Genet. Res., Camb. (1973), **21**, *pp.* 121–132
With 5 text-figures
Printed in Great Britain

A multivariate analysis of subline divergence in the shape of the mandible in C57BL/Gr mice

By MICHAEL FESTING

Medical Research Council Laboratory Animals Centre, Woodmansterne Road, Carshalton, Surrey, U.K.

(*Received* 3 *July* 1972)

SUMMARY

The shape of the mandible in nine sublines of C57BL/Gr, seven other strains of 'C57 ancestry' and four unrelated strains was studied by multivariate techniques. The generalized distance function was used to classify individuals in the groups which they most closely resembled. The degree of misclassification depended on the pedigree relationship between strains and sublines. The generalized distance between pairs of subline centeroids was also highly correlated ($r = 0{\cdot}60$) with the number of generations between them. A canonical variate analysis was used to reduce the dimensionality so that a graphical display of the relationships between strains and sublines could be made. The results agreed closely with the classification analysis. It was concluded that the shape of the mandible could be used for subline identification though the accuracy of this technique depends on how closely the sublines are related.

1. INTRODUCTION

The shape of the mandible is a highly heritable trait, and Festing (1972) has shown that it can be used as a means of identifying different inbred strains of mice. A knowledge of the rate of subline divergence for mandible shape is therefore of practical importance as well as providing additional information on the rate of subline divergence of metrical characters. Subline divergence in mice has been extensively studied for quasi-continuous or threshold variables (see Grewal, 1962; Yong, 1972), but less widely studied for metrical characters, though Bailey (1959) examined subline differentiation of a number of skeletal measurements in BALB/c and C57BL/6 mice. Recently Grüneberg (1970) has reviewed the data on rates of subline divergence in inbred strains of mice, and has questioned whether the relatively high rates can be accounted for by mutation alone. Taylor (1972) studied the relationship between different inbred mouse strains, including different sublines of the same strains, using 16 identifiable genetic loci. He was able to show that strains with a known common ancestry had many genes in common. However, he had to discard some sublines from the study because they were similar at all 16 loci. The purpose of this investigation is to study the relationship between different strains and sublines as evaluated from the shape of the mandible.

2. MATERIALS AND METHODS

The history of the sublines of C57BL/Gr used in these studies has been given by Yong (1972) and Deol *et al.* (1957), so only brief details are given here. According to Staats (1966) all of the 'C57' strains of mice originated in the USA in about 1920. C57BL/Gr was imported into the UK in 1932, and has been maintained in the UK since that time. The relationships between the nine sublines studied are shown in Fig. 1 (from Yong, 1972). Although all mice were 60 days old at the time of the papain preparation of the skeletons, it should be noted that collection of the material covered a period of 20 years and there is no assurance that the environment remained constant during this period.

For comparative purposes data was also obtained on eleven additional strains maintained at the Medical Research Council Laboratory Animals Centre (i.e. in a different laboratory). The origin of these strains has been described by Parrott & Festing (1971) but they include another British subline of C57BL, an American branch of C57BL now known as C57BL/10ScSn, and four 'congenic resistant' lines (B10.LP-a, B10.D2, B10.A and B10.BR) based on this latter strain. Congenic resistant lines are developed by repeated backcrossing with forced incompatibility at one of the histocompatibility loci. Usually the equivalent of about 12 backcross generations are involved, with subsequent brother × sister mating. Strains C57BR and C57L were inbred independently from C57BL but came from the same original stock. Finally, four unrelated strains were included (A, A2G, C3H-*mg* and NZB). The ages of the mice in this latter group of strains varied from 50 to 100 days within each group. Unpublished studies show that there is little growth in the size or change in shape of the mandible after 50 days of age. Only male mice were studied.

