

Genetic Relationships Between Inbred Strains of Mice

B. A. TAYLOR

STANDARD inbred strains of mice are widely used in research; yet the origins of many of these strains are obscure. Often if a quantitative measure of the degree of relationship between strains were available, more cogent interpretations of strain differences could be made. Genetic relationships can be computed when accurate pedigrees are available. In the absence of complete pedigree information, relationships must be estimated empirically. Frequently, attempts are made to discern strain relationships from the distribution of alleles at one or

two loci. Recently, sufficient information on the allelic distribution at many loci has accumulated to make a quantitative analysis possible.

Materials and Methods

The data used in this analysis are the strain distributions of the alleles of 27 strains at 16 polymorphic loci (Table I). Twelve of these loci determine electrophoretic protein variations. These are: esterase-1, *Es-1*; esterase-2, *Es-2*; esterase-3, *Es-3*; NADP malate dehydrogenase-1, *Mod-1*; isocitrate dehydrogenase-1, *Id-1*; autosomal glucose 6-phosphate dehydrogenase-1, *Gpd-1*; glucose phosphate isomerase-1, *Gpi-1*; phosphoglucomutase-1, *Pgm-1*; phosphoglucomutase-2, *Pgm-2*; lactate dehydrogenase regulator gene, *Ldr-1*; hemoglobin beta chain, *Hbb*; and dipeptidase-1, *Dip-1*. The strain distributions are those published by Roderick *et al.*³. Other loci in-

Dr. Taylor is an associate staff scientist at The Jackson Laboratory, Bar Harbor, Maine 04609. This research was supported by an allocation from NIH Research Grant RR 05545-09 from the Division of Research Resources to The Jackson Laboratory. The Jackson Laboratory is fully accredited by the American Association for Accreditation of Laboratory Animal Care. The author is indebted to Dr. T. H. Roderick for making available certain data prior to publication.

Table I. The distribution of alleles at 16 loci in 27 inbred strains

Inbred strain	<i>Es-1</i>	<i>Es-2</i>	<i>Es-3</i>	<i>Mod-1</i>	<i>Id-1</i>	<i>Gpd-1</i>	<i>Gpi-1</i>	<i>Pgm-1</i>	<i>Pgm-2</i>	<i>Ldr-1</i>	<i>Hbb</i>	<i>Dip-1</i>	<i>H-2</i>	<i>Hc</i>	<i>rd</i>	<i>In</i>
A/HeJ	b	b	c	a	a	b	a	a	a	a	d	b	a	0	+	<i>In</i>
AKR/J	b	b	c	b	b	b	a	a	a	a	d	b	k	0	+	<i>In</i>
AU/SsJ	b	b	c	b	b	b	b	b	a	a	p	b	q	0	+	<i>In</i>
BA1B/cJ	b	b	a	a	a	b	a	a	a	a	d	a	d	1	+	<i>In</i>
BRP/J	b	b	a	a	b	a	a	b	a	a	d	b	p	1	<i>rd</i>	<i>In</i>
BUB/BnJ	b	b	c	a	b	b	b	a	a	a	d	b	q	1	<i>rd</i>	<i>In</i>
CBA/J	b	b	c	b	b	b	b	a	a	a	d	b	k	1	<i>rd</i>	<i>In</i>
CE/J	b	b	c	a	a	a	a	a	a	a	s	b	k	0	+	<i>In</i>
C3H/HeJ	b	b	c	a	a	b	b	b	a	a	d	b	k	1	<i>rd</i>	<i>In</i>
C57BL/KsJ	a	b	a	b	a	a	b	a	a	a	s	a	d	1	+	<i>In</i>
C57BL/6J	a	b	a	b	a	a	b	a	a	a	s	a	b	1	+	<i>In</i>
C73R/cdJ	a	b	a	b	b	a	a	a	a	a	s	a	k	1	+	<i>In</i>
C57L/J	a	b	a	b	b	a	a	a	a	a	s	a	b	1	+	<i>In</i>
C58/J	b	b	c	b	a	a	a	a	a	a	s	a	k	1	+	<i>In</i>
DFV/1J	b	b	c	a	b	a	a	b	a	a	d	b	q	1	+	<i>In</i>
DFV/2J	b	b	c	a	b	b	a	b	a	a	d	b	d	0	+	<i>In</i>
LE/J	b	b	c	a	a	b	a	a	a	b	d	a	b	1	+	<i>In</i>
M/J	b	b	c	a	a	b	a	b	a	a	s	b	k	1	+	<i>In</i>
P/J	b	b	a	b	b	a	a	b	a	a	d	b	p	1	<i>rd</i>	<i>In</i>
PH/J	b	c	c	a	b	a	a	b	a	a	d	b	u	1	<i>rd</i>	<i>In</i>
R/J	b	b	c	a	a	a	a	a	a	a	d	b	k	0	+	<i>In</i>
RH/2J	b	b	c	b	a	b	a	a	a	a	s	b	r	1	+	<i>In</i>
S/J	b	b	c	a	b	b	a	b	a	a	s	b	s	1	<i>rd</i>	<i>In</i>
St/J	b	b	c	a	b	b	a	a	b	a	s	b	v	1	+	<i>In</i>
S7bJ	b	b	b	b	a	b	a	a	a	a	d	b	k	0	<i>rd</i>	<i>In</i>
S7R/J	b	b	c	a	a	b	b	b	a	b	s	b	q	0	<i>rd</i>	<i>In</i>
129/J	b	b	c	a	a	a	a	a	a	a	d	b	b	1	+	<i>In</i>

