the Q-strain*

In 1960 a strain of mice of mixed origin was set up for maintenance as a non-inbred strain, with the object of providing a base population for the replicated selection, and also to provide a 'random-bred' strain for other work. The strain, which became known as the Q-strain, has been widely used, but a full account of its origin has not yet been published. The features desired of the base population for the selection experiment were that it should have a diverse origin, so that genetic variation was maximal and gene frequencies as unrestricted as possible; and that it should have passed through as many generations as possible since the

last crossing, so that linkage disequilibrium was minimal. A strain having the first requirement had been made by Dr B. M. Cattanach and, though not ideal in every respect, it was better for the purpose than any strain that could have been constructed in a reasonably short time. Representatives of Cattanach's strain were therefore taken to found the Q-strain, and after expansion in numbers, the strain was kept by minimal inbreeding for 15 generations before the selection was started in 1964.

All lines of ancestry of the Q-strain have been traced back over some 30 generations to about 1948. The ancestry is summarized in Fig. 1. This shows the designations of the strains contributing to the ancestry, with references to published work on them, the number of generations, and the numbers of breeding pairs used in each strain. The following are the main features of the ancestry. One quarter of the gene-pool of Cattanach's strain was derived from a highly inbred strain (JU/Fa). This is what made the strain less than ideal because it restricted the range of possible gene frequencies. There were, however, 23 generations during which gene frequencies could have changed by random drift or natural selection before the selection experiment started. Ultimately the ancestry traces back to the following distinct sources: MacArthur's small and large selected strains, Goodale's large selected strain, Bateman's strain selected for high lactation, and two mutant stocks which had recently been crossed to the inbred $C57BL$. Between these origins and the construction of Cattanach's strain, two lines of ancestry passed through selection experiments, one for high growth rate and one for high litter size.

The segregation of several colour genes was introduced by the original sources and some of these persisted. Those still segregating at the foundation of the Q-strain were *non-agouti* (*a*), *brown* (*b*), *albino* (*c*) and *belted* (*bt*). A new recessive mutant (*dt*) causing a severe neuro-muscular defect appeared in the foundation generation. It persisted at low frequency, despite selection against it, until it was finally eliminated in the setting up of the base population for the selection.

(ii) *Maintenance of the 'random-bred' Q-strain*

To set up the Q-strain, spare mice were obtained from all ten matings of generation 9 of Cattanach's strain, and 27 single-pair matings were made. Progeny from 24 of these were used to make the matings of the foundation generation, designated Q_0. The mating system in this, and all subsequent generations until the selection started, was as follows. Twenty males were mated each to four females, A–D. Females A and B were full sibs to each other, and so were C and D, but A and B were unrelated to C and D, and all the females were unrelated to the male. The litter of one female of each sib-pair was discarded after weighing at 6 weeks. Thus the litters providing parents of the next generation were the progeny of 20 males and 40 females. Parents for the next generation were taken at random and, as far as possible, equally from all litters. Thus each surviving litter provided one sib-pair of females, and the progeny of each male provided one male parent. Matings were made by a rotational scheme which minimized the relationships
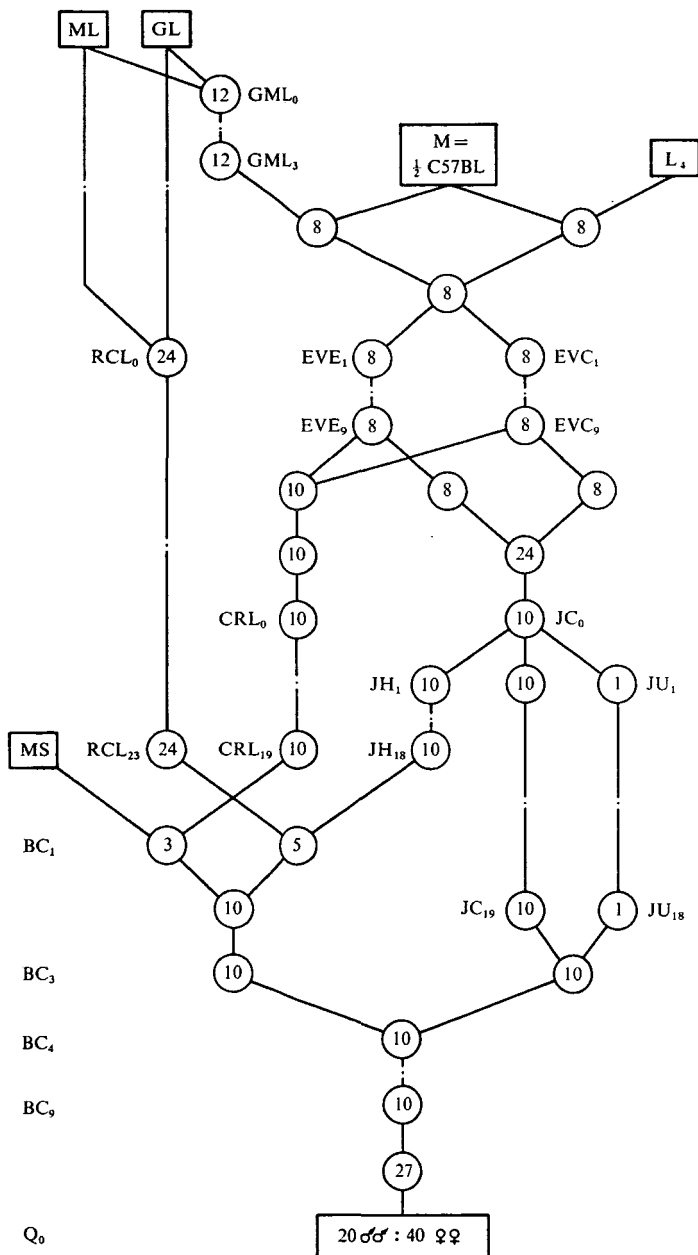
Fig. 1. For legend see facing page.

between the mated individuals. The theoretical rate of inbreeding was 0·547 % per generation. This mating system was followed up to generation Q_{14}, from which the base population for the selection was set up.

The litters were weaned at 3 weeks. Litters born on different days were not housed in the same storage cage at weaning. Three females and two males in

each litter were weighed individually at 3 weeks and 6 weeks of age. Records were also kept of litter size at birth and birth weight of the whole litter. The purpose of the mating system used was to provide full-sib and half-sib groups for the analysis of all these measurements and the maternal effects on them. The analysis of the 6-week weight only is presented here.

There were some changes of management during the maintenance of the random-bred strain and at the start of the selection, made with the aim of improving the health of the mice. The strain suffered throughout from juvenile 'dysentery'. Affected animals were reduced in weight, often severely. The disease tended to be concentrated in litters, though sometimes only a few individuals in a litter were affected. Attempts were made to improve the health of the strain as follows. A modification of the standard pelleted diet was introduced at generation 10; another dietary change – weekly supplements of crushed oats and maize – was made at the start of the selection experiment, and at the same time ectoparasites (mites) were eradicated. Whether as a result of these changes, or for other reasons, the disease disappeared when the selection experiment was started. All diseased mice and others with abnormally low weights were excluded from the analyses.

(iii) *Base population for the replicated selection*

The base population for the selection was made from the progeny of generation 14 of the random-bred Q-strain. Forty-eight single-pair matings were needed to set up the replicates, but 117 matings were made in order to allow replacement of any whose litters were too small, and also to allow rigorous selection against the neuro-muscular mutant that was still present at low frequency. Three matings produced the mutant, and rejection of these and all related matings finally eliminated it. From the remaining matings with acceptable litters six sets of eight matings were chosen to provide the bases of the six replicates, designated A–F. The matings were chosen so as to provide a wide and evenly spread sample of the

Fig. 1. Origin of the Q-strain. Rectangles represent foundation strains. Circles represent generations of derived strains whose designations are shown alongside. The number inside the circle is the number of pair-matings by which the strain was maintained, or the number of pairs used in a cross; all strains were maintained by minimal inbreeding. The subscript number following the strain designation is the generation number of the strain: the number of generations in the pedigree can be deduced from these generation numbers. Breaks in the lines connecting circles indicate the omission of generations from the diagram. The strains, the characters for which they were selected, and references, are given below.

Foundation strains. MS, MacArthur's small body weight (see Roberts, 1966). ML, MacArthur's large body weight; GL, Goodale's large body weight (see Falconer & King, 1953). M, Mutant stocks previously crossed to the standard inbred C57BL/Fa. L, Bateman's high lactation (see Falconer, 1955).

Derived strains. GML, Large body weight (Falconer & King, 1953). RCL, Repeat of GML (King, unpubl.; see Roberts, 1966). EVE, High variability of body weight; EVC, low variability of body weight (Falconer & Robertson, 1956). CRL, High growth-rate on restricted diet (Falconer, 1960a). JH, High litter size; JC, unselected (Falconer, 1960b, 1965). JU, Sib-mated, standard inbred JU/Fa (Bowman & Falconer, 1960). BC, Cattanach's strain.

gene-pool of Q_{14}, and were assigned to the replicates so as to make pairs of replicates as alike as possible in ancestry but as different as possible from the other pairs. Fig. 2 shows schematically how the replicates were derived from the Q_{14} matings. The purpose of setting up the replicates in this way was to make it possible to find out if any differences between the replicates that might appear later were due to sampling within families or to sampling between families of the
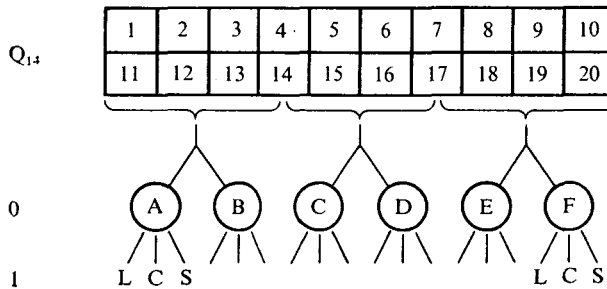

Fig. 2. Construction of base population for replicated selection. Each numbered square represents a sire group of Q_{14}. The circles represent the replicates in the base population, generation 0 of the selection. L, C and S are the Large, Control and Small lines in generation 1, taken from each replicate.