A total of 13 measurements were taken on each right mandible as described by

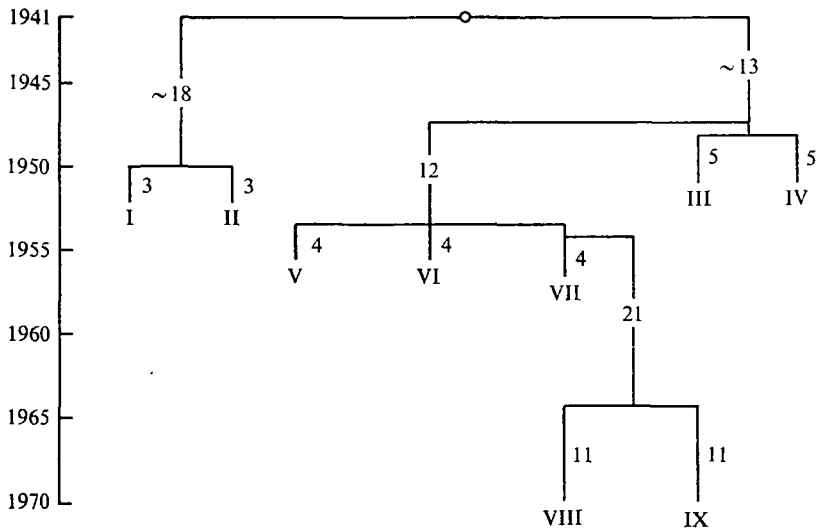

Fig. 1. Sublines of C57BL/Gr used in the study.

https://doi.org/10.1017/S0016672300013306 Published online by Cambridge University Press

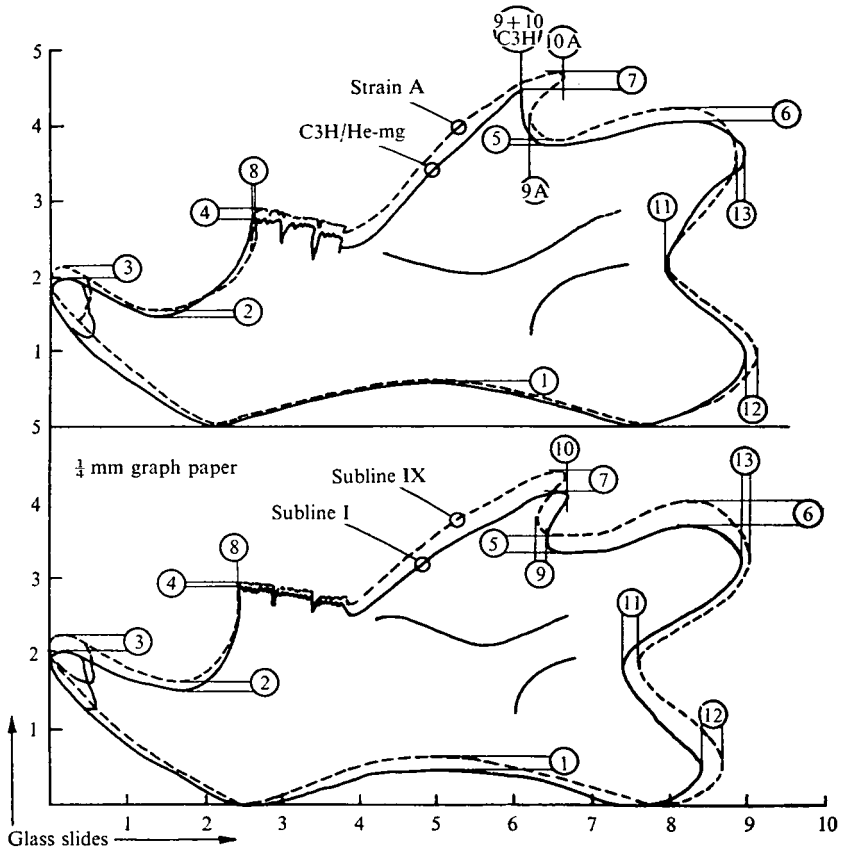

Fig. 2. Measurements used. Note that measurements nos. 1–7 are vertical and 8–13 are horizontal distances, respectively. The figure also shows the mandible shape in strains C3H, A and C57BL/Gr, sublines I and IX.

