Genet. Res., Camb. (1984), **44**, *pp.* 271–281
With 3 text-figures
Printed in Great Britain

Genetic relationships between inbred strains of rats. An analysis based on genetic markers at 28 biochemical loci

BY MICHAEL F. W. FESTING[1] AND KLAUS BENDER[2]

[1] *Medical Research Council, Experimental Embryology and Teratology Unit, Woodmansterne Road, Carshalton, Surrey*
[2] *Institut für Humangenetik und Anthropologie, Albert-Ludwigs-Universität, 7800 Freiburg*

(*Received 18 July 1984 and in revised form 8 August 1984*)

SUMMARY

Genetic similarities among 46 strains of rats based on published data involving 93 samples and 28 biochemical loci were assessed using principal coordinate and cluster analysis techniques. Seventeen strains were represented by more than one colony. In ten of these, nominally identical strains differed, and in four cases this was attributed to genetic contamination. A total of 52 genetically different strains were eventually identified. Strains BN and DA were most dissimilar, while strain BP was the most unusual strain over-all. The principal coordinate and cluster analysis showed three main clusters, which could be explained on the basis of linkage disequilibrium for some of the esterase loci in linkage group 5. Among six of these loci only 12 haplotypes were observed, with 24/52 strains having a single haplotype. Re-analysis of loci in linkage equilibrium failed to reveal any important clusters.

INTRODUCTION

The historical relationship between inbred strains of laboratory rats is poorly documented (Festing, 1979) and of doubtful validity in view of recent studies showing that genetic contamination of inbred strains of laboratory rodents is distressingly common (Festing, 1982; Kahn *et al.* 1982). Moreover, until recently relatively few inbred rat strains had been typed at more than a few biochemical marker loci, so such loci could not be used effectively to monitor genetic integrity, or genetic similarities between strains. Knowledge of such relationships is important for many studies. In some investigations, it would be desirable to have some objective criterion for choosing strains which differ as much as possible. For example, if inbred rats were to become used in toxicological screening, as suggested by Festing (1982), then it would be desirable to choose disparate strains. Similarly, in studies of genetic linkage or in the development of sets of recombinant inbred strains (Taylor, 1981; Bailey, 1971), it is desirable to be able to make crosses which differ for as many markers as possible.

The genetic similarity between 27 inbred strains of mice was investigated by Taylor (1972) using 16 biochemical marker loci. Strains were 'mapped' into a

two-dimensional plane using mathematical techniques similar to Principal Coordinate analysis (Dunn & Everitt, 1982), and it was demonstrated that inbred mouse strains fell into three distinct clusters, one of which consisted largely of strains with a 'C57' ancestry.

The recent cooperative study of biochemical markers in 93 samples of inbred rat strains (Bender *et al.* 1984) now makes it possible to carry out a similar study in inbred strains of rats.

MATERIALS AND METHODS

A full list of alleles at 28 loci among 93 samples of inbred rat strains has been given by Bender *et al.* (1984), together with details of the biochemical techniques involved. The data include 46 named strains, 17 of which were represented by more than one sample, with a maximum of 9 samples of strain BN.

Alleles were coded for numerical analysis, and a matrix of similarities between all pairs of samples was calculated by giving a score of 1 if a pair of strains were identical at a given locus, and zero otherwise (Dunn & Everitt, 1982). The similarities within nominally identical strains were inspected, and any strains which were substantially different at several loci were separated out in the subsequent analysis.

Two techniques were used to analyse the similarity matrix. Multidimensional scaling using Principal Coordinate analysis was used to produce a 2-dimensional map showing the relationships between strains with the minimum degree of distortion. Full details of the techniques are given by Dunn & Everitt (1982), Blackith & Reyment (1971) and Maxwell (1977). The technique involves extraction of the latent roots and vectors of the similarity matrix in order to be able to present the data with reduced dimensionality. Hierarchical cluster analysis using an average-linkage criterion was also used in an attempt to show the similarities between strains. Dunn & Everitt (1982) discuss the relative merits of various hierarchical clustering methods such as single-linkage, complete linkage, average-linkage etc. and concluded that no single method was best in every situation. However, the 'mathematically respectable' single linkage method was usually least successful in practice, and the average clustering technique was generally fairly satisfactory over-all. The average linkage method was therefore chosen in this study.