Relationships between replicates, expressed as the number of grandparents in common

| Replicate | Grandparents | | Grandparents in common | | | |
| | | | Within pairs | | Between pairs | |
	♂	♀	♂	♀	♂	♀
A	8	15				
B	8	14	8	14		
					2	1
C	8	15				
D	8	15	8	12		
					2	0
E	8	15				
F	8	15	8	15		

base population. Thus replicates A and B were made from the same litters of Q_{14}, and so were C and D, and also E and F. But the three pairs, A–B, C–D and E–F, were taken as far as possible from different families. There were 20 sire-groups in Q_{14}, most of them with litters from 4 females. Each pair of replicates was made from the progeny of 8 sire-groups and 14 or 15 different females. Since the 20 sire-groups were not exactly divisible by 3, there was some overlap in those contributing to the three pairs of replicates. Two sire-groups contributed to both A–B and C–D, and two others contributed to both C–D and E–F; there was no common contribution to A–B and E–F. The relationships between the replicates can best be expressed as the number of grandparents (parents of Q_{14}) in common. This is given in the legend to Fig. 2.

(iv) *Breeding and selection methods*

Throughout the selection experiment the litters were reared without reduction or augmentation of the number born, and normally only one litter was reared from each mating. The litters were weaned at 3 weeks of age; the numbers in the storage cages were made up to five or six by mixing litters born on successive days. All mice were weighed at 6 weeks of age and, up to generation 8, also at 3 weeks. Selection was based on the weight at 6 weeks of age and was made within litters throughout the experiment. The first selection, made within each of the six replicates, gave rise to 18 lines, a large (L), small (S) and unselected control (C) for each replicate. Each line was bred from eight single-pair matings in each generation. In order to spread the technical work, the matings were set up over a period of 3 weeks in each generation. Replicates A and B were mated in the first week, C and D in the second, and E and F in the third. Thus the mating date was a possible source of variation between replicate pairs, but not within pairs.

Table 1. *Mating schedule in the selection lines*

(Each mated pair and its progeny was given a number, 1–8. The table shows the families of the mated mice and the new mating number assigned.)

Family of origin		New mating number
♀	♂	
1 × 2		1
3 × 4		2
5 × 6		3
7 × 8		4
2 × 1		5
4 × 3		6
6 × 5		7
8 × 7		8

Matings of least relationship were made by a system suggested by Professor Alan Robertson, as shown in Table 1. This system does not reduce the average rate of inbreeding, but it has two advantages over the cyclical system used in the random-bred Q-strain. One is the practical advantage that the mating schedule is the same in every generation, and the second is the more theoretical advantage that the inbreeding coefficients are the same for all families in a generation, and the rate of inbreeding is the same in all generations. With selection of an equal number from all families, the effective population size was 32, and the theoretical rate of inbreeding was 1/64, or 1·56 %, per generation. This gives theoretical coefficients of inbreeding of 15 % at generation 10, 27 % at generation 20, and 30 % at generation 23. The actual coefficients of inbreeding were higher because the number of parents was sometimes reduced below 8 pairs on account of sterile matings. When a mating failed to produce offspring, substitutes were taken from the reciprocal mating or, failing that, from one with the closest relationship to the mating that failed. Reproductive failure occurred with increasing frequency,

and after generation 10 some lines were reduced to 3, or even 2, breeding pairs on a few occasions. When the continuation of a line was seriously endangered additional matings were made and occasionally second litters were reared and used. From generation 19 onwards ten matings were made in all the Control lines, and in the last generation reported here (23) ten matings were also made in all the Large and Small lines.

3. PARAMETERS OF THE RANDOM-BRED Q-STRAIN

The main reason for keeping the random-bred Q-strain for 15 generations before starting the selection was, as already explained, to reduce linkage disequilibrium. Another reason, however, was to accumulate data from which the genetic parameters could be estimated. In particular it was intended to see how closely the realized heritability observed in the selection would agree with the prediction based on the parameters of the random-bred Q-strain, assuming that linkage disequilibrium would not have much effect.

(i) *Means*

The generation-means of all measurements are shown in Fig. 3, and the first six generations of the control lines of the selection experiment are also shown for comparison (birth-weights of these were not available). Each generation mean of the Q-strain is based on about 180 females (maximum 240) and 130 males (maximum 160) in 70 litters (maximum 80). The features of interest in the graphs are the following. Male and female weights varied closely in parallel. The variation between generations was thus real, and not due simply to sampling variance. It is much greater than the variance expected from genetic drift, and most of it must be attributed to differences of environment between generations. Males were more variable in 6-week weight than females, and this was not due simply to the smaller number of males recorded in each generation. The components of variance estimated by analysis of variance (see Table 2) showed that, relative to the total variance within generations, the between-generation component was twice as great in males as in females. It seems, therefore, that the 6-week weight of males was more sensitive than that of females to environmental differences between generations.

The 3-week and 6-week weights follow the same pattern of changes. The dates of the matings, given at the foot of Fig. 3, show that the variation between generations did not follow a seasonal cycle. The increase of the weights at generation 10 corresponds with the first dietary change introduced. Birth weights did not follow the same pattern as the later weights, but followed the litter size in mirror image, though not very closely. The two graphs of litter size are of live young at birth (above) and at weaning (below). The difference between them indicates the pre-weaning losses. Litter size followed the same trends as 3-week and 6-week weights, and the pre-weaning losses were greatest when the weights were low. Males and females suffered equal pre-weaning losses. The sex ratio over all generations was $47\cdot5 \pm 0\cdot5\%$ of females at birth, and $47\cdot9 \pm 0\cdot5\%$ at weaning.

There was significant ($P = 0.05$) heterogeneity between generations in the sex ratio at weaning.

The managemental changes introduced at the start of the selection experiment do not seem to have had any marked effect on the means of the characters measured. Male 6-week weights increased by about 1 g but female weights were unchanged. Three-week weights were increased by about 0.5 g. Litter size at birth was unchanged, but the weaning rate was improved from 88 to 94 %.
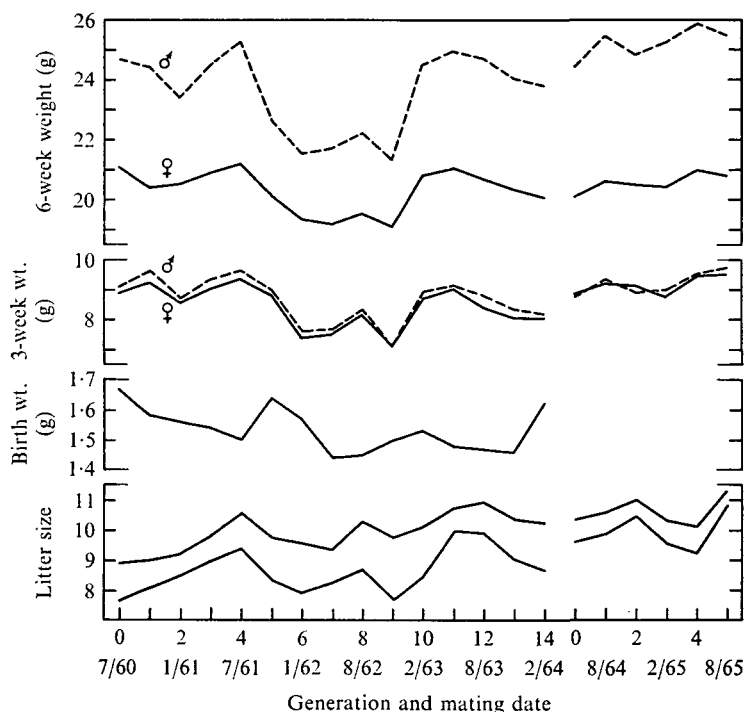

Fig. 3. Mean values in generations 0–14 of the random-bred Q-strain, and in generations 0–5 of the Control lines of the selection experiment. The birth weights are of individuals; the two graphs of litter size are number born alive (above) and number weaned (below). The month and year of mating of alternate generations is shown under the generation number.

(ii) *Variance components and heritability estimates*

Table 2 gives the hierarchical analysis of variance of 6-week weight, and the components of variance derived from it with adjustment for unequal class numbers. The heritabilities, calculated from the sire-components, are much lower than would be expected from previous work, particularly in males. The prediction of the response to selection would be based on the heritability within litters, which was 17 % in females and 5.5 % in males. There is no obvious reason why these estimates are so low. The standard errors, however, even with 281 degrees of freedom for sires, are so large that prediction of the response has very little value.

In view of the very imprecise results of the variance analysis, the heritability

of 6-week weight was estimated also from the regression of offspring on parents, within generations, and the results are given in Table 3. Separate estimates were obtained from female and male offspring, and from the regression on sires and on dams within sires. Unweighted regressions were calculated from the mean of the offspring of each dam or of each sire. Weighted regressions were also calculated but the estimates differed very little from the unweighted ones and are not given. The variances of the two sexes differed, as can be seen from Table 2, males having

Table 2. *Components of variance (g^2) of 6-week weight in a hierarchical analysis of variance of the random-bred Q-strain*

Source	Degrees of freedom		Mean square		Variance component	
	♀	♂	♀	♂	♀	♂
Generations	14	14	92·36	233·50	0·41	1·59
Sires	281	281	16·68	25·00	0·27	0·14
Unrelated dam-pairs	283	281	13·77	23·84	0·27	0·80
Sib-dams	452	451	12·25	20·75	3·51	8·36
Individuals	1730	938	3·11	5·04	3·11	5·04
Total within generations	2746	1951	—	—	7·15	14·34

Within generations	Females	Males	Mean
Full-sib correlation	0·57 ± 0·03	0·65 ± 0·04	0·61 ± 0·03
Additive genetic variance* (g^2)	1·06 ± 0·81	0·56 ± 1·79	0·81 ± 0·98
Heritability, overall† (%)	14·8 ± 11·3	3·9 ± 12·5	9·4 ± 8·4
Heritability, within litters‡ (%)	17·1 ± 13·0	5·5 ± 17·8	11·3 ± 11·0

The standard errors of the heritabilities are based on the method described by Dickerson (1969). The standard errors of the full-sib correlations are based on the assumption of a balanced design and are therefore approximate.

   * 4 × sire-component.

   † (4 × sire-component)/(total within generations).

   ‡ (2 × sire-component)/(individuals).

Table 3. *Heritability (%) of 6-week weight (± S.E.) estimated from the regression of offspring on parents in the random-bred Q-strain*

(Figures in bold type are adjusted for the sex-difference in variance; means are unweighted. Further explanation in text.)