Festing (1972). These measurements are shown in Fig. 2 (for four different strains). The method was to place the mandible over a sheet of 0·25 mm graph paper touching two glass slides in a standard way. The measurements shown in Fig. 2 were then read off directly to the nearest 0·125 mm under a 10X monocular microscope. It should be noted that there is an unknown bias associated with these measurements due to inaccuracies in fixing the glass slides to the graph paper. Such a bias will not influence strain comparisons since all measurements were done on the same apparatus, but the strain means are given in arbitrary units rather than in units of 0·125 mm because of this bias.

3. STATISTICAL ANALYSIS

The volume of the data and the presence of significant correlations between the variables make a conventional univariate analysis impractical. Accordingly, the vector of 13 measurements for each individual was considered to be a single multivariate character which defines the 'shape' of the mandible (where 'shape'

https://doi.org/10.1017/S0016672300013306 Published online by Cambridge University Press

also includes size), and a multivariate analysis was carried out in order to discover how closely the shapes of the mandibles resembled each other in the sublines, the related strains and the unrelated strains. Three techniques were used. First, a classification analysis was carried out as described by Cooley & Lohnes (1971, chapter 10). Secondly, the 'generalized distances' were calculated between each pair of the sublines in order to gain some idea of the rate of subline divergence. Thirdly, a canonical variate or discriminant function analysis was carried out in order to reduce the dimensionality of the problem so that the results could be displayed graphically. This latter technique is described by Cooley & Lohnes (1971, chapter 9) and Blackith & Reyment (1971, chapter 8), who give computer programs for these analyses, though a different computer program (BMD07M) was used in the present studies. The technique has also been used by Delany & Whittaker (1969) to study natural populations of *Apodemus sylvaticus*. As the techniques are described fully in the references given above, only a brief description will be given here.

1. The classification analysis.

It is assumed that the vector of observations of the ith individual of the jth strain X_{ji} follows a multivariate normal distribution. In this case each individual can, conceptually, be located in n-dimensional space by treating the measurements as co-ordinates of a single point in space (n being the number of measurements per individual). The multivariate normal swarm of points representing a single strain will therefore fall within a hyper-ellipsoid in n-dimensional space which may be elongated in some directions due to correlations between traits. The problem is to find out what proportion of individuals lie at various distances from the group centroid so that the probability that an individual belongs to a given swarm can be calculated.

The generalized distance (also known as the Mahalanobis distance) is defined as:

$$d^2 = (X_{ji} - m_j)' D^{-1} (X_{ji} - m_j)$$

where m_j is the vector of means of the jth population and D is the pooled within-group variance-covariance matrix. The generalized distance is distributed approximately as a chi-squared variable with n degrees of freedom, and can therefore be used to assess the probability of the hypothesis H_j given observation vector X_i (i.e. $\Pr(H_j/X_i)$, $i = 1, 2, \ldots N$, $j = 1, 2, \ldots g$) that an individual belongs to the jth group (assuming g groups and equal *a priori* probabilities of group membership).

2. The generalized distances between each pair of sublines of C57BL/Gr were calculated in order to study the rate at which sublines diverge. The distance d^2 between each pair of sublines was therefore graphed against the known pedigree relationship between the sublines.

3. The aim of a multiple discriminant function or canonical variate analysis was to reduce the dimensionality of the problem so that the results could be shown graphically. The statistical methods are described elsewhere, but the principle is to extract a linear function of the set of observations which gives 'best' discrimination between strains. 'Best' in this case means the function which will maximise

the ratio of the among-groups sum of squares to the within-groups sum of squares, so that among-group differences will be large relative to within-group scatter. This is the first canonical variate, and it usually accounts for a high proportion of the total variation between strains. A second canonical variate is then extracted in such a way as to be independent of the first, but giving next best discrimination between strains. Using these two functions, the strains can be located in a discriminant plane. Additional canonical variates can be calculated up to $n-1$, but the first two or three usually account for a very high proportion of the variation between strains.

4. RESULTS

Means for each of the 13 traits in the 20 different strains or sublines are given in Table 1. There were no significant differences between sublines of C57BL/Gr for traits Nos. 8, 9 and 10, but there were significant differences between all other traits.

