Protein variation in strains of mice differing in body size

BY I. GARNETT* AND D. S. FALCONER†

Institute of Animal Genetics, Edinburgh 9, Scotland

SUMMARY

Nine (41%) of the 22 enzymic and non-enzymic loci examined in a strain of mice divergently selected for six-week body weight (six lines selected in each direction and six controls) were found to be polymorphic. The degree of polymorphism varied between the replicates from a maximum of 38% to a minimum of 14% with an average individual heterozygous at 7·7% of its loci. There was no obvious association between any of the isozyme variants and body size. The frequency distribution among the 18 lines was adequately accounted for by random genetic drift. However, an association was observed between body size and the *Hbb* locus; the *Hbb^s* allele was found to be fixed in all of the six Large lines. An examination of the variance of gene frequencies at this locus excluded random genetic drift as an explanation for the fixation.

1. INTRODUCTION

The question whether protein variants contribute to the genetic variation of quantitative characters has been the subject of several studies but so far no clear answer to it has emerged. Associations between protein variants and quantitative characters have been examined by comparing allelic frequencies in lines selected in opposite directions for the quantitative character. When an association is found in this way, however, it is not easy to decide between the three possible causes of association: (i) a real effect of the protein variant on the character selected, (ii) a correlation due to linkage disequilibrium, or (iii) a chance effect of random drift. For example, Brown & Nordskog (1962) observed correlated changes in the gene frequencies of two blood-group loci in lines of poultry selected for body weight and for egg weight; the gene frequency changes were greater than could be accounted for by random drift, but linkage disequilibrium was not excluded. Grühn & Dinklage (1971) reported gene frequency changes in the E and L blood-group systems in a line of pigs selected for low body weight, but neither random drift nor linkage were excluded. Negative results were reported by Kraeling, Gerrits & Young (1971) who found no association between the *Tfr* and *PA* loci and back-fat thickness in pigs. Brown (1971) studied six enzyme loci in lines of *Zea mays* selected for protein content and for oil content, and found the gene frequency changes to be compatible with random drift.

The present study was made on the Q-strain of mice maintained at the Institute of Animal Genetics, Edinburgh, and selected for body weight at six weeks of age. The main advantage of this material was that the selection was replicated: there were six lines selected for large size, six selected for small size, and six unselected controls, all derived from the same base population. The replication provided an assessment of the consistency of any association, and allowed the effects of random drift to be assessed by comparing the observed and expected variances of gene frequencies.

The main object of the study was to look for associations of electrophoretically detectable protein variants with body size. A subsidiary object was to assess the degree of polymorphism in a non-inbred laboratory mouse population for comparison with wild populations.

2. MATERIALS AND METHODS

(i) *Population*

The origin and history of the Q-strain of mice is described by Falconer (1973). The features relevant to the present study are the following. The strain was genetically heterogeneous, having originated from a variety of pre-existing non-inbred strains. It was maintained with minimal inbreeding and random choice of parents for 15 generations before selection for 6-week weight was started. Any linkage disequilibrium, except between fairly closely linked loci, was expected to have been broken down in this period. The base population for selection, designated generation 0, was divided into six replicates, A to F. Each replicate was then divided into three, a line (L) selected for large size, a line (S) selected for small size, and an unselected control line (C). There is no evidence from the selection responses (Falconer, 1973) that lines of the same replicate (e.g. LA, SA and CA) were genetically more alike than lines of different replicates. Thus the 18 lines are to be regarded as being random samples of six large lines, six small lines, and six unselected lines. The sets of large, small and control lines will here be referred to as 'size-groups'. Selection, which was made within families, was applied from generation 0 to 23, and each line was maintained by eight single-pair matings with minimal inbreeding. At generation 24 the selected lines were crossed with sub-lines in which selection had been suspended since generation 18, and were continued by 16 matings without further selection. At the same time the control lines were expanded to 16 matings. At generation 31 the selected lines were split in two and selection was applied to one sub-line of each. The animals for the present study came from several generations between 27 and 35, as detailed below. At generation 27 there was approximately a twofold difference in body weight between the large and small lines. The overall mean weights of the Large, Small and Control lines were then 30·91, 15·76 and 22·47 g respectively.

(ii) *Protein variants*

In total 20 enzymic and 2 non-enzymic loci were examined. Table 1 lists all the loci with the relevant information. In the initial survey, conducted in generations 27 and 28, the lines were examined at all the loci except *Adh*, *Tip* and *α-Gpd*.