RESULTS

1. *Variation within nominally identical strains*

Seventeen of the strains were represented by 2 or more independent samples. Table 1 lists these strains together with the numerical measure of mean similarity. In 7 strains there was no difference among nominally identical samples. The source of the variation among the other samples is also noted in the table. Within strains BN and COP, the only variation was at the Pep-3 locus. Similarly, samples of strains SHR, WF and WKY only differed at a single locus. Such differences are most likely to have arisen as a result of residual heterozygosity or mutation (Bailey, 1977). In contrast, in strain LEW, the Kyo subline differed from the others at 4 loci, in LOU the Iap subline differed from the other two at 8 loci and in WAG the

Cpb subline differed from the other 4 sublines at 8 loci. Such differences are most likely to have occurred as a result of genetic contamination of one of the sublines. In the case of SD, the four sublines differed at a total of 7 loci, with no two sublines being genetically identical. This variation may have arisen because each line was independently inbred from the same Sprague–Dawley outbred colony, with unimaginative choice of strain designation. Finally, the Orl subline of strain LE differed from the other two at two loci. It is not immediately obvious how this occurred.

Table 1. *Variation within nominally identical strains*

Strain	No. of samples	Mean % similarity	Comment
ACI	5	100	—
AO	2	100	—
AUG	2	100	—
BN	9	98	Variation only at Pep-3, 6 strains a, 3 strains b
COP	2	96	Variation only at Pep-3
DA	2	100	—
F344	4	100	—
LE	3	94	Orl subline differs from rest at Akp-1 and Pgd
LEW	8	93	Kyo subline differs from rest at Es-1, Es-2, Es-3, Es-9 & Es-10
LOU	3	78	BW and Max sublines identical, differ from Iap at 8 loci
PVG	2	100	—
R	4	100	—
SD	4	84	Sublines differ at 7 loci, no two identical
SHR	4	99	Kyo & Cpb sublines differ at Ahd-2 (others untyped)
WAG	5	85	Cpb subline differs from rest at 8 loci
WF	2	96	Differ at Es-6
WKY	2	96	Differ at Pgd

As a result of this preliminary analysis, LEW/Kyo, LOU/Iap, WAG/Cpb and the four samples of SD rats were regarded as independent inbred strains in all subsequent analyses, giving a total of 52 different strains. The remaining within-strain variation was eliminated by arbitrarily taking the first strain listed by Bender *et al.* (1984) as the 'type strain' in all subsequent analyses.

2. *Similarities between strains*

An abbreviated similarity matrix between all 52 strains is given in Table 2. In this table, the similarity is shown by a single number. Strains have been ordered so as to bring similar strains together as far as this is possible. The lowest level of similarity in this table is 3, representing a similarity of 30–40 %. Inspection of the full matrix (not shown) shows that the greatest dissimilarity between any two strains is that between BN and DA, at a level of 31 % similarity across all loci. However, the most 'unusual' strain, i.e. the strain which on average differs most from all other strains is BP, as is clearly shown in the cluster analysis, below.

Table 2. *Abbreviated similarity matrix between all 52 strains*

(Each number represents the percent similarity to the nearest 10 units, i.e. '7' indicates 70–80% similarity.)