Some of the differences between strains were so pronounced that they could be detected by the naked eye. For example, the coranoid process in C3H was very short so that there was virtually no difference between measurements Nos. 9 and 10. In strains of C57 ancestry, the condaloid process (measurement No. 13) was always considerably longer than the angle process (measurement No. 12), whereas in other strains the two processes were more equal in length. The overall length of the mandibles of NZB (say measurement No. 13) was greater than for any other strain, though the total height (measurement No. 7) was not much above average. Some idea of the type of variation observed is given in Fig. 2.

Mandibles of the C57BL/Gr sublines tended to be smaller than average, but it is not known whether this was due to genetic or environmental causes. The latter is a possibility since these sublines were maintained in a different laboratory from the rest of the strains.

The results of the classification analysis are given in Table 2. Correct classifications are given on the main diagonal. In general, the degree of misclassification between the sublines and strains corresponded very closely with the known relationships between groups. Thus there was considerable overlap between closely related sublines such as I to IV, V to VII and VIII to IX, but less overlap between more distantly related groups. For example, there was no overlap between sublines I to IV and VIII to IX, which were separated by 51 to 79 generations. Three individuals of the C57BL/Gr sublines were misclassified as belonging to one of the other groups of C57 ancestry, and one C57BR was misclassified as being of subline VIII. Overall, 63% of the C57BL/Gr individuals were correctly classified to the strain and subline.

There was less misclassification of the other strains of C57 ancestry, as might be expected. Overall 85% of individuals were correctly classified, but all of these misclassifications were amongst strains of C57 origin. Finally, only a single individual of the remaining four strains was misclassified, giving 99% correct classification. It must be concluded therefore, that the classification analysis gives

Table 1. *Strain means for each of the variables given in Fig. 2*

Strain	Variable number 1	2	3	4	5	6	7	8	9	10	11	12	13	*n*
C57BL/Gr I	4·6	14·9	20·4	28·8	33·3	37·0	41·6	24·1	64·1	66·9	74·3	84·4	89·4	16
C57BL/Gr II	4·7	15·3	21·1	28·8	33·8	37·4	42·7	23·8	63·8	67·9	75·2	85·2	90·1	15
C57BL/Gr III	4·5	15·2	20·4	29·0	33·0	37·0	41·8	23·7	62·8	66·7	74·5	83·0	89·1	15
C57BL/Gr IV	5·0	15·2	21·1	29·3	33·4	37·3	42·1	23·8	63·2	67·3	75·1	84·3	90·3	15
C57BL/Gr V	5·6	15·0	21·3	28·5	34·8	39·4	43·9	24·6	63·5	68·0	77·2	87·2	91·3	11
C57BL/Gr VI	5·6	15·1	20·9	28·8	35·1	39·2	44·1	24·6	63·5	67·8	77·2	87·4	90·8	30
C57BL/Gr VII	5·8	15·1	21·0	28·3	35·0	39·3	43·6	24·2	62·9	67·3	76·7	86·6	90·9	24
C57BL/Gr VIII	6·4	16·4	22·8	29·7	36·2	40·8	45·6	24·6	63·5	67·2	77·6	88·1	91·5	17
C57BL/Gr IX	6·4	16·1	22·5	29·3	35·5	40·2	44·8	24·2	62·9	66·5	76·4	86·8	90·5	31
B10.LP-a	6·3	17·0	22·9	30·3	36·6	41·0	46·5	26·3	67·4	72·0	80·0	91·1	94·3	8
B10.D2	7·5	17·6	23·4	32·0	39·9	45·0	49·6	25·0	67·8	71·9	81·5	92·8	97·6	8
B10.A	6·4	17·3	23·6	32·0	39·0	44·1	48·9	26·3	68·9	73·3	84·4	95·6	99·5	8
B10.BR	6·6	16·7	22·9	31·3	38·1	44·3	48·3	25·3	65·4	70·4	82·3	94·4	96·7	7
C57BL	6·6	16·8	23·1	31·4	37·6	42·1	48·3	26·3	66·0	71·0	80·6	91·4	95·6	8
C57BR	6·1	16·4	22·6	30·3	38·3	43·6	49·3	27·3	66·9	72·4	82·9	96·4	96·8	7
C57L	6·1	17·2	24·3	30·8	38·0	44·0	49·4	27·0	66·5	72·3	81·5	94·1	96·5	8
NZB	6·0	15·4	21·8	30·2	38·7	45·2	48·9	28·2	67·5	71·2	88·2	101·8	99·3	30
A2G	6·5	14·4	20·0	28·9	39·2	43·4	47·9	28·7	65·1	70·3	84·3	95·1	94·0	30
A	6·6	15·1	21·2	28·8	37·8	42·3	46·8	26·6	62·3	66·4	79·2	92·2	88·9	29
C3H-mg	5·7	14·3	19·9	27·4	37·1	40·8	44·6	26·2	61·4	61·8	79·8	91·3	89·9	30
Weighted mean	5·9	15·5	21·4	29·2	36·4	41·0	45·6	25·6	64·1	68·0	79·4	90·4	92·3	347
S.D.*	0·48	0·39	0·52	0·66	0·87	0·95	0·95	0·63	1·22	1·08	1·27	1·35	1·37	—