cluded in this study are: the major histocompatibility locus, $H-2^b$; presence of one component of complement in serum, Hc^1 ; retinal degeneration, rd^b ; and 7-12 dimethylbenzanthracene inflammatory response, In^8 . These loci were chosen because they permit the formation of a large, complete, strain-locus matrix, and because they should be relatively immune to conscious selection during the formation of these strains. Coat color loci were excluded on the basis of the latter criterion. The chief basis for choice of strains was that complete information was available for the 16 loci. In cases where complete information was available on two or more sublines that were identical at all sixteen loci, only one of the sublines was included in the analysis.

Each allele at all loci was coded 1, 2, 3, . . . etc., for numerical analysis. Since there is no basis for considering one allelic difference as being more important than another, the number of loci at which two strains share identical alleles was used as a criterion of similarity. In this fashion, a 27 times 27 similarity matrix was formed with the number of shared alleles between the i^{th} and j^{th} strains entered into the ij^{th} location in the matrix.

Two methods of analyzing the results were used. The first method was to compute the average number of shared alleles of each strain with every other strain. Strains were then characterized as relatively common or uncommon with regard to their genotype array. The maximum number of shared alleles was also found to be a useful measure of commonness in this set of data.

In order to evaluate relationships objectively, the eigenvalues and eigenvectors of this matrix were calculated. These calculations were made by a standard computer routine on an IBM 1130 computer in extended precision. This procedure is formally similar to a principle components analysis of a correlation matrix⁴. The first two eigenvectors were transformed so that the variance of each vector was proportional to the magnitude of the corresponding eigenvalue. The new vectors were taken to be the first and second

coordinates of the strains, respectively, for the purpose of mapping the strains in two dimensions.