ACI	–
BDV	7–
BDIX	79–
R/A	999–
LOU/Iap	8768–
AVN	78797–
HMT	788879–
CPBB	7778798–
F344	68886999–
LOU/Bw	688869999–
PVG	7778798888–
DHK	76676889897–
WE	777869888888–
WEK	8778687777779–
W/Kyo	66676878888777–
COP	788767778867766–
DA	7777666667665577–
AUG	88887787777667778–
BUF	788867878887777779–
LEW/Kyo	8777778788866667798–
BH	77886787888778666888–
BS	788868878877896668879–
WAG/Mb1	7999699899878877688889–
SD/Cpb	88897897888678767998999–
SPRD	8889789788878767988989–
LE	6776567666656766688787778–
TM	67776777887677766777888887–
LEP	677665667665556677777667777–
DONRYU	4665567677566566566556665455–
SHR	57776887777776766776677776568–
SHRSP	577667766766656676655676655678–
IS	6567767766866665577776677656686–
BP	555565544454445555544445544455554–
AO	666757666666665446656677656556455–
PD	566667676666765535555766563565559–
SD/Rij	5666677666656565466666775755655589–
SD/A	666766555565565456556667756545556888–
SD/Waa	6667676665656664466666777575454559999–
BDII	66666756565655564545556654535554689888–
U	666656565556656456545555554546556877777–
WAG/Cpb	655657676667765535555566555446459877779–
AS	566545566645456666666666656565545586767677–
AS2	467645666555556566566766666656554878787778–
BDE	455545556655555544445655545454534677676666–
WKY	4655565666555455544445655355556457777677777778–
LEW/Lac	554555655564444465655566454454467787776776777–
MAXX	46655666666565553566666665545655588877777877889–
WKS	455545656655546555555556656555555777877667777887–
E3	46555555554455343444565554544553467777665777677–
OM	455555556544554545555666557543433777676567866666 7–
BN	44454665556555433444455553435544577656667556677554–
WF	455546656656654534545565544455534677667675666676757–

https://doi.org/10.1017/S0016672300026513 Published online by Cambridge University Press

3. *Hierarchical cluster analysis*

The results of the hierarchical cluster analysis using the average-linkage criterion are displayed in Fig. 1. It will be noted that at the 50–60 % similarity levels there are three main clusters with 32 strains ACI-IS/Kyo in the first cluster, strain BP in the second and the 19 strains AO-WF in the third, though strains WF and BN are rather dissimilar from the rest.

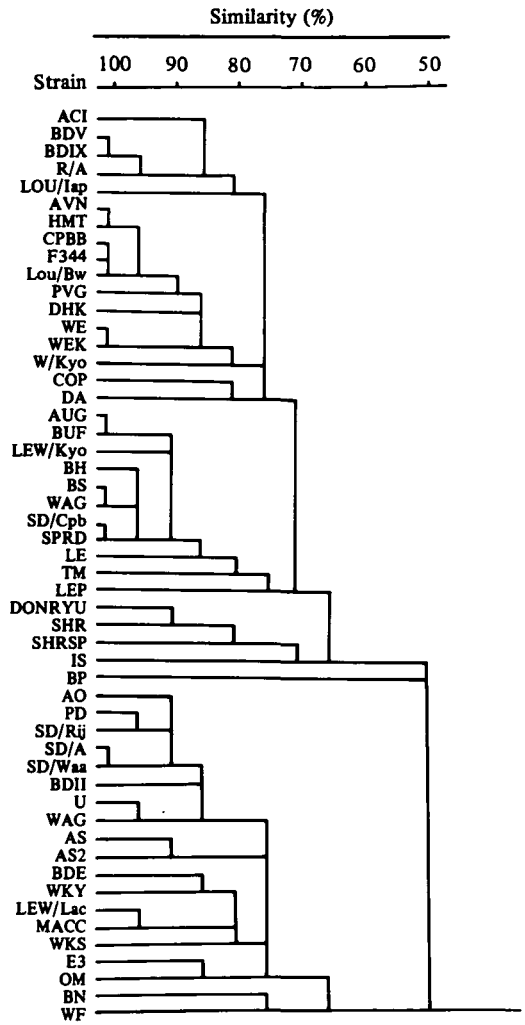

Fig. 1. Average linkage dendrogram of relationship between rat strains based on alleles at 28 biochemical loci.