Sex of parent	Sex of offspring		
	Female	Male	Mean
Female	47·3 ± 8·5	64·9 ± 12·8	56·1 ± 7·7
	45·8 ± 9·0		**46·6 ± 6·2**
Male	22·2 ± 5·8	21·9 ± 8·1	22·1 ± 5·0
	31·5 ± 8·2		**26·7 ± 5·7**
Mean	34·8 ± 5·2	43·4 ± 7·6	39·1 ± 4·6
	39·4 ± 5·9	**33·9 ± 6·1**	**36·6 ± 4·2**
Within-litters (h_w^2)	40·0 ± 6·7	61·7 ± 12·7	50·9 ± 7·2
	45·4 ± 7·7	**48·2 ± 10·1**	**46·8 ± 6·3**

about twice the variance of females. This calls for some care in interpreting the regressions when the sex of the offspring and the sex of the parent differ. In these cases two estimates of the heritability are given in Table 3, one, in roman type, being unadjusted, and the other, in bold, being adjusted by the ratio of the standard deviations. (For example, the adjusted estimate from the regression, b, of male offspring on female parents is given by $2 \times b \times \sigma_f/\sigma_m$.) The adjusted estimates are more appropriate for comparison with the estimates from the variance components, and the unadjusted estimates are more appropriate for prediction of the response to selection.

The estimates from female parents are much higher than those from male parents, presumably because of a maternal effect. The estimates from the regressions are more realistic than those from the variance components, and as will be shown, they agreed with the response to selection. When the sexes of parents are averaged, the estimates from the adjusted regressions and from the variance-components differed significantly from each other, the P value being about 0·05 for each sex of offspring separately and 0·01 for the mean of sexes. Since the estimates from the variance components derive from the sire-components, a fairer comparison would be with the regressions on male parents only. The differences are then not significant, the P values being about 0·2 for the separate sexes of offspring and 0·1 for the mean. The very low estimates obtained from the variance components are therefore acceptable as chance deviations. An independent analysis of the random-bred Q-strain was made by Dr L. S. Monteiro, with mice contemporaneous with the base generation of the selection. The heritability of 6-week weight estimated from the sire-component in an analysis of variance was 34 % in females and 22 % in males (Monteiro & Falconer, 1966). These values agree very closely with the estimates from the regressions on male parents (Table 3), which supports the conclusion that the variance components here gave erroneously low estimates.

For comparison with the response to selection, the within-litter heritabilities were calculated from the regression estimates by the relationship $h_w^2 = h^2(1-r)/(1-t)$, where $r = \frac{1}{2}$ and t is the full-sib correlation from the analysis of variance (Table 2). The within-litter heritability, sexes averaged, was $51 \pm 7 \%$ from the unadjusted regressions and $47 \pm 6 \%$ from the adjusted regressions. Neither of these estimates differs significantly from the realized heritability, which was $37 \pm 3 \%$ (Table 5). The maternal effect, which is included in the regression estimates, would be expected to result in these being higher than the realized heritability.

4. SELECTION RESPONSES

(i) *Responses of 6-week weight*

The results of the selection are presented graphically in Figs. 4–7. We start by looking at the overall picture given by the means of all replicates. Looked at in this way, the three groups – Large, Control and Small – represent populations bred from 48 matings, with each generation-mean based on measurements of

21

between 300 and 500 individuals in the Large and Control groups and between 200 and 400 individuals in the Small group. Apart from the effects of any additional inbreeding depression arising from the subdivision, the results are what might have been expected from a single two-way selection experiment on this scale. Fig. 4 shows the means of all Large lines, of all Small lines and of all Control lines plotted against the generation number, with the sexes shown separately. The weight-scale is logarithmic, with the values for males and females superimposed at generation 0. The features of interest are as follows:

(1) The Controls maintained a steady level, with no trend. There was, thus, no indication of any inbreeding depression, though the coefficient of inbreeding
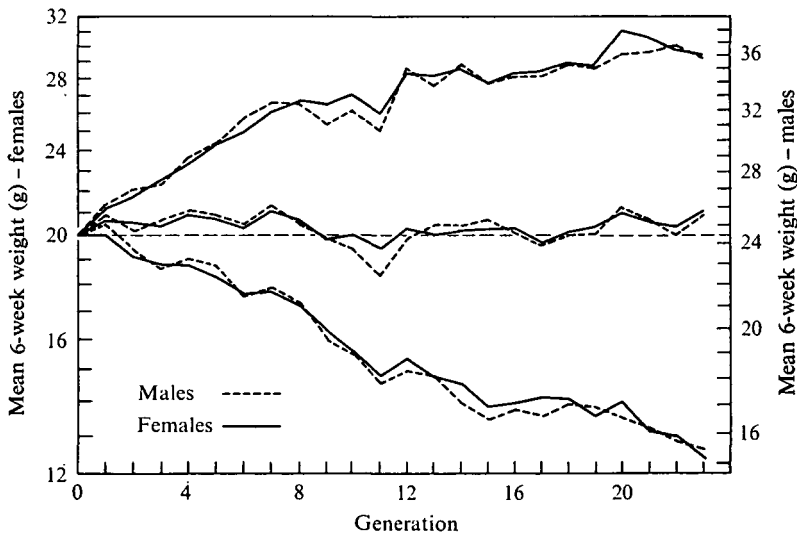

Fig. 4. Mean 6-week weights of the six replicates in each group, Large, Control and Small. The weights are plotted on a logarithmic scale, with the weights of males and females superimposed at generation 0.

reached at least 30 % by the end. It will be shown later, however, that there was inbreeding depression of litter size which concealed what might otherwise have been a depression of body weight.

(2) The responses to selection were proportionately the same in males and in females, so that ratio of male weight to female weight remained constant at 1·2.

(3) The two sexes paralleled each other closely in the major variations from generation to generation. These variations were therefore mainly real and not due to errors in estimating the generation mean, though whether they were due to genetic drift or environmental changes cannot be decided at this stage.

(4) The responses to selection in the two directions were proportionately not very different. At the end the Large lines had increased by about 45 %, and the Small lines had decreased by about 38 % of the Control weights. The ratio of the weights of Large to Small lines at the end was 2·3.

In Fig. 5 the responses are plotted against the cumulated selection differentials.

Here, and in all subsequent graphs, the generation means are the means of the two sexes, and the weight-scales are arithmetic. Fig. 5(a) shows the means of all Large lines and the means of all Small lines, both as deviations from the means of all Control lines. Fig. 5(b) shows the divergence between Large and Small lines. The conclusions are as follows:

(1) The responses approximate very closely to straight lines up to generation 10, at least. Thereafter, the rate of response per unit of selection differential decreased in the Large lines and increased in the Small lines.
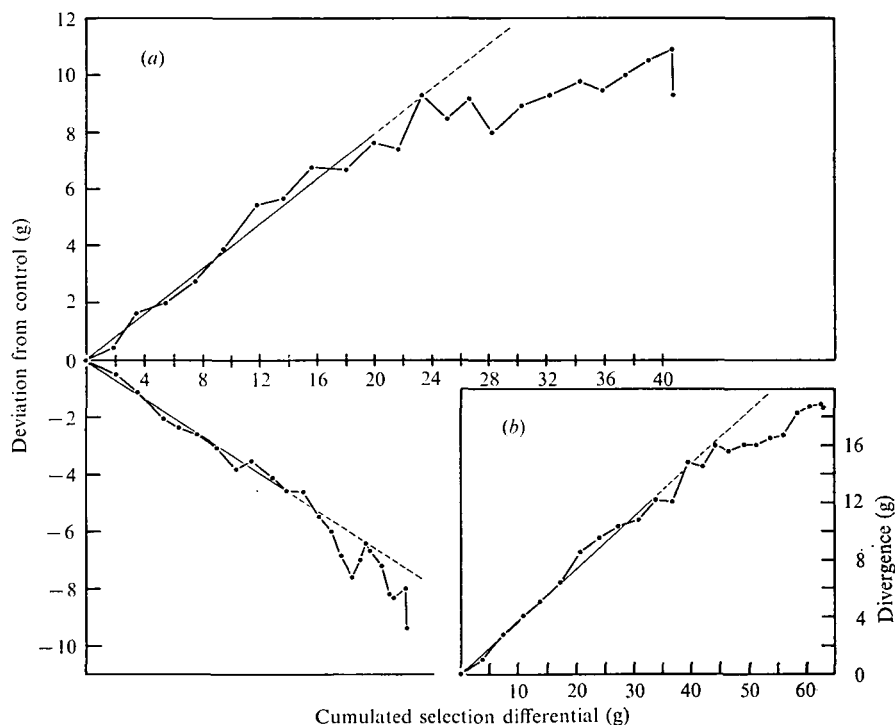

Fig. 5. Responses (sexes averaged) plotted against cumulated selection differentials. The straight lines are the regressions fitted up to generation 10, and not forced through the origin. (a) Means of all Large and of all Small lines, as deviations from the mean of all Control lines (the selection differentials in the Control lines are disregarded). (b) Divergence between the means of all Large and of all Small lines, plotted against the sum of the mean selection differentials.

(2) The realized heritabilities, estimated from the regression of response on cumulated selection differential up to generation 10, were 39·8 % for upward selection, 32·8 % for downward selection and 36·9 % for the divergence. The standard errors of these estimates, and the reality of the asymmetry, will be considered later after the replicates have been described separately. The estimate of 36·9 % from the divergence is consistent with the prediction from the offspring–parent regression in the random-bred strain, as has already been noted.