Results and Discussion

If these inbred strains are considered a representative sample of mouse genes, then we can decide whether a strain is relatively common or distinctive on the basis of the average number of shared alleles with the 26 other strains. The results are presented in Table II. Strains are arranged in order of distinctiveness, the more distinctive strains listed first. Strains SWR/J and AU/SsJ appear to possess the most rare alleles. These strains do not seem to be related to each other (Table I). The closely related C57-strains are also very distinctive, but their distinctiveness is somewhat masked by their close relationship to each other. Also distinctive is the PL/J strain. The SM/J strain is the only strain so far detected that carries a variant allele at the $Pgm-2$ and $Mod-2$ loci³. Evidently, one component of the inheritance of this strain is not shared by other inbred strains, whereas the other components are evidently quite common. This is consistent with the known origin of this strain². Thus, the SM/J strain is a good source in searching for new variants, even though its genotype has an overall similarity to other strains.

Eigenvalues and eigenvectors of the similarity matrix were computed. The relative magnitudes of the eigenvalues represent the proportion of variation accounted for by the first, second, third, . . . , n^{th} dimensions, and thereby provide a measure of the complexity of the relationship structure. If the strains were all unrelated, all of the eigenvalues would tend to be small and approximately equal. On the other hand, if all of the strains could be placed into two or three tight clusters of related strains, the first few eigenvalues would account for nearly all of the variation. The first 10 eigenvalues expressed as a percentage of the sum of all n eigenvalues are shown in Table III. The first two dimensions account for over 70 percent of the differences between strains. The next five dimensions contribute smaller and similar proportions, whereas the other dimensions add little information. This can be interpreted to mean that many of the strains belong to clusters of related strains that are delineated in the first two dimensions. Other strains are relatively unrelated to these clusters and to each other, and defy simple mapping into a few related groups. This interpretation is consistent with what is known about strain origins⁷. The first

Table II. Strains ranked according to distinctiveness

Strain	Mean no. shared alleles	Strain	Mean no. shared alleles
SWR/J	7.65	BALB/cJ	9.81
AU/SsJ	7.73	C3H/HeJ	9.81
C57BL/Ks	8.04	ST/bJ	9.81
C57BL/6J	8.08	AKR/J	9.85
C57L/J	8.54	CBA/J	9.85
C57BR/cdJ	8.73	LP/J	9.85
PL/J	8.81	C58/J	9.88
P/J	9.23	DBA/1J	9.88
RF/J	9.35	MA/J	10.00
SJL/J	9.38	129/J	10.11
BDP/J	9.42	CE/J	10.15
SM/J	9.46	A/HeJ	10.19
DBA/2J	9.54	RIII/2J	10.27
BUB/BnJ	9.73		

Table III. Proportion of variation accounted for by the first 10 orthogonal vectors of the similarity matrix

Vector	Percent	Vector	Percent
1	60.4	6	3.3
2	9.7	7	2.7
3	5.3	8	2.2
4	3.9	9	1.7
5	3.5	10	1.3