4. *Principal coordinate analysis*

The first four principal coordinates of the full data matrix only accounted for 27, 10, 9 and 8 % of the variance, respectively. In such circumstances, representation of the data in two-dimensions inevitably leads to considerable distortions so that

strains which map close to each other may not actually be very closely related, though closely related strains will map close to each other. A plot of the first versus the second principal coordinate scores is presented in Fig. 2. The results are striking, with the strains falling into the three main clusters shown also in Fig. 1, i.e. the 32 strains ACI to IS/Kyo falling in one cluster, 19 strains AO-WF in the second, and strain BP in a 'cluster' of its own.

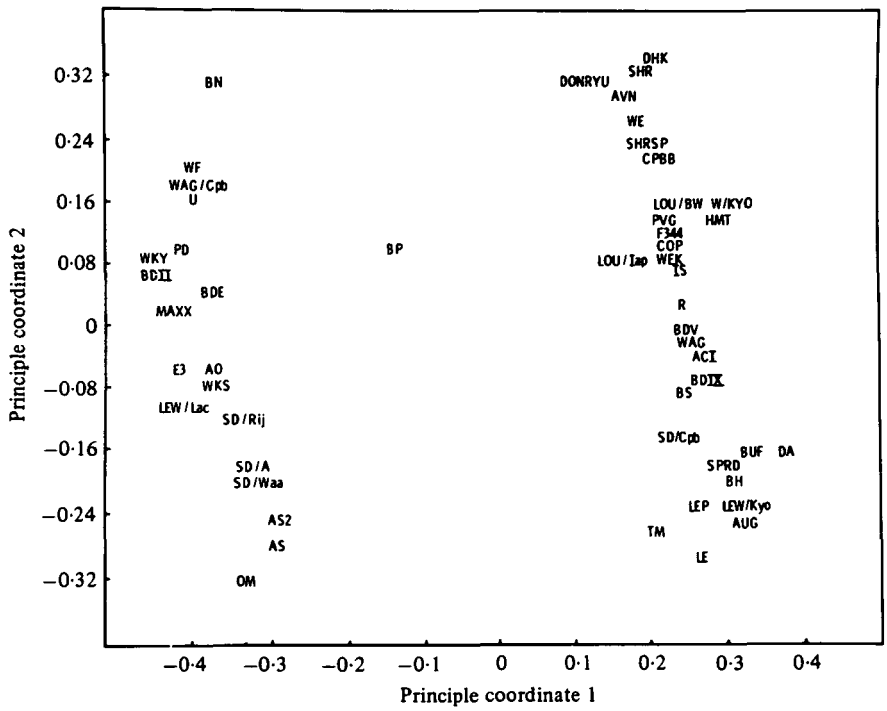

Fig. 2. Plot of the first two principal coordinates calculated from data on 28 biochemical loci in 52 strains of rat.

The reason for such a strong clustering of the strains is clarified in Table 3. A group of esterases loci are known to be closely linked in group V in the rat. Linkage data published by Yamada, Nikaido & Matsumoto (1980), Matsumoto (1980), Moutier, Toyama & Charrier (1973), and also Bender (unpublished) suggests that the gene order is Es-3, 9, 7, 10 (and 8), 2 and 4. Among the 52 strains there are only 12 different haplotypes for this chromosome segment. No fewer than 24 of the strains have the first haplotype, and the clustering is largely picking out strains that are Es-9a, Es-7b, Es-10a Es-2a and Es-4b or a as the first cluster, and those which are Es-9c, Es-7b, Es-10b, Es-2d or c and Es-4b as the third. It is clear from these data that among inbred rat strains there is strong linkage disequilibrium involving these and other esterase loci. This may reflect a common ancestry of some of the strains, or a functional role for some esterase haplotypes. In view of these results, a systematic study of linkage disequilibrium involving all 28 loci was carried out. The results are summarised in Table 4. The only significant disequilibria