(3) The graphs do not suggest that the selection limits have been reached after

21-2

23 generations of selection. The fall-off in the response of the Large lines suggests that they are approaching an asymptote of about 12 g above the Controls, corresponding to a mean weight of about 34 g. The Small lines, on the contrary, show no sign of approaching a limit: the selection differentials became much reduced toward the end, with the result that progress per generation was very small.
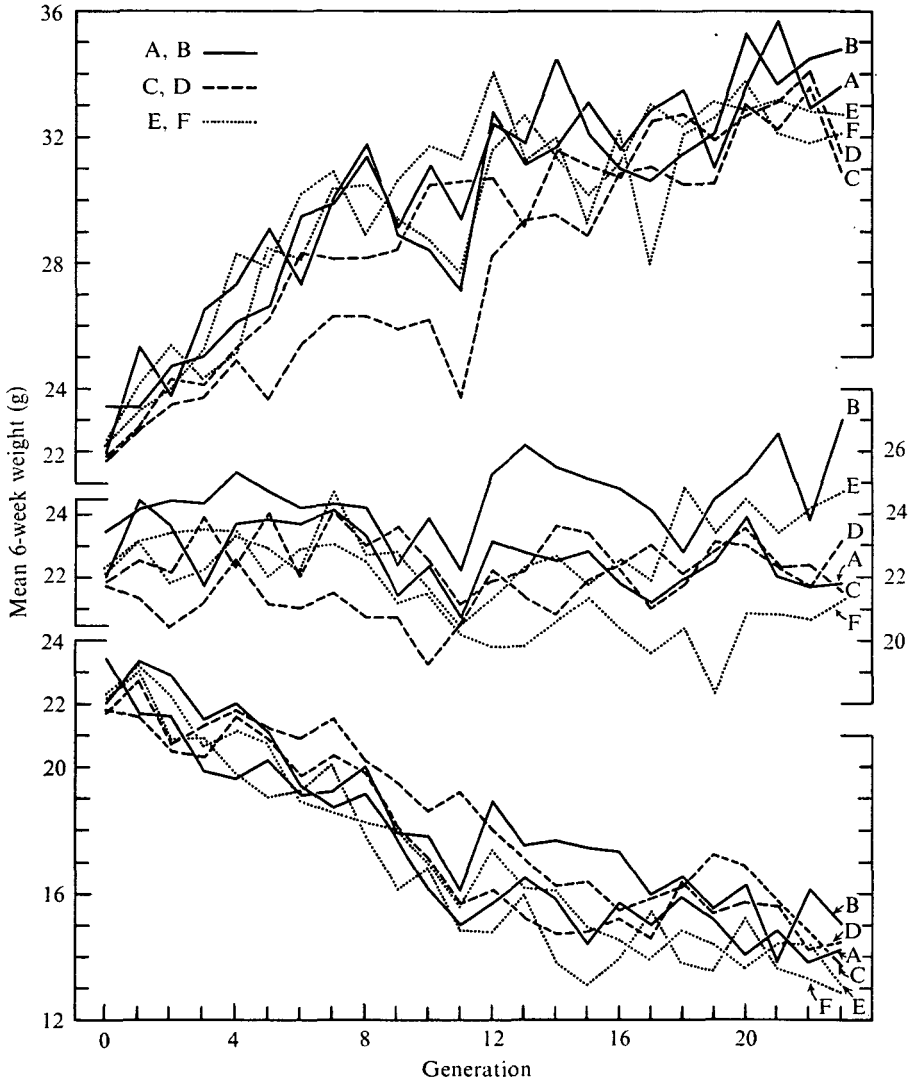

Fig. 6. Generation means (sexes averaged) of all replicates. The vertical scale is divided so as to avoid overlap of Large, Control and Small lines, which of course all started from the same points.

The replicate lines are all shown separately in Fig. 6, with the generation means plotted against the generation number. This graph gives a visual impression of the 'repeatability' of the selection response. We are interested primarily in the

differentiation between the lines. The variance component between lines was calculated in each generation separately, up to generation 20. Table 4(*a*) gives the arithmetic mean of the estimates over 5-generation periods. In order to make some adjustment for the scale effect due to differences in mean weight, coefficients of variation are also given in Table 4(*a*). These are the square root of the component expressed as a percentage of the mean. The conclusions to be drawn from the graph and the variance between lines are as follows, taking the Controls first.

(1) The variation between lines increased, as would be expected from random drift.

Table 4

(*a*) Components of variance (g²) between lines, averaged over 5-generation periods, ± empirical standard errors based on the variance of the component between generations. The between-line coefficients of variation (%) are shown in parentheses

Period	Large	Control	Small
0–4	$0\cdot07 \pm 0\cdot24$ (1·1)	$0\cdot80 \pm 0\cdot26$ (3·9)	$0\cdot18 \pm 0\cdot11$ (2·0)
5–9	$2\cdot88 \pm 0\cdot37$ (5·9)	$0\cdot79 \pm 0\cdot18$ (3·9)	$0\cdot36 \pm 0\cdot13$ (3·1)
10–14	$2\cdot59 \pm 0\cdot88$ (5·3)	$1\cdot55 \pm 0\cdot41$ (5·7)	$1\cdot39 \pm 0\cdot29$ (7·1)
15–20	$0\cdot01 \pm 0\cdot24$ (0·3)	$2\cdot33 \pm 0\cdot54$ (6·8)	$1\cdot22 \pm 0\cdot25$ (7·2)

(*b*) Components of variance (g²) within generations in the Control lines, averaged over generations 0–20

	Females	Males
Between litters	2·27	4·50
Within litters	2·36	4·03
Full-sib correlation	0·49	0·53

(2) The lines tended to maintain their relative positions over considerable periods of time. For example, line B was the highest almost all the way through, line C was lowest over the first ten generations, and line F was lowest over the last 13 generations. The differences between the Control lines thus represented real differentiation and not just sampling error in the estimation of the means. The reality of the differentiation is shown further by the fact that the variance component between lines was significantly different from zero in all four periods, as judged by the empirical standard error.

(3) There was a considerable degree of differentiation from the beginning. This suggests that more differentiation arose from the sampling of the base population than occurred from one generation of drift in the established lines.

(4) There was little, if any, discernible similarity between the pairs of lines that had similar origins and mating dates. This suggests that sampling within families of the base population led to similar amounts of differentiation, and that the dates of mating were not an important source of variation.

(5) There was some parallelism of the lines in the variation between generations. For example, there was a downward trend from generation 4 to 11, and an upward trend from then till the end. This shows that some of the variation between generations was due to real environmental differences. There was, however, much

less detailed parallelism between the lines than was seen between the sexes in Fig. 4. This shows that much of the short-term variation from generation to generation was due to random drift rather than to environmental differences.

Did the differentiation between the Control lines correspond with what would be expected from random drift in lines with an effective population size of 32? The expected variance between lines can be taken to be approximately $2FV_A$, when F is the inbreeding coefficient and V_A the additive variance. The additive variance can be estimated as $V_A = 2h_w^2\sigma_w^2$, where h_w^2 is the realized heritability (0·369) and σ_w^2 is the phenotypic variance within litters. The variance within litters in the Control lines did not change consistently over the course of the experiment and the means of the estimates in each generation are given in Table 4 (b). The mean of the sexes is $\sigma_w^2 = 3\cdot19$, and from this the estimate of the additive variance is $V_A = 2\cdot36\ \mathrm{g}^2$. The theoretical inbreeding coefficients at generations 10 and 20 were 0·15 and 0·27. The expected variance between lines was thus $0\cdot71\ \mathrm{g}^2$ at generation 10 and $1\cdot27\ \mathrm{g}^2$ at generation 20. These values are not very different from the observed values given in Table 4 (a), and the differentiation observed is therefore compatible with the expectation from random drift.

Now consider the selected lines. Lines selected in the same direction did not differ much in the selection differentials applied, so the graphs plotted against the generation number give a fair comparison of the rates of response. The selection differentials will be detailed further below. The conclusions to be drawn are as follows:

(1) The most important feature is that there was no more, and perhaps less, differentiation among the selected replicates than there was among the Control replicates. This suggests in general that the lines did not differ in the genetic variation giving rise to the responses, and that the differentiation between the lines was due primarily to random drift. In the Large lines differentiation increased during the first half of the selection response, but it decreased toward the end, and over generations 15–20 it was virtually zero. This suggests that the Large lines were approaching a common selection limit. The absence of differences between replicates at the limit means that the genetic variation is attributable to a small number of genes with large effects rather than to a large number with small effects (Robertson, 1960). The differentiation between the Small lines increased over the whole experiment and did not decrease at the end, at least when expressed as a coefficient of variation. It was seen earlier that the responses in the Small lines showed no sign of approaching a limit. The differentiation between them suggests that at the end they were in a state comparable to the Large lines in the middle period.

(2) With one exception, there was not much tendency for lines to maintain their relative levels over more than a few generations. On the whole, the graphs of the lines seem to cross over each other more than do those of the Controls. This suggests again that the differences between the replicate lines were mainly random ones due to drift and sampling, rather than to real differences in the genetic

variation. The exception is the C-replicate. Over the first half of the experiment it responded very much more slowly than the other replicates to both upward and downward selection. During the second half of the experiment, however, it caught up with the other replicates. There was nothing in the phenotypic variance or in the selection differentials to account for its slow initial response.

(3) The general conclusion about the 'repeatability', as far as it can be drawn without detailed analysis, is that the selection responses were at least as repeatable
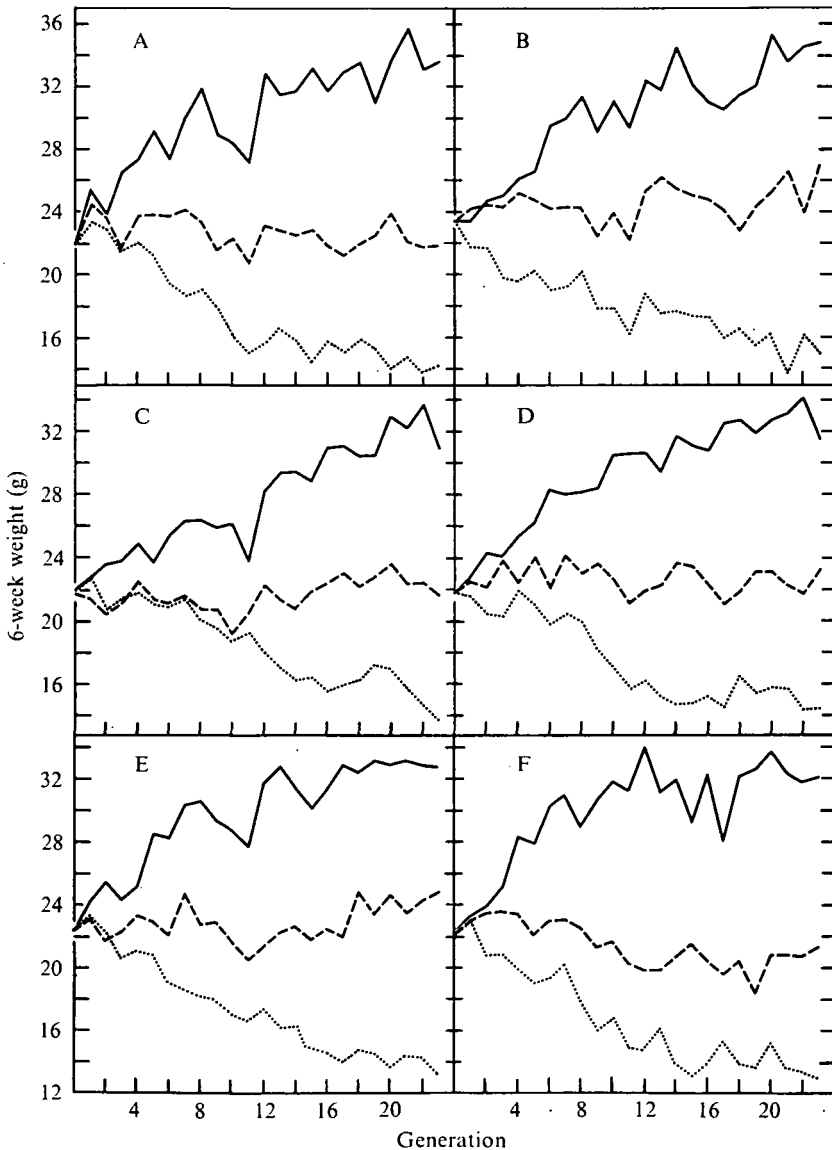