Units are 0·125 mm + unknown bias which is constant across all strains.

* Standard deviation, pooled within groups.

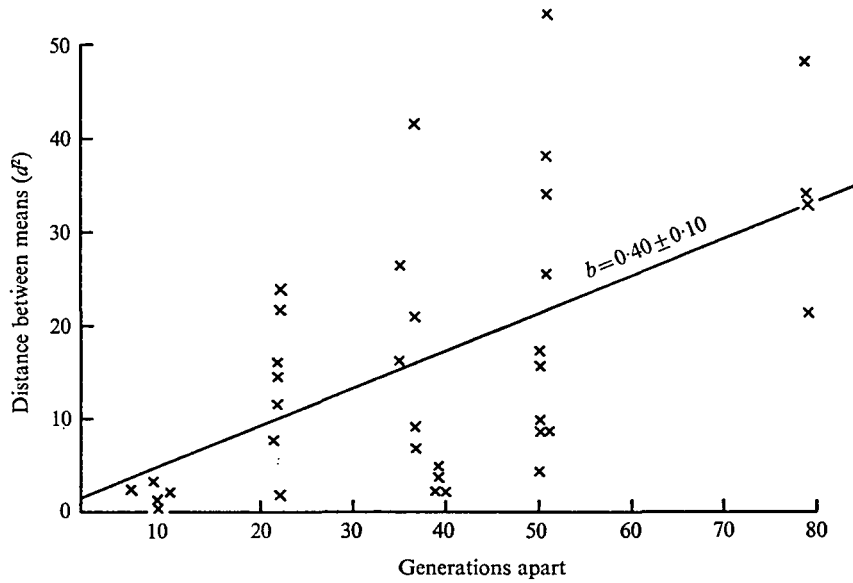

Fig. 3. Generalized distance and pedigree relationship between the sublines.

results which are in close agreement with the known genetic relationships between strains and sublines.

The rate of subline divergence in the nine sublines of C57BL/Gr is shown graphically in Fig. 3 which gives the generalized distance between pairs of strains as a function of the number of generations apart. The correlation between the two was 0·60 (95% confidence interval 0·31–0·74). A linear regression line has been fitted to the data, though as the rate of divergence presumably depends on the chance occurrence and fixation of new mutations, a close fit to the regression line could not be expected. The results do suggest, however, that subline divergence occurs as a result of the fixation of many mutations each having a minor effect.

The effects of environment should also be considered. Yong (1972) concluded that environmental factors had had a relatively small influence on the subline data, and these results give some support to this conclusion. Although the subline data was collected over a 20-year period, the overall sizes of the mandibles did not change, yet the similarities in shape were apparently dependent on the pedigree relationships. It would be difficult to think of any environmental factor which could produce changes in shape without associated changes in size, though the converse would not necessarily be true.