Table 1. *Summary of biochemical variants assayed*

Protein	Gene[1] symbol	Tissue[2]	Staining method
Alcohol dehydrogenase	*Adh*	L	Briscoe (pers. commun.)
Dipeptidase	*Dip-1*	K	Lewis & Truslove (1969)
Esterase-1	*Es-1*	K	Ruddle *et al.* (1969*a*)
Esterase-2	*Es-2*	K	Ruddle *et al.* (1969*a*)
Esterase-3	*Es-3*	K	Ruddle *et al.* (1969*a*)
Esterase-5	*Es-5*	K	Ruddle *et al.* (1969*a*)
Glucosephosphate isomerase	*Gpi-1*	K + B	Delorenzo & Ruddle (1969)
Glucose 6-phosphate dehydrogenase (auto.)[3]	*Gpd-1*	K	Ruddle *et al.* (1968)
Glucose 6-phosphate dehydrogenase (X-linked)	*Gpd-2*	K	Ruddle *et al.* (1968)
Glutamic oxoacetic transaminase (super.)	*Got-1*	K	Delorenzo & Ruddle (1970)
Glutamic oxoacetic transaminase (mito.)	*Got-2*	K	Delorenzo & Ruddle (1970)
Glycerophosphate dehydrogenase	*Gpd*	L	Briscoe (pers. commun.)
Haemoglobin	*Hbb*	B	Briscoe (pers. commun.)
Isocitrate dehydrogenase	*Id-1*	K	Henderson (1965)
Malic enzyme (super.)	*Mod-1*	K	Shows & Ruddle (1968)
Malic enzyme (mito.)	*Mod-2*	K	Shows & Ruddle (1968)
Malate dehydrogenase (super.)	*Mor-1*	K	Shows & Ruddle (1968)
Phosphoglucomutase (super.)	*Pgm-1*	K + B	Shows *et al.* (1969)
Phosphoglucomutase (mito.)	*Pgm-2*	K	Shows *et al.* (1969)
6-Phosphogluconate dehydrogenase	*Pgd-1*	K	Shaw & Prasad (1970)
Transferrin	*Tfr*	B	Shreffler (1960)
Tripeptidase	*Tip*	L	Lewis & Truslove (1969)

[1] Variants have not been reported at either *Pgd-1* or *Tip*, the remaining loci all exhibit codominance.

[2] Refers to the tissue used in the assay. L, liver; K, kidney; B, blood.

[3] Auto., autosomal; super., supernatant form; mito., mitochondrial form.

The lines of the C and D replicates were surveyed in generation 27 and the remaining lines in generation 28. In the initial survey it was originally intended to type all 32 parents of each line, in order to get exact counts of the gene frequencies. This was done for all segregating loci, but the numbers typed were reduced to save both time and resources when no variant had been found after a number of lines had been fully typed.

Difficulty was encountered with the *Pgm-1* locus through denaturation during prolonged storage, and the data on this locus were supplemented from generations 31 and 34. Additional data on the *Hbb* locus were obtained from both the selected

and relaxed sub-lines in the L and S lines, and in all the C lines in generation 32. The *Adh*, *Tip* and *α-Gpd* loci were typed in generation 35, but not in all lines. *Adh* was typed in 7 lines (LA, LC, LE, CA, CC, CE, SA), *Tip* in 5 lines (LE, CA, CC, CE, SA) and *α-Gpd* in 3 lines (CA, CC, CE); in every case 10 animals of a line were typed.

(iii) *Tissue preparation*

The kidney and/or liver was removed from the animal and stored at − 20 °C until required. Except in the case of *Pgm-1*, storage at this temperature for up to 7 months did not adversely affect the quality of separation on the gel. The tissue was homogenized by a Teflon homogenizer in a volume of distilled water equal to the volume of the specimen. During this operation the homogenization tube was immersed in an ice bath. The homogenates were centrifuged at 4 °C at 10000 *g* for one hour. Following centrifugation a sample of the supernatant was applied to the gel.

(iv) *Haemoglobin and plasma preparation*

Blood samples were obtained either from the caudal artery or by cardiac puncture. All samples were collected into cold, heparinized tubes. Following centrifugation at 100 *g* for 5 min, the plasma was decanted and stored separately. The red-blood-cell pellets were lysed in an equal volume of distilled water. Both the red cell lysate and plasma samples were stored at − 20 °C. It was found that the separation on starch gel of both the haemoglobin and transferrin variants was affected by prolonged storage. However, up to 3 months' storage did not adversely affect separation.

To distinguish clearly all three *Hbb* phenotypes the haemoglobin samples were treated according to a modification of the procedure outlined by Wegmann & Gilman (1970). The lysates were brought to pH 8·7 with a borate–NaOH–EDTA buffer (1 M boric acid; 0·4 M NaOH; and 0·04 M EDTA). To 0·1 ml of the adjusted lysate was added 6 μl of 0·14 M 2-mercaptoethanol and 0·05 ml 2·25 M cystamine diHCl. The lysate was then applied directly to the gel.