Table 3. *Esterase haplotypes among 52 rat strains*

	Locus (listed in presumed gene order)						
Haplotype	Es-3	Es-9	Es-7	Es-10	Es-2	Es-4	Cluster
1	a	a	b	a	a	b	1
2	b	a	b	a	a	b	1
3	c	a	b	a	a	b	1
4	d	a	b	a	a	b	1
5	b	a	b	a	a	a	1
6	b	a	b	a	c	a	1
7	a	b	a	a	a	b	1
8	c	c	b	b	d	b	3
9	d	c	b	b	d	b	3
10	d	c	b	b	c	b	3
11	c	c	b	b	c	b	3
12	d	d	b	c	e	b	2

Strains with each haplotype

(1) ACI, AUG, AVN, BDV, BDIX, BH, BS, BUF, CPBB, COP, DA, DHK, F344, HMT, LEW/Kyo, LOU/Bw, PVG, R, SD/Cpb, SPRD, WAG, W, WE, WEK
(2) SHRSP, LOU/Iap
(3) TM
(4) LEP
(5) IS, SHR
(6) DONRYU
(7) LE
(8) AS2, BDII, E3, OM, PD, SD/A, SD/Rij, SD/Waa
(9) AO, AS, LEW, MAXX, U, WAG/Cpb, WKS, WKY
(10) BN, BDE
(11) WF
(12) BP

Table 4. *Loci in which there are statistically significant associations among the 52 strains studied*

	Chi-squared value, (degrees of freedom) for associated locus				
Locus	Es-3	Es-8	Es-9	Es-10	Es-Si
Es-1	—	—	—	—	12 (1)
Es-2	44 (9)	95 (6)	96 (9)	94 (6)	—
Es-3	—	44 (6)	46 (9)	44 (6)	—

involved the loci noted above and a slight disequilibrium between Es-1 and Hbb which was just statistically significant at the 5 % level of probability. As these two loci are unlinked, this is attributed to type 1 error (i.e. a false positive result) which is not unexpected in view of the number of comparisons (252) that were made.

5. *Reanalysis of independent loci*

In view of the above linkage disequilibrium, the esterase loci are not providing independent assessments of the similarities between strains. Accordingly, the data were re-analysed, but without the data on Es-Si, Es-3, 8, 9 and 10. A revised similarity matrix and average-linkage dendrogram are presented in Table 5 and

Table 5. *Abbreviated similarity matrix between rat strains after elimination of data on Es-3, 8, 9, 10 and Si*

(The number indicates similarity to the nearest 10 units, i.e. 9 indicates 90–100% similarity.)

ACI	–
BDII	7–
LOU/Iap	88–
BP	777–
COP	7766–
DA	65677–
LEP	546467–
AO	7866555–
PD	68766449–
AVN	787665589–
R/A	8887666888–
SD/A	88775668879–
SD/Waa	777655698889–
BS	6665656878877–
WAG/Mbi	77656668888789–
HMT	677666688987889–
SD/Rij	6776656888788789–
SD/Cpb	77766678888898999–
SPRD	777666787888889989–
BDV	7777876778877788788–
BDIX	77667777778878887889–
AS	675577776666677667777–
AS2	5755767877777887788889–
AUG	76777777677777887888887–
BUF	756667776777788878878779–
LEW/Kyo	76756777577677777886768 8–
BH	766556776777799887897777888–
LE	655556765666677767877778879–
TM	65656667777778888789776777788–
LEW/Lac	67775547776776788876576778756–
MAXX	576564588877778888777776777668–
PVG	7776555788878778888765677876788–
CPBB	677575688876777877777677767678–
F344	67657677887678898888778777687879–
LOU/Bw	676576678876788988887877776878799–
DHK	666575577765666866665666566567888–
U/A	77667557777765666657766655555767667–
WAG/Cpb	776564588877767877666666665578787788–
WE	77656447887678787777656665767577777778–
WEK	766564587887788767777566667675677777679–
W	66666666787676677776667676667567777765666–
WKS	56566666766676688776667666657776678656557–
IS	6575446766767667676565577676556876666666666–
DONRYU	3546655666555667665666766455556567765555565–
SHR	565665577875677877777677856657777777767678–
SHRSP	566667656765566766677566655555655675655565568–
BDE	575565566666576666677675556576766775556656564655–
E3	576554567666676776676565546566755675666656455 67–
WKY	57667667777667777677675555566767886666567565778–
OM	566465576666677667656776666576656775556656454567 6–
BN	46564346776565676655545455544777666656655556656554–
WF	565564566765667877666565656556766777677656566566657–

https://doi.org/10.1017/S0016672300026513 Published online by Cambridge University Press