Fig. 7. Generation means (sexes averaged) of each replicate plotted separately.
——, Large lines; – –, Control lines; . . . . . , Small lines.

as were the levels maintained by unselected Controls, and that the differences between the replicates were due mainly to random drift and sampling. Though the replicates seem from this superficial examination to have all reached very much the same end-points, in the character selected, detailed study has shown that they differ in many ways. In particular, the segregating colour genes and protein variants show clearly the effects of random drift (I. Garnett, unpublished).

Finally, Fig. 7 shows the replicates separately, each as a single two-way selection experiment. While Fig. 6 emphasized the similarities between the lines selected in the same direction, Fig. 7 shows forcibly how dissimilar the replicates were over the first part of the selection response, particularly when the upward and downward responses are compared. Replicate C, for example, responded very slowly up to generation 11, and judged by the deviation from the Control there was virtually no downward response during the first 10 generations. In contrast, replicate B showed very little upward response until generation 5 or 6. Comparison of the replicates thus leads to a clear warning: single selection experiments on the scale of one of these replicates can be very misleading about the rate of response, and particularly about the asymmetry, if judged from the first 5 or even 10 generations.

Table 5. *Realized heritabilities in the separate replicates up to generation 10, with responses taken as deviations from the mean of all Controls*

(b = regression coefficient of response on cumulated selection differential; S.E. = standard error of regression, except where marked * and explained in the footnote.)

Rep.	Large		Small		Divergence	
	b	S.E.	b	S.E.	b	S.E.
A	0·390	0·066	0·501	0·058	0·434	0·042
B	0·438	0·047	0·301	0·058	0·375	0·024
C	0·251	0·025	0·159	0·024	0·212	0·020
D	0·457	0·041	0·288	0·041	0·392	0·037
E	0·385	0·051	0·365	0·039	0·379	0·033
F	0·448	0·043	0·376	0·046	0·420	0·032
Pooled*	0·398	0·020	0·328	0·014	0·369	0·014
Mean†	0·395	0·031*	0·331	0·046*	0·369	0·033*
$F_{\frac{5}{54}}(P)$‡	2·59 (< 0.05)		5·90 (< 0.001)		6·40 (< 0.001)	

* Regression of mean of lines on mean selection differential.

† Arithmetic mean of b's with empirical standard error based on variance of b between replicates.

‡ F-ratio testing heterogeneity of b's between replicates, with probability in parentheses.

(ii) *Realized heritability and its sampling error*

The realized heritabilities were calculated for each replicate separately from the regression of response on cumulated selection differential up to generation 10. The response was taken as the deviation of the line-mean from the mean of all Control lines. The regressions and their standard errors are given in Table 5.

Hill (1971, 1972a, b) has shown that the regression coefficient is an unbiased estimate of the heritability, but its standard error is not a valid estimate of the standard error of the heritability. The response estimating the heritability includes the cumulated deviation due to random drift, and so the sampling variance of the heritability is larger than that of the regression coefficient. The replication of the selected lines allows the sampling variance of the realized heritability to be estimated empirically from the observed variance of the regression coefficients between the replicates. In Table 5 the 'pooled' estimates are in each case the regression of the mean of lines on the mean selection differential (as shown in Fig. 5), with the standard error of the regression coefficient, which is not a valid standard error of the realized heritability. The 'mean' estimates in Table 5 are in each case the unweighted mean of the separate regression coefficients in each replicate, with the empirical standard error of this mean. These empirical standard errors are the valid standard errors of the realized heritabilities, and they are about three times the standard errors of the pooled regression coefficients. The conclusions from this analysis are that the realized heritabilities, with their empirical standard errors, were $39\cdot8 \pm 3\cdot1 \%$ and $32\cdot8 \pm 4\cdot6 \%$ for the upward and downward responses respectively, and $36\cdot9 \pm 3\cdot3 \%$ for the divergence. The heritabilities of upward and downward responses were not significantly different ($t_{10} = 1\cdot14$; $P = 0\cdot3$). There is therefore no evidence of a real asymmetry of response. The replicates differed significantly in their regression coefficients both for upward and for downward selection and for divergence. The F-ratios and probabilities are given in Table 5. Without further analysis it is not possible to say whether these differences between the replicates were due entirely to random drift, or whether they were in part due to real differences of heritability.

(iii) *Selection differentials*

The selection differentials calculated were the mean differences in 6-week weight between the selected individuals and the mean of their sex in their litter. Parents that produced no offspring that survived to be weighed at 6 weeks were excluded. This was the only weighting needed since the generation-mean was calculated as the unweighted mean of litter-means. The selection differentials in the Control lines were also calculated, as a check on the randomness of the choice of parents. The total cumulated selection differential in each line is given in Table 6, and the mean selection differentials over 5-generation periods are shown in Fig. 8. The points to be noted are as follows. (1) There was no consistent selection of any importance in the Controls: the lack of inbreeding depression therefore cannot be attributed to unconscious selection. (2) The replicates selected in the same direction did not differ much in their selection differentials, except for one Large line (LF) in the last period. The similarity of the replicates in this respect justifies the comparisons of responses on a per-generation basis made in the previous section. (3) In the Large lines the selection differentials increased in the second period, decreased in the third period and, except for LF, recovered in the last period. The initial increase was due to an increase of variance following the increase of

mean. The reduction in the third period was due to poor productivity, and the further fall in LF was due partly to natural selection, as will be detailed below. (4) In the Small lines the selection differentials fell off continuously to about one third of their initial values. This reduction was due to reduced variance, reduced productivity, and natural selection.

Table 6. *Total cumulated selection differentials* (g), *generations 1–23* (*i.e. parents selected from generations 0–22*)

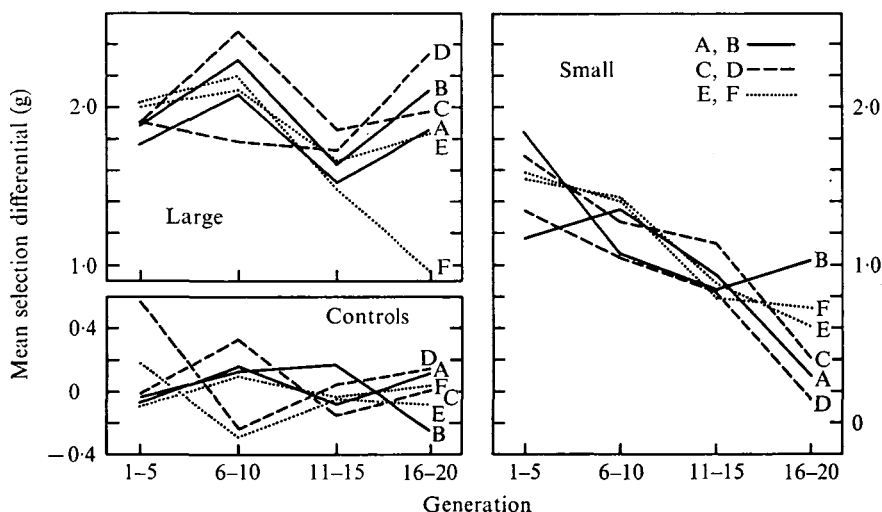| | Replicate | | | | | | |
	A	B	C	D	E	F	Mean
Large	39·3	42·4	42·6	42·7	41·0	35·9	40·6
Small	18·3	25·8	24·4	18·7	22·9	23·2	22·2
Control	+ 1·3	+ 1·2	+ 2·4	+ 1·5	− 1·6	+ 0·4	+ 0·9

Fig. 8. Selection differentials, averaged over 5-generation periods.

The number of matings that produced no offspring surviving to 6 weeks (to be called 'infertile' matings) increased throughout the experiment. Details of these will be given in the next section: here we are concerned with their effects on the selection differentials. When infertile matings occurred in one generation, the selection differential on the next generation was inevitably reduced, because a larger proportion of the offspring of the fertile matings had to be selected to make up the required number of matings. There is also the possibility that infertile matings were the more extreme of the selected individuals, and that they might therefore cause some natural selection opposing the artificial selection. To find out if this was so, the weights of the infertile pairs were compared with those of the fertile pairs in the period from generation 11 to the end. The results are summarized in Table 7. The proportion of infertile matings varied between the replicates: in the Large lines it ranged from 6 % in LD to 31 % in LF, and in the

Small lines from 13 % in SB to 26 % in SE. In considering the differences in weight it must be remembered that these are weights of pairs of mice, and in most cases only one member of the pair is probably responsible for the infertility. Therefore the differences calculated are probably only about half as great as the real differences associated with infertile individuals. In the Large lines the weight differences are not consistent. In three of the replicates the infertile pairs were slightly less heavy than the fertile pairs. In the other three they were considerably heavier – from 0·8 g to nearly 1·0 g. In two of these three (LB and LC) the number

Table 7. *Infertile matings and their effect on the selection differentials in generations 11–23*

	Replicate					
	· A	B	C	D	E	F
Large						
Infertile pairs						
No.	25	20	10	6	20	35
%	24	19	9	6	19	31
Mean wt.-diff.* (g)	− 0·15	+ 0·96	+ 0·80	− 0·33	− 0·20	+ 0·87
Sel. diff.† (%)	102	90	96	101	102	80
Small						
Infertile pairs						
No.	23	14	23	19	27	23
%	21	13	21	18	26	21
Mean wt.-diff.* (g)	− 0·65	− 0·50	− 0·69	0·00	+ 0·05	− 0·81
Sel. diff.† (%)	75	93	84	100	102	79

* Mean difference in 6-week weight between infertile and fertile pairs. (A +ve difference indicates that the infertile mice were heavier.)