(v) *Electrophoretic procedures*

Electrophoresis was conducted on horizontal starch gel. Two types of starch were used; hydrolysed potato starch (Sigma Chemical Co., London) and Electro-starch (Electrostarch Co., Madison, Wis., U.S.A.). For the enzymes and trans-ferrin a gel concentration of 13% was used and for haemoglobin a concentration of 10%. All electrophoresis was carried out at 4 °C with an ice pack placed on the gel. The length of time of the electrophoretic run depended on the size and thick-ness of the gel. Two different-sized gels were employed, 0·6 cm × 20 cm × 23 cm and 0·3 cm × 19·5 cm × 15·5 cm. For the larger gel the duration of electrophoresis was 16 hours with a potential gradient of 300 V over the entire apparatus and a current of 20 mA. For the smaller gel electrophoresis was conducted for six hours with a voltage drop of 250 V and a current of 20 to 25 mA.

Electrophoresis was carried out on three different buffer systems. A tris-citrate,

pH 7·00, buffer (Shaw & Prasad, 1970 (buffer no. III)) was used for the following loci: *Adh, Got-1,2, Gpi-1, Id-1, Mod-1,2, Mor-1,2, Pgm-1,2* and α-*Gpd*. With the exception of transferrin a tris-EDTA-borate, pH 8·60, buffer (Ruddle, Shows & Roderick, 1968) was used for the remaining systems. Transferrin was separated on a 0·3 M borate, pH 8·50 buffer (Shaw & Prasad, 1970 (buffer no. II)).

3. RESULTS

(i) *Enzymic loci*

In the initial survey of generations 27 and 28 variants were found at 7 of the 17 enzymic loci for which all lines were surveyed. No variant was found at any of the 10 other enzymic loci. In addition *Adh, Tip* and α-*Gpd* were invariant in the

Table 2. *Number of individuals typed and the frequencies of the* a *allele of six isozyme loci in the 18 lines*

Size group	Line	Dip-1		Got-2		Gpd-1		Gpi-1		Id-1		Pgm-1	
		n	q_a	n	q_a	n	q_a	n	q_a	n	q_a	n	q_a
Large	A	31	0·73	31	0·36	31	0·0	31	0·42	31	1·00	31	1·00
	B	32	1·00	32	0·0	32	0·31	32	0·0	32	0·33	32	0·77
	C	32	1·00	32	0·12	32	1·00	32	0·16	32	0·0	32	0·48
	D	32	0·73	31	0·27	32	0·09	32	0·52	32	0·12	30	0·25
	E	32	0·89	32	0·0	32	0·25	32	0·47	32	0·0	32	1·00
	F	30	0·82	32	0·0	30	0·13	32	0·0	30	0·77	30	0·60
	Mean	189	0·86	190	0·12	189	0·13	191	0·26	189	0·37	187	0·68
Control	A	31	0·27	30	0·23	31	0·42	31	0·27	26	1·00	29	0·21
	B	31	0·19	31	0·24	31	0·21	31	1·00	31	0·19	12	0·88
	C	31	1·00	30	0·0	31	0·53	30	0·20	30	0·53	16	0·84
	D	30	0·37	30	0·05	30	0·22	30	0·07	30	0·0	29	0·40
	E	30	0·88	31	0·0	30	0·75	31	0·56	30	0·52	13	0·77
	F	32	0·97	32	0·0	31	0·16	32	0·17	32	0·42	13	0·88
	Mean	185	0·61	184	0·09	184	0·38	185	0·38	179	0·44	112	0·66
Small	A	32	0·91	32	0·0	32	0·37	32	0·11	30	0·07	29	1·00
	B	32	1·00	32	0·06	32	0·23	32	0·44	32	0·17	20	1·00
	C	32	1·00	32	0·91	32	0·0	32	0·58	32	0·16	19	0·54
	D	32	0·41	32	0·30	32	0·19	32	0·33	32	0·42	14	0·57
	E	32	0·98	32	0·0	32	0·0	32	0·03	32	0·38	9	1·00
	F	32	0·61	32	0·0	32	0·0	32	0·0	32	0·58	27	0·63
	Mean	192	0·82	192	0·21	192	0·10	192	0·25	190	0·30	118	0·79