The results of the canonical variate analysis are shown graphically in Figs. 4 and 5, and the coefficients of these canonical variates are given in Table 3. No obvious interpretation can be placed on these figures. There is no evidence, for example, that there is a general size factor which is useful in discriminating between strains. Fig. 4 shows that the canonical variate analysis gives good discrimination between strains of C57 origin and the other four strains. (Lines have been drawn round the groups in Figs. 4 and 5 to clarify the relationships, but these

Table 2. *Results of the classification analysis*

Number of cases classified in each strain (numbers correspond to those in 1st column)

Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
1 C57BL/Gr I	9	3	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16
2 C57BL/Gr II	2	11	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	15
3 C57BL/Gr III	.	1	13	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	15
4 C57BL/Gr IV	2	2	.	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	15
5 C57BL/Gr V	.	.	.	.	5	3	2	1	.	.	.	.	.	.	.	.	.	.	.	.	11
6 C57BL/Gr VI	.	1	.	2	4	18	4	1	.	.	.	.	.	.	.	.	.	.	.	.	30
7 C57BL/Gr VII	.	1	.	.	7	3	12	.	1	.	.	.	.	.	.	.	.	.	.	.	24
8 C57BL/Gr VIII	.	.	.	.	.	.	.	13	3	1	.	.	.	.	.	.	.	.	.	.	17
9 C57BL/Gr IX	.	.	.	.	.	.	1	10	18	1	.	.	.	1	.	.	.	.	.	.	31
10 B10.LP-a	.	.	.	.	.	.	.	.	.	7	.	.	.	1	.	.	.	.	.	.	8
11 B10.D2	.	.	.	.	.	.	.	.	.	.	8	.	.	.	.	.	.	.	.	.	8
12 B10.A	.	.	.	.	.	.	.	.	.	.	.	7	1	.	.	.	.	.	.	.	8
13 B10.BR	.	.	.	.	.	.	.	.	.	.	.	.	7	.	.	.	.	.	.	.	7
14 C57BL	.	.	.	.	.	.	.	.	.	2	.	.	.	6	.	.	.	.	.	.	8
15 C57BR	.	.	.	.	.	.	.	1	.	.	.	.	.	.	6	.	.	.	.	.	7
16 C57BL	.	.	.	.	.	.	.	.	.	1	.	.	.	.	2	5	.	.	.	.	8
17 NZB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.	.	30
18 A2G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	29	1	.	30
19 A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	29	.	39
20 C3H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	30	30

All individuals off the leading diagonal are misclassified. 347

Table 3. *Coefficients for the canonical variables*

Variable No.	Coefficients 1st	2nd	3rd
1	−0·39	0·79	0·78
2	0·67	0·70	−1·18
3	−0·93	1·51	1·09
4	1·03	−0·44	−0·08
5	0·13	−0·68	−0·22
6	−0·37	1·04	−0·13
7	−0·49	0·16	0·01
8	−0·49	−0·23	0·26
9	−0·15	0·09	−0·29
10	0·48	−0·03	−0·63
11	0·01	−0·61	0·44
12	−0·91	−0·06	−0·16
13	0·65	0·04	−0·27
Proportion of total dispersion	0·67	0·19	0·07

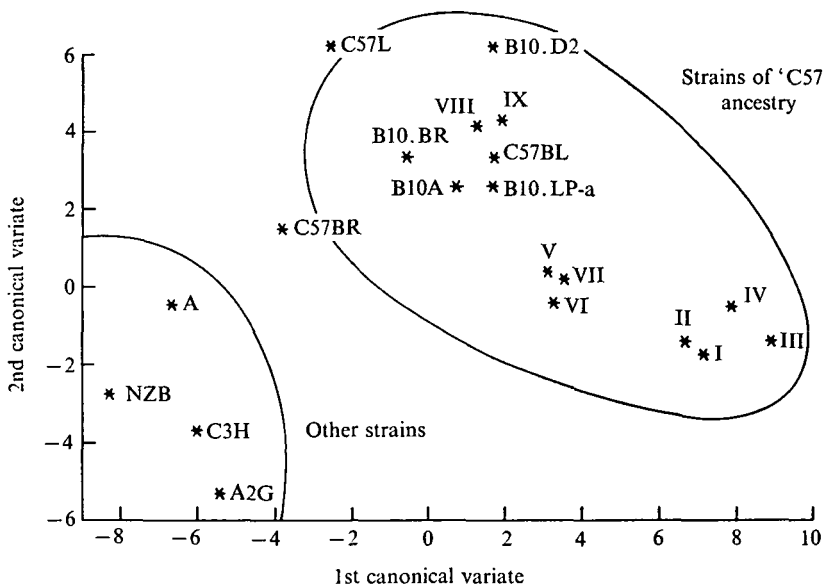

Fig. 4. First and second canonical variates.