† Ratio (%) of actual to potential selection differential, the potential being the selection differential that would have been achieved if all matings had been fertile.

of infertile pairs was not enough for this natural selection to have much effect on the selection differential, but in one (LF) it was. Here the selection differential achieved over generations 11–23 was only 80 % of what it would have been if all pairs had been fertile. This effect was all in the period from generation 16 to 23, when the mean weight difference was + 1·73 g and the selection differential achieved was only 63 % of what it would have been with no infertility. This is clear evidence of natural selection opposing the artificial selection in one Large line. The Small lines are more consistent in showing a fairly large negative weight difference in four of the six lines, i.e. the infertile pairs were less heavy by between 0·5 and 0·8 g. In the other two lines there was virtually no weight difference. In three of the lines with a weight difference, the number of infertile matings was enough to reduce the selection differential to 84 %, 79 %, and 75 % respectively of what it would have been with no infertility. Thus the Small lines give evidence of a more general effect of natural selection opposing artificial selection for small size. In an earlier experiment (Falconer, 1955) natural selection was found to have opposed downward but not upward selection.

(iv) Litter size and productivity

The 'productivity' of the mated pairs provides the best available measure of natural fitness that can be followed through the course of the experiment. It is of interest to see how the productivity was influenced by the selection and by the inbreeding. Three components of productivity can be separately assessed: (a) the proportion of productive matings, i.e. matings that produced one or more live young at birth; (b) the litter size, i.e. the number of live young at birth in litters containing at least one live young; and (c) the weaning rate, i.e. the proportion surviving to weaning of those born alive. The 'productivity' is the number of young weaned per mating made, and is the product of the three components. The productivity measured is restricted to data on first litters, since second litters
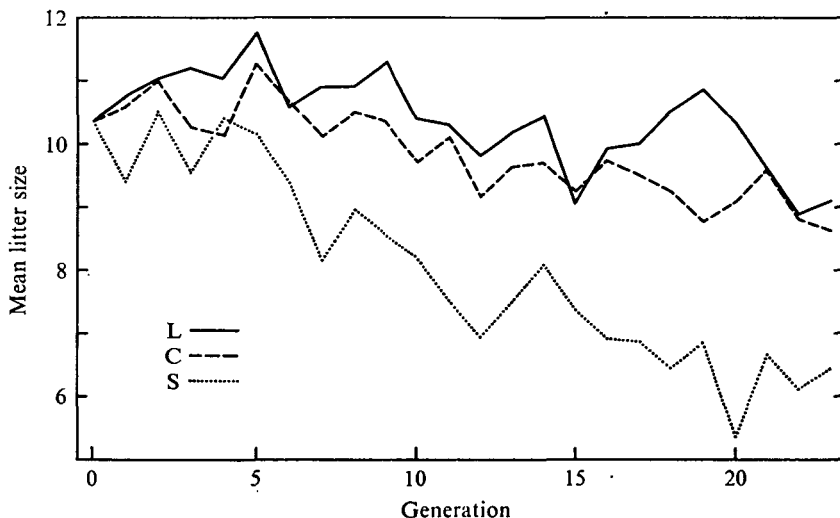
Fig. 9. Mean litter size: i.e. number of live young in first litters containing one or more live young. The generation means shown are the means of line-means in Large, Control and Small groups. Generation 0 is based on 117 matings, the others on about 48 in each group.

were reared only very occasionally. (The number of 'productive' matings dis-cussed here is slightly different from the number of 'fertile' matings discussed in connexion with the selection differential. There were a few matings that reared no young from the first litter but reared young used for selection from a second litter. These were counted as 'fertile' but are here excluded from the 'productive' matings.) It would be possible to take account of post-weaning survival as another component, but very few animals died between weaning and 6 weeks of age and so the survival from birth to six weeks differs little from the weaning rate.

There were seldom more than eight matings per line per generation from which to estimate the components of productivity, and so the sampling variance of the generation means of individual lines was rather large. The lines will therefore not be described individually in detail. The data for generation 0 were based on the

total of 117 matings made, and not on the 48 that formed the foundations of the replicates.

Fig. 9 shows the litter size in each generation in the Large, Control and Small groups of lines, calculated as the means of line means. Litter size is of interest because of two causal relationships with 6-week weight: (1) an individual's 6-week weight is influenced by the size of litter in which it was reared, i.e. its mother's litter size; and (2) a female's 6-week weight influences the size of the litter it subsequently produces. The values of these relationships found in studies of the JC strain, which formed part of the ancestry of the Q-strain, were as follows (Falconer, 1965): (1) regression of female's 6-week weight on mother's litter size $= -0.34$ g per unit of litter size; (2) regression of female's litter size on her 6-week weight, with mother's litter size held constant, $= +0.29$ young per gram. Two expected consequences for the present experiment follow from these relationships: (1) depression of litter size as a result of the inbreeding would be expected to increase body weight, and so to compensate, at least partially, for any inbreeding depression of body weight; and (2) litter size would be expected to increase in the Large lines and decrease in the Small lines, as correlated responses, unless the phenotypic correlation is wholly environmental.

Consider first the Control lines. The litter size declined fairly steadily throughout the course of the experiment, presumably as a result of the inbreeding. The decline was from about 10·5 to about 9·0, and the coefficient of inbreeding at the end was about $F = 30\%$. The rate of inbreeding depression was thus 0·5 young per 10 % increase of F. This agrees well with the value of 0·58 obtained from rapid inbreeding of the JC strain (Bowman & Falconer, 1960). The observed decline of litter size is therefore explained satisfactorily by inbreeding depression. A reduction of 1·5 in litter size would be expected to increase the 6-week weight of females, from the relationship quoted above, by $-1.5 \times -0.34 = +0.51$ g. This may, however, be an underestimate. The regression of 6-week weight on mother's litter size was estimated from the QLA line by Al Murrani (unpublished) as -0.74 in females and -0.97 in males. For application to the Control lines, the regressions should probably be expressed as percentages of the mean weight, and these were -2.6 and -2.8%. The increase in weight of the Control lines expected from the reduced litter size worked out to be $+0.8$ g in females and $+1.0$ g in males. The mean weights of the Controls did not change, as shown in Fig. 1. Therefore the total inbreeding depression of 6-week weight that was masked by the reduced litter size could have been roughly between 0·5 and 1 g, or about 0·2 to 0·3 g per 10 % increase of F. This is a reasonable figure, and so the apparent lack of inbreeding depression of weight can be accounted for by the depression of litter size.

Now consider the selected lines. The initial changes of litter size, over the first five generations or thereabouts, are not inconsistent with the expected correlated responses, in an increase in the Large lines and a decrease in the Small lines. After that, however, the Large lines decreased in parallel with the Controls and roughly 0·5 young per litter above them, while the Small lines decreased

more rapidly. The overall picture is thus one of strong asymmetry of the correlated response.

Fig. 10 shows the other two components and the overall productivity. The points are again the mean of line-means. Generation 0 is shown separately, but thereafter the means of 5-generation periods are shown. Both of the components and the overall productivity declined in all three groups of lines. The proportion of productive matings (Fig. 10*a*) declined equally in the Large and Small lines, and much more in them than in the Controls. The weaning rate (Fig. 10*b*) declined more in the Small lines than in the Large and Control lines which did not differ much. The overall productivity (Fig. 10*c*) presents a clear and interesting picture.
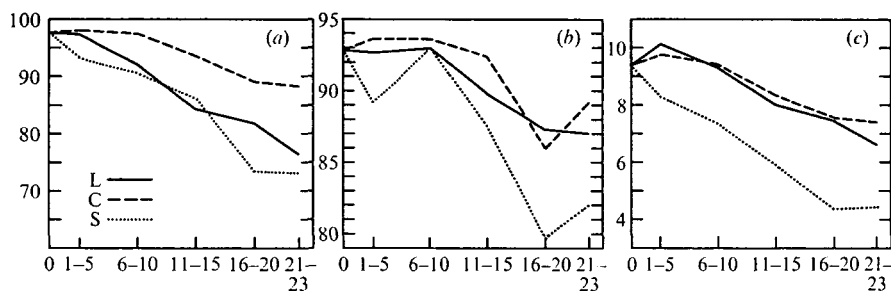
Fig. 10. Aspects of productivity averaged over 5-generation intervals in Large, Control and Small groups. In each case generation 0 is based on 117 matings, the others on about 48 per generation in each group. (*a*) Productive matings (%), i.e. number of matings that weaned at least one young, as percentage of total number of matings made. (*b*) Weaning rate (%), i.e. number weaned as percentage of number born alive. (*c*) Productivity, i.e. mean number of young weaned per mating made.

The Large lines gained a small initial advantage over the Controls, as a result of their increased litter size. But then they fell slightly below the Controls and dropped from a maximum of 10 young per mating to final level of 6·6, while the Controls ended at 7·4. The productivity of the Small lines dropped more rapidly and continuously, to a final level of 4·4 young per mating. Thus natural fitness, as measured by productivity, was adversely affected by selection in both directions, but much more by selection for small size. It cannot, however, be concluded that the body size itself is the cause of the reduced fitness. The reduced fitness may be the consequence of increased homozygosity, and the differences between the Large, Control and Small lines, may be due to differences of homozygosity resulting from the selection.

Looking at the replicates separately shows that they differed in many respects. Some examples of the ways in which certain replicates or lines differed from the overall pattern are as follows. In litter size: replicate A showed no differentiation between the lines until after generation 10; replicate B showed a symmetrical response. In the proportion of productive matings: line LF was particularly bad from generation 11 onward, with under 70 %. Analyses of variance, or χ^2 tests,

showed that the replicates within selection-groups differed significantly in overall productivity and in all three of its components.