limited number of lines surveyed. The loci that were polymorphic in the Q-strain as a whole, i.e. at which variants were found, were *Dip-1, Es-3, Gpi-1, Gpd-1, Got-2, Id-1* and *Pgm-1*. With the exception of the LD and SD lines none of these loci segregated in every line. Table 2 gives the number of animals typed in each line and the gene frequency of each of the variant loci, with the exception of *Es-3* which segregated in only two lines. The variation in gene frequency between lines and size groups is shown diagrammatically in Fig. 1. Except for *Es-3*, every locus
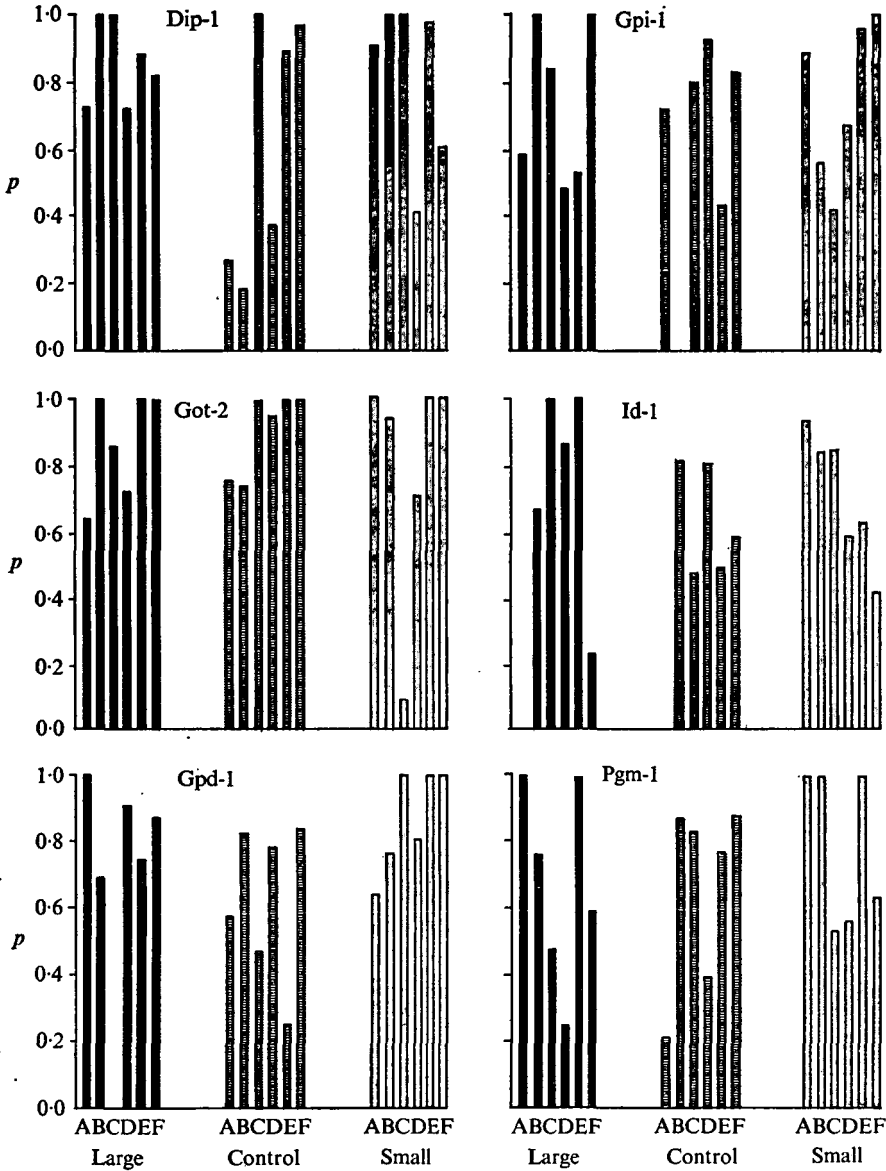
4-2

Fig. 1. Distributions of gene frequencies of the six polymorphic enzyme loci in the Large, Control and Small lines of the Q-strain. The gene frequency (*p*) in each line is shown by the height of the vertical bar.

showed a wide range of gene frequencies. The three size-groups, however, showed little difference in their mean gene frequencies at any locus. There was thus no evidence of association between any of the enzymic loci and body size. The question whether the variation between lines can be adequately accounted for by random drift will be examined in a later section.

The genotype frequencies at all loci where the number permitted were tested

by χ^2 for agreement with Hardy–Weinberg expectations. Altogether 59 such tests were made of which 6 yielded individual probabilities less than 5% and one a probability less than 0·5%. Applying the criterion of Cooper (1968) for multiple sampling, none was found to be significant at the 5% level. There was thus no evidence of any departure from random genotype frequencies within any line.

(ii) *Non-enzymic loci*

The transferrin locus was segregating in seven of the lines, in all cases the Tfr^b allele was at low frequency and only appeared in the heterozygous state. Gene frequencies were not calculated because the separation on the starch gel was

Table 3. *Frequencies of the* Hbb^s *allele* (q_s) *and of the albino allele* (q_c)