Fig. 3, respectively. The revised principal coordinate analysis failed to reveal any important clustering when these loci were omitted, and the first two coordinates only represented 25 % of the total variation so that two-dimensional representation would lead to considerable distortion of the similarities between strains. The revised principal coordinate analysis is not shown.

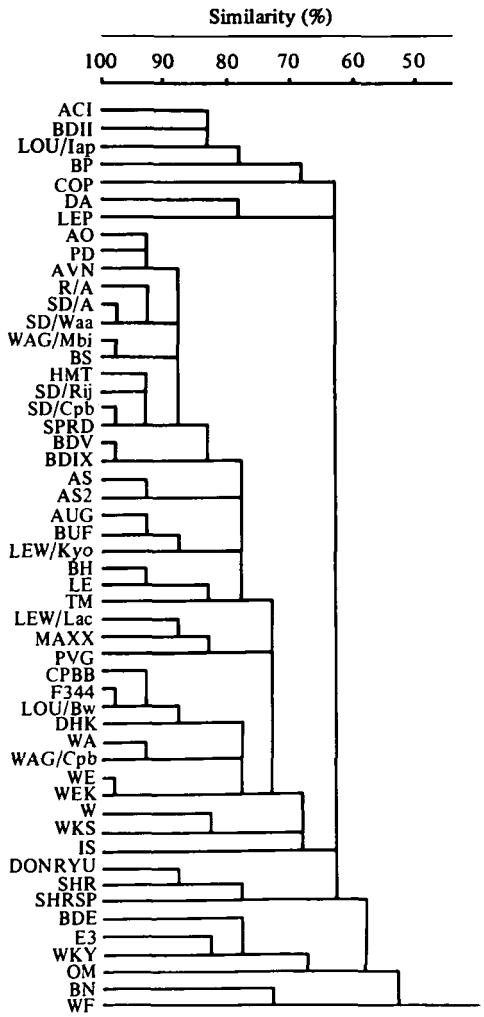

Fig. 3. Average linkage dendrogram of relationship between rat strains based on alleles at 23 biochemical loci not in linkage disequilibrium.

DISCUSSION

The variation between different samples of nominally identical inbred rat strains has already been commented on by Bender *et al.* (1984). It could have arisen as a result of residual heterozygosity, mutation, and/or genetic contamination. As only 7/17 strains were entirely homogeneous, it is clear that much more attention

will need to be paid to genetic quality control in the future. Fortunately, these data provide a useful background for future genetic quality control using biochemical techniques.

The most striking finding of this analysis is that there is strong linkage disequilibrium for several of the esterase loci among inbred strains of rats, with only 12 haplotypes among the 52 strains, and no fewer than 24 strains having a single haplotype. Theoretically, among six loci with 4, 4, 2, 3, 4 and 2 alleles (Table 3) there could be 768 different haplotypes, though as some alleles are relatively uncommon not all combinations would have an equal probability of being observed. Such disequilibria for esterases have been reported on a number of occasions in studies of wild mice (Peters, 1982). It could arise in inbred rat strains either because each strain is not an independent sample from a homogeneous gene pool (i.e. the strains are related by descent), or because some esterase haplotypes survive better than others in laboratory conditions, or as a result of a combination of both of these. Unfortunately, the data on inbred strains is not sufficient to distinguish between these possible causes.