lines have no statistical meaning. The canonical variates have been standardized to give unit variances.) Moreover, the sublines fall into three distinct groups, as in the classification analysis, covering approximately seven standard deviation units of the first canonical variate.

The third canonical variate graphed against the first gives good separation of the C57BL/Gr sublines from the other strains of C57 ancestry. Thus, in general, the canonical variate analysis gives graphical results which are in good agreement with the classification analysis. The results of the canonical variate analysis are

https://doi.org/10.1017/S0016672300013306 Published online by Cambridge University Press

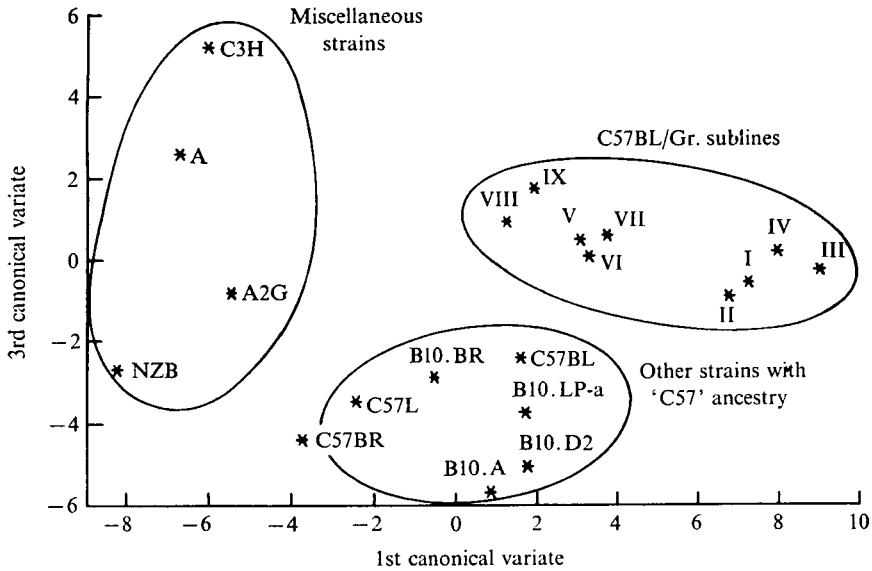

Fig. 5. First and third canonical variates.

also remarkably similar to those of Taylor (1972), whose study was based on 16 identifiable genetic loci. In both these studies strains of C57 ancestry were shown to have many features in common. It is reasonable to conclude that a multivariate analysis of thirteen measurements defining the shape of the mandible may be used to identify different sublines of the same strain, though the accuracy will depend on how closely the sublines are related.

5. DISCUSSION

Although the subline data were collected over a long period, and the eleven other strains were reared in a different laboratory from the sublines, there is some indirect evidence that environmental factors have had relatively little influence in the interpretation of these results. The means of traits 8, 9 and 10 did not differ significantly between sublines, and it is unlikely that environmental factors could cause changes in shape without influencing overall size. Moreover, there was a linear relationship between the generalized distance between pairs of strains and the pedigree relationship between them. If environmental factors were of any great importance it is unlikely that they would have a linear trend. The small size of the subline mandibles compared with those of other strains of C57 ancestry may have been caused partly by environmental influences. However, the analyses tended to emphasize the similarities of all the strains of C57 origin in spite of these size differences, so the interpretation would not have been very different if account had to be taken of environmental influences.