Size group	Line	Generations 27 and 28[1]			Generation 32[2]		
		n	s/s	q_s	n	q_s	q_c
Large	A	30	30	1·00	32	1·00	0·03
	B	32	32	1·00	32	1·00	0·0
	C	32	31	0·98	32	1·00	0·0
	D	31	30	0·98	32	1·00	0·08
	E	32	32	1·00	32	1·00	0·29
	F	29	29	1·00	32	1·00	0·0
	Mean	186	184	0·99	192	1·00	0·07
Control	A	30	8	0·50	30	0·33	0·05
	B	30	0	0·0	32	0·12	1·00
	C	25	0	0·0	26	0·0	0·50
	D	29	0	0·0	32	0·0	0·0
	E	30	17	0·74	32	0·67	0·0
	F	31	1	0·18	30	0·20	0·59
	Mean	175	26	0·24	182	0·22	0·36
Small	A	32	3	0·31	32	0·42	0·91
	B	30	1	0·18	32	0·06	0·97
	C	31	2	0·25	32	0·02	0·08
	D	32	1	0·18	31	0·39	0·59
	E	32	32	1·00	32	1·00	0·45
	F	32	0	0·0	32	0·0	0·0
	Mean	189	39	0·32	191	0·30	0·50

[1] Replicates C and D were types in generation 28, all the others in generation 27. The gene frequency, q_s, was estimated as the square root of the proportion of Hbb^s/Hbb^s individuals.

[2] The gene frequency, q_s, was estimated from counts of the three genotypes.

unreliable. However, a sufficient number of satisfactory gels were obtained to be certain that both alleles were present in the Q-strain at generations 27–28. A comparison with freshly collected samples of plasma suggested that prolonged storage could have adversely affected separation. In some instances the samples had been stored for nearly 12 months.

The haemoglobin samples from generations 27 and 28 were also affected by prolonged storage resulting in inadequate separation of the Hbb^d/Hbb^d and the

Hbb^d/Hbb^s phenotypes. As a result, the gene frequencies in these generations have been estimated from the frequency of the Hbb^s/Hbb^s individuals, on the assumption of Hardy–Weinberg frequencies. In the repeated survey of generation 32 the typing was done with minimum storage and the differences between the three genotypes were clearly discernible. The gene frequencies were therefore determined by simple counting. Table 3 gives the numbers typed and the gene frequencies in the two surveys; the gene frequencies in the second survey are shown in Fig. 2.
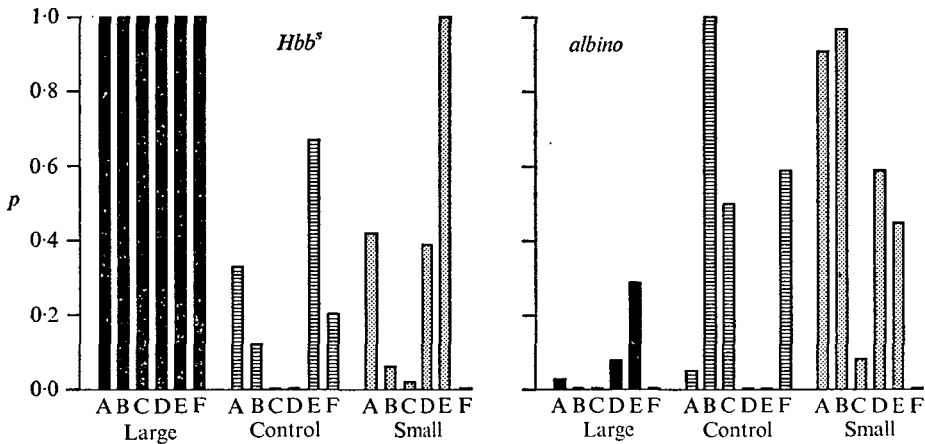

Fig. 2. Distributions of gene frequencies of the Hbb^s and the *albino* alleles in the 18 lines of the Q-strain at generation 32. The gene frequency (p) in each line is shown by the height of the vertical bar.

In contrast to all the other loci studied, the Hbb locus showed an apparent association with body size. In generation 32 the Large lines were all fixed for the Hbb^s allele, whereas the Control lines showed a fairly wide range of gene frequencies with a mean of 0·22. The association with size was, however, not straightforward because the Small lines did not, as would be expected, have frequencies of Hbb^s lower than the Controls. One Small line, SE, was fixed for Hbb^s and the mean gene frequency in the Small lines was 0·30. The reality of the association of Hbb^s with large body size was tested by the Mann–Whitney U-test. This non-parametric test was used because the distribution of gene frequencies is very far from being normal. The difference between the Large and Control groups of lines in generation 32 was highly significant ($U = 36$; $P < 0.01$), and so was the difference between the Large and Small lines ($U = 33$; $P < 0.02$). It is thus clear that the fixation of Hbb^s in all the Large lines cannot be ascribed to chance. The difference between the Small and Control lines was, however, not significant ($U = 21$; $P < 0.2$). The reason for the association of the Hbb locus with large body size but not with small will be discussed later. The genotype frequencies at the Hbb locus were tested against the Hardy–Weinberg expectations in four of the lines in generation 32; no significant departure was found.