This linkage disequilibrium also presents a number of problems in assessing the similarities between the strains, as it is difficult to know what weight to attach to an observation that two strains have similar esterase haplotypes. If it did imply that two strains had similar ancestry, then this might be useful information. Unfortunately, this does not seem to be the case as if two strains are derived from the same outbred stock, which is itself not in linkage equilibrium, they may end up with entirely different esterase haplotypes. For example, strains WKY and SHR are reported to have been derived from the same outbred Wistar stock and WKY is used as the normotensive control for SHR (Festing, 1979), yet they fall in different clusters in Fig. 1 as a result of having different esterase haplotypes. A number of other examples could be quoted. On the other hand, if the esterase haplotypes have some functional significance, then a clustering based largely on these haplotypes may be more valid, as it implies biological similarity.

In view of the difficulty of interpretation of these results, the analysis was also carried out using a reduced data matrix in which data on some of the esterases had been discarded. Those wishing to use the dendrogram for assessing the similarities between strains may find it safer to use this second analysis.

REFERENCES

Bailey, D. W. (1971). Recombinant inbred strains, an aid to finding identity, linkage and function of histocompatibility and other genes. *Transplantation* **11**, 325–327.

Bailey, D. W. (1977). Genetic drift: the problem and its possible solution by frozen embryo storage. In The Freezing of Mammalian Embryos. (ed. K. Elliot and J. Whelan), pp. 291–303. Ciba Foundation symposium 52, New Series.

Bender, K., Adams, M., Baverstock, P. R., den Bieman, M., Bissbort, S., Brdicka, R., Butcher, G. W., Cramer, D. V., von Deimling, O., Festing, M. F. W., Gunther, E., Guttmann, R. D., Hedrich, H. J., Kendall, P. B., Kluge, R., Moutier, R., Simon, B., Womack, J. E., Yamada, J. & van Zutphen, B. (1984). Biochemical markers in inbred strains of the rat. *Immunogenetics* **19**, 257–266.

Blackith, R. E. & Reyment, R. A. (1971). *Multivariate Morphometrics*. London, New York: Academic Press.

DUNN, G. & EVERITT, B. S. (1982). An introduction to mathematical taxonomy. Cambridge, London: Cambridge University Press.

FESTING, M. F. W. (1979). *Inbred Strains in Biomedical Research.* London, Basingstoke: Macmillan Press.

FESTING, M. F. W. (1982). Inbred strains and the factorial experimental design in toxicological screening: In *Proceeding of the 7th ICLAS Symposium, Utrecht 1979.* Stuttgart, New York: Gustav Fischer.

FESTING, M. F. W. (1982). Genetic contamination of laboratory animal colonies: an increasingly serious problem. *ILAR News 25.* 6–10.

KAHN, B., AUERBACH, R., ALTER, B. J. & BACH, F. H. (1982). Histocompatibility and isoenzyme differences in commercially supplied 'BALB/c' mice. *Science* **217**, 379–381.

MATSUMOTO, K. (1980). New polymorphisms of esterase 7 and esterase 8 in inbred strains of rats: tissue expression and linkage studies. *Biochemical Genetics* **18**, 879–887.

MAXWELL, A. E. (1977). Multivariate analysis in behavioural research. London: Chapman Hall.

MOUTIER, R., TOYAMA, K. & CHARRIER, M. F. (1973). Biochemical polymorphism in the rat, Rattus norvegicus: genetic study of four markers. *Biochemical Genetics* **8**, 321–328.

PETERS, J. (1982). Nonspecific esterases of Mus musculus. *Biochemical Genetics* **20**, 585–606.

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

TAYLOR, B. A. (1981). Recombinant inbred strains. In Genetic Variants and Strains of the Laboratory Mouse. M. C. Green (ed.), pp. 397–407. Stuttgart, New York: Gustav Fisher Verlag.

YAMADA, J., NIKAIDO, H. & MATSUMOTO, S. (1980). Linkage analyses among five esterase loci in the laboratory rat (Rattus norvegicus). *Biochemical Genetics* **18**, 433–438.