(v) *Weaning weight and post-weaning gain*

The weaning weights, at 3 weeks of age, were recorded up to generation 8 in all the lines. It is interesting to find out how much of the responses of 6-week weight were attributable to changes of the weaning weight, which is mainly a characteristic of the mother, and how much to the post-weaning growth of the individuals. The weaning weight is itself influenced by the litter size, and the interrelations of the various characters are complex. These interrelations will not be discussed because their general nature is well enough understood.
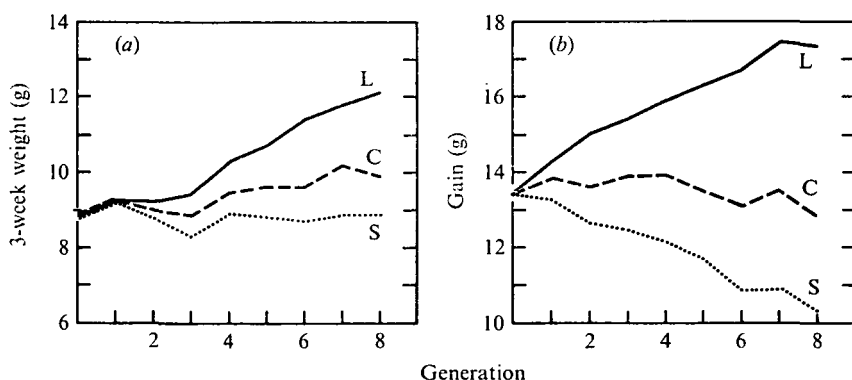

Fig. 11. (*a*) Weight at weaning (3 weeks). (*b*) Post-weaning gain (3–6 weeks). Means of all Large, all Control and all Small lines.

The means of all Large, all Control, and all Small lines are shown in Fig. 11. There was some asymmetry in the correlated responses of weaning weight. Starting from about 9 g, the weights increased to 12 g in the Large lines, and increased also in the Controls, to 10 g; the Small lines, however, showed no change. Growth from 3 to 6 weeks, starting at 13·4 g, increased by 4 g in the Large lines, and decreased by 3 g in the Small lines. The Control lines showed a decrease of about 0·5 g. The conclusions about how the responses of 6-week weight were obtained are as follows. (1) In the Large lines there was an increase of both 3-week weight and gain; in proportion to their initial values, 3-week weight increased by 38 % and gain by 29 %. (2) In the Small lines all the response of 6-week weight came from a reduction of 23 % in the gain. (3) In the Control lines the constancy of 6-week weight masked a 12 % increase of weaning weight and a 4 % reduction of gain. (4) The ratio of post-weaning gain to weight at weaning started at 1·52 and decreased to 1·43 in the Large lines, 1·30 in the Control, and 1·16 in the Small lines, at generation 8.

The replicates are shown separately in Figs. 12 and 13. The post-weaning gain (Fig. 13) does not add much to what was seen in the responses of 6-week weight.

The weaning weights (Fig. 12), however, show very marked differences between the replicates. In particular, replicate C showed almost no differentiation between the three lines.
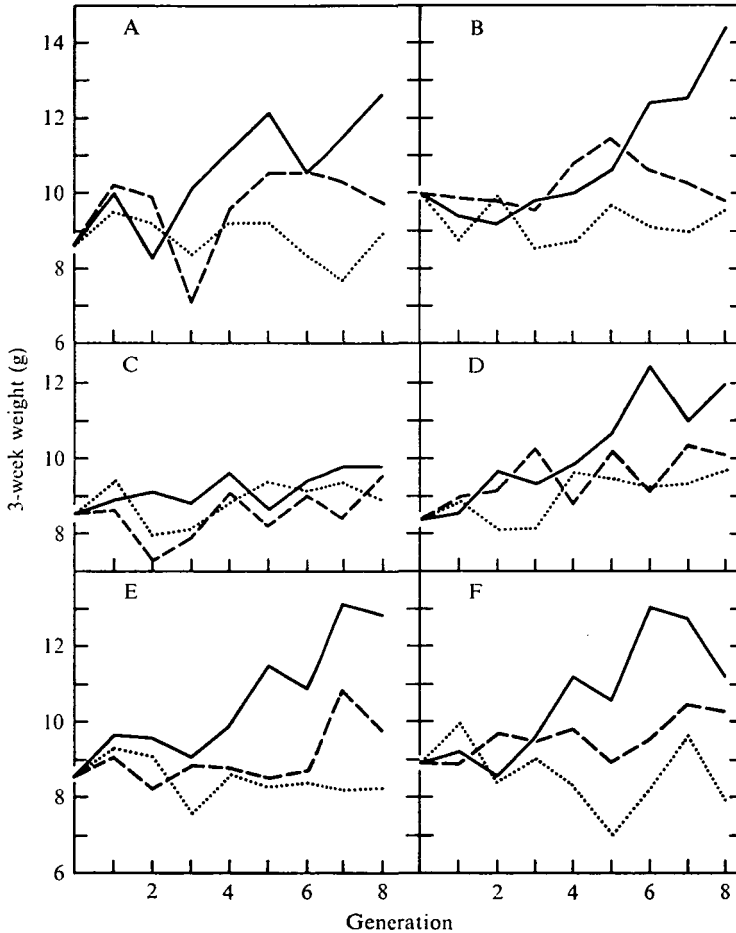

Fig. 12. Weaning weights in the replicates separately. ——, Large lines; – –, Control lines; . . . . ., Small lines.

5. DISCUSSION

One of the aims of the experiment was to test the prediction of the response to selection made from estimates of the heritability in the base population. If the heritability is estimated from the regression of offspring on parents, this may provide the best prediction, in the sense of foreknowledge, of the response to be expected. But the test of the genetic theory that it provides is limited because the regression estimate and the realized heritability are both essentially measures of the degree of resemblance between offspring and parents. Furthermore, the regression estimate and the response in the first generation of selection are derived from data that are very little different: the former includes all the points in the
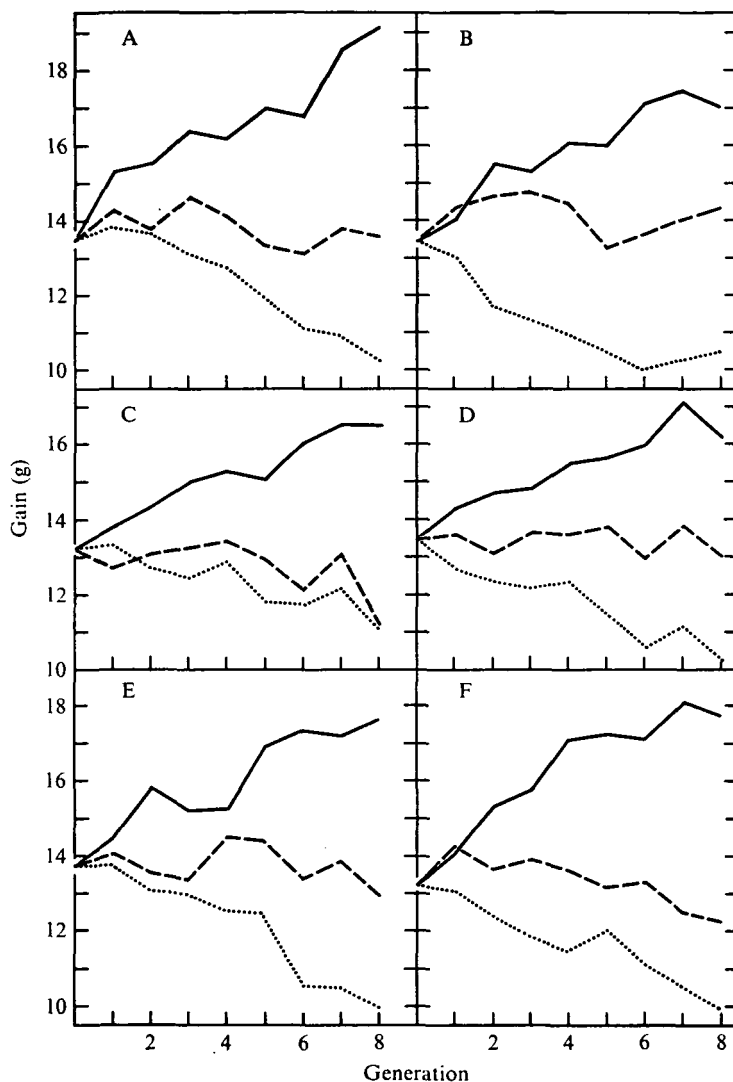
Fig. 13. Post-weaning growth (gain from 3 to 6 weeks) in the replicates separately.
———, Large lines; – –, Control lines; . . . . ., Small lines.

bivariate distribution of parents and offspring, while the latter excludes the more centrally placed points. However, if the realized heritability is estimated from more than just one generation, the comparison tests for the constancy of the genetic parameters. The agreement found between the regression estimate and the realized heritability shows that the prediction held good over more than just the theoretically valid first generation, and that the changes of management between the base population and the selected lines did not materially affect the heritability.

If the heritability in the base population is estimated from the components of variance, the comparison with either the regression estimate or the realized

heritability does provide a test of the genetic theory because the data are analysed in different ways. The comparison tests the theory that accounts for the variance components and for the resemblance between offspring and parents in terms of one parameter, the heritability. The results of these comparisons were unsatisfactory because the estimates of the heritability from the variance components had very large standard errors. The variance components gave estimates that were much lower than the regression estimates and the realized heritability, but the differences were not significant and cannot be regarded as throwing any serious doubt on the validity of the theory, especially in view of the fact that an independent analysis (Monteiro & Falconer, 1966) gave estimates from variance components that agreed very closely with the regression estimates.

The main conclusion from this experiment concerns the importance of random drift in lines selected with an effective population size of 32. Over the first ten generations the six replicates gave widely divergent estimates of the realized heritability and of the asymmetry of the response. The realized heritability estimated from the divergence between upward and downward selection varied between 21 and 43 %. The asymmetry ranged from upward selection being 59 % more effective than downward, to downward being 28 % more effective than upward. The mean responses of all the replicates were very regular and essentially linear up to generation 10. The mean responses represent the results of selection in a population with an effective size of 192.