(iii) *Variance of gene frequency*

The random nature of the distributions of gene frequencies among the lines can be tested by comparison of the observed variance of gene frequency with the expected variance derived from the inbreeding coefficient. The expected variance of gene frequencies is given by $\sigma_E^2 = p_O (1-p_O)F$, where p_O is the initial gene frequency, which is equal to $\bar{p}$, the mean gene frequency among the lines in the

Table 4. *Expected and observed variance of gene frequencies for the five isozyme loci,* haemoglobin *and the three coat-colour genes*

Locus	$\bar{p}$	σ_E^2	σ_O^2	σ_O^2/σ_E^2	χ^2 (17)
Dip-1	0·764	0·068	0·077	1·13	19·21
Gpd-1	0·215	0·064	0·042	0·66	11·22
Gpi-1	0·297	0·079	0·072	0·91	15·47
Got-2	0·141	0·046	0·052	1·13	19·21
Id-1	0·370	0·088	0·102	1·16	19·72
a	0·721	0·087	0·051	0·74	7·62[1]
b	0·765	0·079	0·076	0·96	12·48[1]
c	0·308	0·095	0·135	1·42	24·14
Hbb	0·512	0·110	1·190	1·72	29·27

[1] df = 13.

random situation, and F is the coefficient of inbreeding. If the frequency distribution has been solely a function of the sampling process unaffected by any systematic processes—selection, migration and mutation—a close correspondence is expected between the observed variance (σ_O^2) and the expected variance (σ_E^2) based on $\bar{p}$ and F.

An expectation of the variance of gene frequencies was calculated for five of the isozymes and haemoglobin and compared with the observed variances. *Pgm-1* was excluded since the frequency estimates were not contemporaneous. The allele frequencies estimated in generation 32 were used for the *Hbb* locus. To allow a comparison of the *Hbb* locus with gene frequencies calculated in a contemporary generation, three coat-colour genes were included in the analysis (*non-agouti* (*a*), *brown* (*b*) and *albino* (*c*)). The initial gene frequencies of these were available from population records for the coat-colour loci, and were used in place of the mean gene frequencies for calculating the expected variance. The inbreeding coefficients were calculated from pedigree records (Cruden, 1949) from generation 0 to 30 for each line. The average level of inbreeding for the purpose of calculating the expected variance for the isozyme loci was based on the average of the 18 lines at generation 27. The average inbreeding for the strain was 0·379. A coefficient of inbreeding had not been calculated by pedigree for the individuals of generation 32. However, extrapolation over two generations was thought to be sufficiently accurate. This gave an average value of $F = 0·44$.

The gene frequencies, observed and expected variances, the ratio of the variances and χ^2 values are presented in Table 4. Except in the case of *Gpd-1* the isozyme loci show close agreement between the two sets of values. The χ^2 in all cases show

no significant difference between the observed and expected variances. The *haemoglobin* locus shows an inflated variance, as was to be expected from its association with body size, and the increased variance is significant by the χ^2 test ($P < 0.05$). The *non-agouti* and *brown* loci, like the isozyme loci, show reasonable agreement between the observed and expected variances, but the *albino* locus, like *Hbb*, shows an inflated variance which approaches significance ($P \sim 0.10$).

A non-random distribution of the *albino* gene frequencies might be expected because *albino* and *Hbb* are closely linked (about 6 crossover units apart) on chromosome 7, i.e. linkage group I (Hutton & Roderick, 1970). If there were linkage disequilibrium, *albino* might also show association with body size or with the *Hbb* locus. The *albino* gene frequencies were therefore looked at more closely. The gene frequencies in all the lines in generation 32 are given in Table 3 and Fig. 2. The Large, Control and Small groups of lines differ in a regular manner in mean gene frequency, the Large lines having a lower frequency of the *albino* gene, and the Small lines a higher frequency, than the Controls. Comparison of the gene frequencies in the Large and Small groups of lines by the Mann–Whitney test showed the difference to be approaching significance ($U = 30$; $P < 0.1$). The association with body size, if real, would be enough to account for the inflated variance. When the variance of gene frequency was recalculated within the Large, Control and Small groups and pooled, the ratio of observed to expected variance was reduced from 1·42 to 1·21. Association between *albino* and *Hbb* was tested by Spearman's rank correlation between the gene frequencies of *albino* and *Hbb^s* in each line. The correlation was -0.18, which is far from being significant. Thus, in summary, the *albino* locus shows an over-dispersion among lines, a regular association with body size (though neither of these were proved significant), and almost no association with *Hbb*.

4. DISCUSSION

(i) *Association with body weight*

To consider the simplest case, a protein variant contributing to six-week body weight would be expected to increase in frequency to fixation in the corresponding selected line. The gene frequency distributions revealed no obvious correlation between any of the enzyme loci and body size. This observation was confirmed in a comparison of the expected and observed variance of gene frequencies. The close agreement between the two sets of variances clearly showed that random genetic drift was a sufficient explanation of the distributions of gene frequencies at the five isozyme loci. There was thus no evidence that either the selection for body weight or any form of stabilizing natural selection was operating on these loci.