The purposes of this investigation were to study the rate of subline divergence for a metric character and to discover the accuracy of predicting which strain an individual belongs to from the shape of the mandible when closely related sublines

are involved. Previous studies of subline divergence have tended to emphasize that inbred strains do not remain genetically stable over long periods when propagated by continued brother × sister mating. These results confirm this finding in that after only 51 generations there was no overlap at all between over 100 individuals of sublines I–IV on the one hand and VIII to IX on the other. However, even after more than 40 years separation (probably well over 100 generations) different sublines of C57 origin are still quite similar to one another, at least in the shape of the mandible, and they differ markedly from other unrelated strains. Moreover, changes tend to occur in small discrete steps, giving an overall linear trend, in separation between lines as a function of the number of generations apart though chance obviously plays an important part in controlling the rate of divergence. It is probably worth noting that if long-term stability rather than complete homozygosity is the objective in maintaining inbred strains, then a random mating system based on a large effective population size rather than continued brother × sister mating should be used.

The predictability of group membership, which is of practical importance if mandible shape is to be used as a method of strain identification as suggested by Festing (1972), depended on how closely related the strains and sublines were to each other. Unrelated strains were sufficiently dissimilar for there to be little chance of incorrect identification of individuals. Sublines separated for over 50–60 generations were also sufficiently different for there to be less than about 5% overlap, though in large groups of related sublines and congenic lines based on the same inbred strain and reared in the same environment, the misclassification rate of individuals appeared to be about 15%. Often very high levels of accuracy at the individual level are not necessary since in most practical applications there is usually more than one individual per group that can be studied. In any case, it is improbable that any other technique described so far would be more accurate. Biochemical methods, for example, would probably fail to show up any differences between such closely related sublines. Moreover, there is no reason why the accuracy of this method should not be increased by taking into account additional variables if such a procedure was thought to be worthwhile. The same apparatus can be used to measure more bones, and any other variables that were thought to be important could be included. This approach might prove more successful than Taylor's (1972) method of detecting relationships between inbred strains since morphological characters are usually relatively easy to measure. The canonical variate analysis also gives a clear graphical description of the relationship between different strains and substrains.

This paper is dedicated to Professor H. Grüneberg, F.R.S., on his retirement from the Chair of Animal Genetics at University College London. I should also like to thank Professor Grüneberg for giving me access to the subline skeletons used in this study.

REFERENCES

BAILEY, D. W. (1959). Rates of subline divergence in highly inbred strains of mice. *Journal of Heredity* **50**, 26–30.

BLACKITH, R. E. & REYMENT, R. A. (1971). *Multivariate Morphometrics*. Academic Press, London and New York.

COOLEY, W. W. & LOHNES P. R. (1971). Multivariate data analysis. John Wiley and Sons Inc. New York and London.

DELANY, M. J. & WHITTAKER, H. M. (1969). Variation in the skull of the long-tailed field-mouse. *Apodemus sylvaticus* in mainland Britain. *Journal of Zoology* **157**, 147–157.

DEOL, M. S., GRÜNEBERG, H., SEARLE, A. G. & TRUSLOVE, G. M. (1957). Genetical differentiation involving morphological characters in an inbred strain of mice. I. A British branch of the C57BL strain. *Journal of Morphology* **100**, 345–376.

FESTING, M. (1972). Mouse strain identification. *Nature* **238**, 351–352.

GREWAL, M. S. (1962). The rate of genetic divergence of sublines in the C57BL strain of mice. *Genetical Research* **3**, 226–237.

GRÜNEBERG, H. (1970). Is there a viral component in the genetic background? *Nature* **225**, 39–41.

PARROTT, R. F. & FESTING, M. (1971). *Standardised Laboratory Animals, Manual* No. 2, MRC Laboratory Animals Centre, Carshalton, Surrey.

STAATS, J. (1966). The Laboratory Mouse. In E. L. Green (ed) *The Biology of the Laboratory Mouse*. McGraw-Hill, New York and London.

TAYLOR, B. A. (1972). Genetic relationships between inbred strains of mice. *Journal of Heredity* **63**, 83–86.

YONG, HOI-SEN (1972). Is subline differentiation a continuing process in inbred strains of mice? *Genetical Research* **19**, 53–59.