The response to selection has been widely used to estimate the heritability of the character selected, the term 'realized heritability' being used to distinguish this method from others. The estimation is usually made from the regression of the cumulated response on the cumulated selection differential, and the standard error of the estimate is derived from the standard error of the regression coefficient. In the simplest situation the regression coefficient itself estimates the realized heritability, and the standard error of the regression coefficient is assumed to be a valid estimate of the standard error of the heritability. Hill (1971) has shown, however, that this assumption is not justified because the deviations due to random drift are cumulative. There is consequently more deviation between replicates than would be predicted from the deviations of each replicate about its own regression line. The standard error of the regression coefficient therefore underestimates the standard error of the heritability. This experiment has fully confirmed Hill's ideas. Thus, considering first the divergence between upward and downward selection, when the replicates were pooled and regarded as a single large experiment, the overall regression was 0·369 with a standard error of 0·014. The regressions in the replicates separately, however, differed significantly and the empirical standard error of the mean was 0·033. Thus the standard error of the estimate of the realized heritability was 2·4 times that of the overall regression coefficient.

When the upward and downward responses were considered separately, the overall regression upwards was significantly greater than the regression downwards. But the realized heritabilities in the two directions were not significantly different.

It is clear that in small populations strongly asymmetrical responses can be generated by random drift. The asymmetry of the regression coefficients was highly significant ($P = 0.01$) in two of the six replicates (C and D), and approaching significance in a third (B). Yet no overall asymmetry of the realized heritability was demonstrated. The application of Hill's methods should make it possible to evaluate the standard error of the realized heritability estimated from each line separately, and thus to see if the lines differed in their heritabilities. This remains to be done.

The regression coefficients were estimated from the responses over the first ten generations of selection because the overall regressions, both upward and downward, were linear over this period. After generation 10 the upward regression decreased but the downward regression increased. The significance of this asymmetry in the later stages was not tested because the regressions were no longer clearly linear. This later asymmetry, if real, could be accounted for by the gene frequencies approaching fixation in the large lines but reaching intermediate values in the small lines, or by the selected lines suffering more inbreeding depression than the controls.

Not much can be said at this stage about the limits to selection. The small lines made very little progress at the end, but this was due to the small selection differentials and not to a reduced coefficient of response. The large lines were clearly approaching their limits. The interesting point here is that they all seemed to be approaching the same level at their limits. If borne out by the later generations, this would mean that most of the favourable alleles had been fixed by selection and few of the unfavourable alleles had been fixed by random drift. This, in turn, would indicate that most of the genetic variance was due to few genes with large effects rather than many with small effects (Robertson, 1960). It may not prove possible to get much more information about the levels at the limits because selection had to be suspended after generation 23 on account of the low productivity of some of the lines.

When characters other than 6-week weight were looked at, widespread differences between the replicates were found. The characters described in this paper were the components of productivity, and the replicates differed in the proportion of fertile matings, litter size, and weaning rate. Some of the characters studied, but not yet published, are uncorrelated with 6-week weight and the differentiation in these can readily be attributed to random drift. Others, however, are correlated with 6-week weight, and the differentiation in these suggests that selection has not led to the fixation of the same alleles in all the large lines, as is suggested by the apparent equality of the limits. To account for this contradictory evidence, we may have to consider the regulatory mechanism of growth and body size. It may be that selection can alter the component growth processes more readily than the regulatory process, so that different lines undergo different genetic changes but regulate to roughly the same final size.

Examination of the productivity of the lines showed that selection for small size seriously impaired reproductive fitness. All three components (proportion of fertile

matings, litter size, and weaning rate) were reduced, so that the overall productivity at the end was only about half that of the controls. Selection for large size reduced the proportion of fertile matings and, to a lesser degree, the weaning rate. But litter size was somewhat increased, so that the overall productivity at the end was only a little below that of the controls. The control lines themselves declined in all the components and their overall productivity was reduced by about 20 %. This reduced productivity of the controls was presumably the result of inbreeding which reached a level of about 30 % at the end.

The effect of inbreeding on the litter size of the controls throws light on an earlier experiment with the JC strain. The rate of reduction of litter size in all the control lines together was roughly the same as was found with rapid inbreeding in the JC strain (Bowman & Falconer, 1960). But the JC-control line, maintained by ten pairs per generation, showed no decline of litter size over 30 generations and up to an inbreeding coefficient of 32 % (Falconer, 1960 b). This apparent immunity of the JC-control line from inbreeding depression has remained a puzzle. However, when the separate Control replicates of the present experiment are examined, it is found that they varied in the extent of the inbreeding depression of litter size. In particular, one of the lines (replicate A) showed no decline. This line therefore behaved just like the JC-control. The variation between the replicates is presumably the result of drift variance around an overall mean decline, with the chance that, very roughly one line in six will drift upwards at a rate equal to the mean decline. Thus the lack of inbreeding depression in the JC-control line can be seen as a lucky, but not very improbable, consequence of random drift.

REFERENCES

BOHREN, B. B., HILL, W. G. & ROBERTSON, A. (1966). Some observations on asymmetrical correlated responses to selection. *Genetical Research* 7, 44–57.

BOWMAN, J. C. & FALCONER, D. S. (1960). Inbreeding depression and heterosis of litter size in mice. *Genetical Research* 1, 262–274.

DICKERSON, G. E. (1969). Techniques for research in quantitative animal genetics. In *Techniques and Procedures in Animal Science Research*, pp. 36–79. American Society of Animal Science.

FALCONER, D. S. (1955). Patterns of response in selection experiments with mice. *Cold Spring Harbor Symposia on Quantitative Biology* 20, 178–196.

FALCONER, D. S. (1960a). Selection of mice for growth on high and low planes of nutrition. *Genetical Research* 1, 91–113.

FALCONER, D. S. (1960b). The genetics of litter size in mice. *Journal of Cellular and Comparative Physiology* 56 (Suppl. 1), 153–167.

FALCONER, D. S. (1965). Maternal effects and selection response. *Proceedings XIth International Congress of Genetics* 3, pp. 763–774.

FALCONER, D. S. & KING, J. W. B. (1953). A study of selection limits in the mouse. *Journal of Genetics* 51, 561–581.

FALCONER, D. S. & ROBERTSON, A. (1956). Selection for environmental variability of body size in mice. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 87, 385–391.

HILL, W. G. (1971). Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics* **27**, 293–311.

HILL, W. G. (1972a). Estimation of realised heritabilities from selection experiments. I. Divergent selection. *Biometrics* **28**, 747–765.

HILL, W. G. (1972b). Estimation of realised heritabilities from selection experiments. II. Selection in one direction. *Biometrics* **28**, 767–780.

MONTEIRO, L. S. & FALCONER, D. S. (1966). Compensatory growth and sexual maturity in mice. *Animal Production* **8**, 179–192.

ROBERTS, R. C. (1966). The limits to artificial selection for body weight in the mouse. I. The limits attained in earlier experiments. *Genetical Research* **8**, 347–360.

ROBERTSON, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society* B **153**, 234–249.
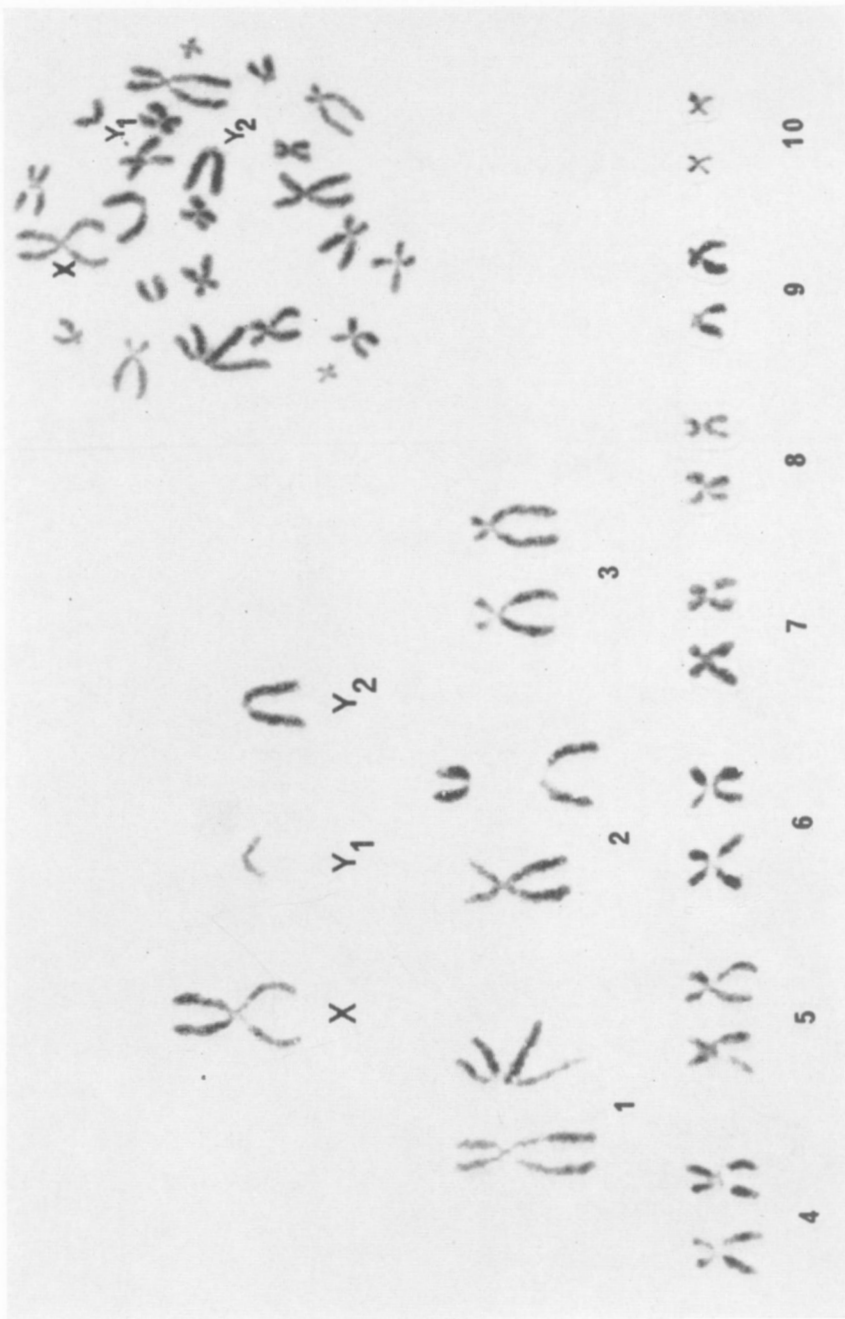
Karyotype and metaphase of the male shrew, *Sorex gemellus*, with dissociated chromosomes = no. 2.

J. OLERT