On the other hand, the distribution of gene frequencies at the *haemoglobin* locus, and probably also the closely linked *albino* locus, cannot be attributed to random drift. The association of both loci with body weight shows that both were in some way affected by the selection for weight. The effect on *albino* could be readily explained by supposing that the *albino* allele itself reduces body size so that the gene frequency increased in the Small lines and decreased in the Large lines.

Evidence suggests, however, that the effect on both loci did not result from a direct contribution of either gene to body size. A study on the Small lines of the Q-strain failed to detect any effect of the *albino* genotype on body size (Al-Murrani, 1973). The gene frequency at the *Hbb* locus was changed only by selection for large size and not by selection for small size. Further experiments, which will be described in a later paper, showed that the two *Hbb* alleles did not themselves affect weight. One must therefore postulate linkage disequilibrium between the two loci and a neighbouring locus that affected weight. The diverse origin of the Q-strain (Falconer, 1973) makes it quite possible for the *Hbb* and *albino* loci to have entered the base population in linkage disequilibrium with a size locus. The crosses most likely to have introduced the disequilibrium were made 25 generations before the selection was started. Even after this period of random mating closely linked loci would be expected to have retained a considerable amount of the original disequilibrium. For example loci six units apart, such as *Hbb* and *albino*, would still have 20 per cent of the original disequilibrium. It is therefore not unreasonable to postulate linkage disequilibrium as a cause of the non-random distributions of gene frequencies.

(ii) *Extent of electrophoretic protein variation*

Extensive surveys of biochemical variation in non-inbred strains of mice have not appeared in the literature. From two extensive surveys on feral *Mus* populations in the south-western United States and Denmark, Selander, Hunt & Yang

Table 5. *The number and percentage (in parentheses) of enzymic and non-enzymic loci segregating in each line, out of 21[1] loci surveyed*

	Line					
Size group	A	B	C	D	E	F
L	3(14)	3(14)	4(18)	8(38)	3(14)	4(18)
C	6(27)	6(27)	5(23)	6(27)	6(27)	5(23)
S	5(23)	5(23)	5(23)	8(38)	3(14)	4(18)

[1] The *Tfr* locus has not been included (see text).

(1969) and Selander & Yang (1969) estimated that the average population is segregating at 26% of its loci and an individual is heterozygous at 8·5% of its loci. In a study covering a limited sample of mice from various North American locations Ruddle *et al.* (1969*b*) found 29% of the loci they examined to be polymorphic.

Of the 22 loci surveyed in the Q-strain, nine (41%) were found to be polymorphic. The degree of polymorphism varied considerably between the lines from a maximum of 38% to a minimum of 14%, as shown in Table 5. Transferrin is not included in these estimates since in some cases the starch gels could not be interpreted with complete accuracy. For the purpose of comparison with feral populations the average number of polymorphic loci was estimated from the control lines. An average of 23% of the loci were found to be polymorphic and the average

individual was heterozygous at 7·7% of its loci. This last value does not include *Pgm-1* since the survey of this locus was carried out over several generations. It is interesting to note the similarity in the degree of variation between the Q-strain, a population with an extensive laboratory background, and feral populations.

On first consideration, such forces as the founder effect, population number and consistency of environment would lead one to expect a lesser degree of variation in a laboratory than in a natural population. However, each laboratory strain must be considered a unique genetic sample. The extent of biochemical variation would depend primarily on several factors: (i) the 'history' of the strain; (ii) the effective size at which the population is maintained and the consequent degree of inbreeding; and (iii) the loci surveyed.

Abundant ecological data has shown that feral *Mus* populations are highly structured and consist essentially of closed breeding units of small size with little migration occurring (see Anderson, 1970). Petras (1967) has estimated the effective size of the breeding unit to be between 6 and 80 individuals. Thus if one considers, on the one hand, the structure of the feral population and, on the other, the heterogeneity and the effective size of the Q-strain, the degree of variation observed should not be unexpected.

The Q-strain was established with the objective of achieving a broad spectrum of genetic variation, and consequently it had a substantially heterogeneous background. Prior to selection the Q-strain had been random bred for 15 generations with an estimated rate of inbreeding of 0·54% per generation. During the course of selection the 'idealized' effective number (N_e) varied from 32 to 64. Due to differential fertility, fecundity and progeny viability the 'actual' N_e would fall short of the 'idealized' N_e. However, a comparison of the theoretical inbreeding coefficient with that obtained by pedigree data, 0·33 and 0·38 respectively, indicates that the discrepancy between the 'idealized' and 'actual' N_e was not large.

In any study attempting to assess the extent of biochemical variation in a population it is essential that a random sample of proteins be assayed. Since this study was not specifically designed to estimate the extent of polymorphism, the choice of loci was biased towards those systems previously shown to be segregating in *Mus* populations. This bias could possibly have inflated the degree of variation observed in the strain. Had a larger sample of loci been surveyed it is conceivable that the laboratory and natural situation might have shown less agreement.

REFERENCES

AL-MURRANI, W. (1973). Ph.D. Thesis, University of Edinburgh.

ANDERSON, P. K. (1970). Ecological structure and gene flow in small mammals. *Variation in Mammalian Populations*, ed. E. J. Berry and H. N. Southern. *Zoological Soc. of London Symposia*, vol. 26, pp. 299–325.

BROWN, A. H. D. (1971). Isozyme variation under selection in *Zea mays. Nature* **232**, 570.

BROWN, R. V. & NORDSKOG, A. W. (1962). Correlated responses in blood group frequencies with selection for body weight and egg weight in the fowl. *Genetics* **47**, 945 (abstr.).

COOPER, D. W. (1968). The significance level in multiple tests made simultaneously. *Heredity* **23**, 614–617.

CRUDEN, D. (1949). The computation of inbreeding coefficients in closed populations. *Journal of Heredity* **40**, 248–251.

DELORENZO, R. J. & RUDDLE, F. H. (1969). Genetic control of two electrophoretic variants of glucosephosphate isomerase in the mouse *(Mus musculus). Biochemical Genetics* **3**, 151–162.

DELORENZO, R. J. & RUDDLE, R. H. (1970). Glutamate oxalate transaminase *(Got)* genetics in *Mus musculus*: linkage, polymorphism and phenotypes of *Got-2* and *Got-1* loci. *Biochemical Genetics* **4**, 259–273.

FALCONER, D. S. (1973). Replicated selection for body weight in mice. *Genetical Research* **22**, 291–321.

GRÜHN, R. & DINKLAGE, H. (1971). Blutgruppen- und proteinpolymorphismus beim Gottingen miniaturschwein. *Zeitschchrift fur Versuchstierkunde* **13**, 179–187.

HENDERSON, N. S. (1965). Isozymes of citrate dehydrogenase: subunit structure and intracellular location. *Journal of Experimental Zoology* **158**, 263–274.

HUTTON, J. J. & RODERICK, T. H. (1970). Linkage analyses using biochemical variants in mice. III. Linkage relationships of eleven biochemical markers. *Biochemical Genetics* **4**, 339–350.

KRAELING, R. R., GERRITS, R. J. & YOUNG, E. P. (1971). Transferrin and pre-albumin polymorphisms in swine selected for backfat thickness. *Journal of Animal Science* **32**, 174–178.

LEWIS, W. H. P. & TRUSLOVE, G. H. (1969). Electrophoretic heterogeneity of mouse erythrocyte peptidases. *Biochemical Genetics* **3**, 493–498.

PETRAS, M. L. (1967). Studies of natural populations of *Mus*. I. Biochemical polymorphisms and their bearing on breeding structure. *Evolution* **21**, 259–274.

RUDDLE, F. H., SHOWS, T. B. & RODERICK, T. H. (1968). Autosomal control of an electrophoretic variant glucose-6-phosphate dehydrogenase in the mouse *(Mus musculus). Genetics* **58**, 599–606.

RUDDLE, F. H., SHOWS, T. B. & RODERICK, T. H. (1969a). Esterase genetics in *Mus musculus*: expression, linkage and polymorphism of locus *Es-2. Genetics* **62**, 393–399.

RUDDLE, F. H., RODERICK, T. H., SHOWS, T. B., WEIGL, P. G., CHIPMAN, R. K. & ANDERSON, P. K. (1969b). Measurement of genetic heterogeneity by means of enzyme polymorphisms. *Journal of Heredity* **60**, 321–322.

SELANDER, R. K., HUNT, W. G. & YANG, S. Y. (1969). Protein polymorphism and genic heterozygosity in two European subspecies of the house mouse. *Evolution* **23**, 379–390.

SELANDER, R. K. & YANG, S. Y. (1969). Protein polymorphism and genic heterozygosity in a wild population of the house mouse *(Mus musculus). Genetics* **63**, 653–667.

SHAW, C. R. & PRASAD, R. (1970). Starch gel electrophoresis of enzymes – a compilation of recipes. *Biochemical Genetics* **4**, 297–320.

SHOWS, T. B. & RUDDLE, F. H. (1958). Malate dehydrogenase: evidence for tetrameric structure in *Mus musculus. Science* **160**, 1356–1357.

SHREFFLER, D. C. (1969). Genetic control of serum transferrin types in mice. *Proceedings of the National Academy of Science, U.S.A.* **46**, 1376–1384.

WEGMANN, T. C. & GILMAN, J. G. (1970). Chimerism for three genetic systems in tetraparental mice. *Developmental Biology* **21**, 281–291.