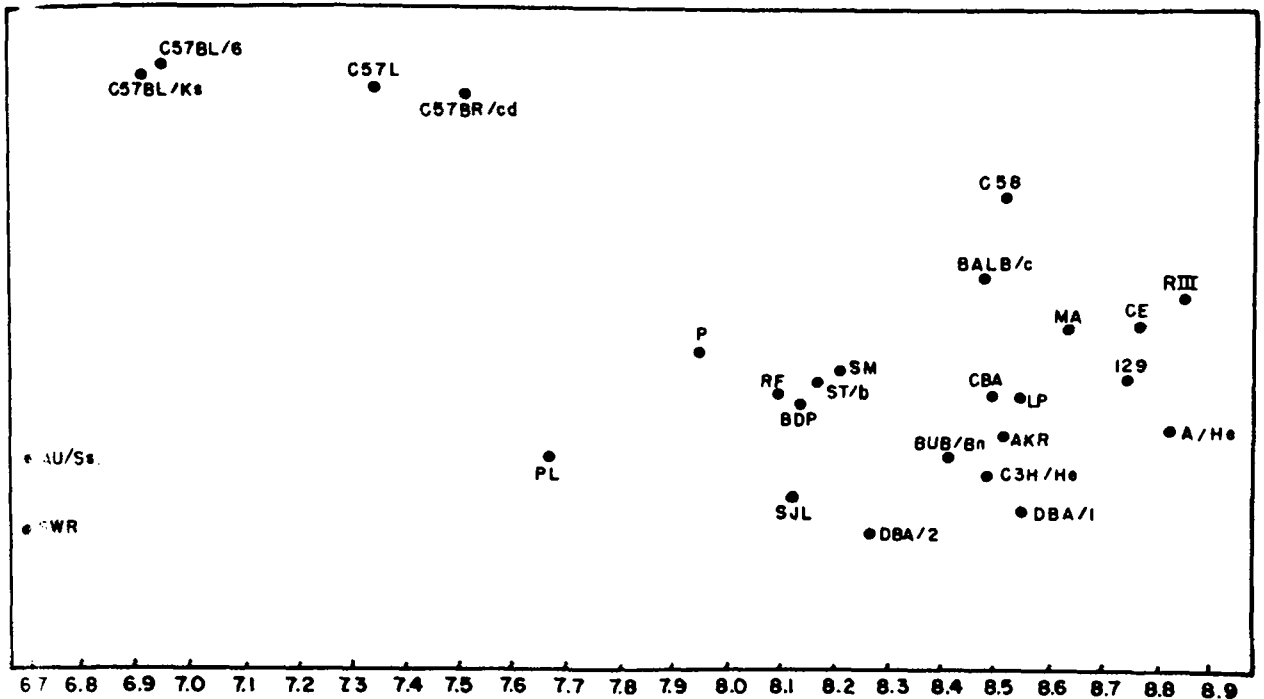


FIGURE 1—Positions of 27 inbred strains in two dimensions as determined by an eigenvector analysis of

similarity matrix. See text for details of the scale of coordinates and the linear distance between strains.

and second transformed eigenvectors were used to map the position of 27 strains in two dimensions (Figure 1). The scale of the coordinates in Figure 1 is arbitrary; the linear distance ($d = \sqrt{x^2 + y^2}$) between strains represents the best estimate of the relative genetic distance between strains possible in two dimensions.

Substrains of inbred strains map closer to each other than they do to other strains, e.g., C57BL/6J and C57BL/Ks, C57BR/cdJ and C57L/J, DBA/1J and DBA/2J. Furthermore, the distances separating substrains are less than those separating other strains. One striking feature of this arrangement is that the C-7- strains are quite separate from the others. It is ironical that the "standard" inbred mouse strain, C57BL, appears to be very atypical genetically. Strains DBA, CBA, and C3H share some common ancestry, C3H and CBA being derived from a cross involving DBA. Although these strains appear in the same area in the present mapping scheme, other strains intervene. In particular, the similarity of SJL/J to C3H/HeJ would not have been predicted, since these are thought to be of independent origin. Actually, this strain-pair differs at 4 of the 16 loci; hence, this is an example of the kind of distortion that occurs when a multidimensional problem is represented in only a few dimensions. Also, if more loci had been included in the study it might be that the apparent similarity of these strains would decrease. It should be emphasized that the spatial relationships repre-

sented in Figure 1 suffer from at least two limitations. The first is the relatively few loci sampled. Clearly, if information on more loci were available, the positions would change. Eventually, as more loci are added, stable relationships should emerge. The other limitation is the unavoidable distortion produced by representing a multidimensional problem in two dimensions.

Certain substrains have been excluded from this analysis because they are identical at all 13 loci to one of the other strains. Strains C57BL/10J, C3-HeB/FeJ, and A/J are indistinguishable from C57BL/6J, C3H/HeJ, and A/HeJ, respectively, at all 16 sampled loci. This identity of sublines testifies to the genetic homogeneity of the strain at the time of separation of sublines, and to the mutational stability of these polymorphic loci. Where related substrains differ at more than one independent locus, e.g., DBA/1J and DBA/2J, C57BR/cdJ and C57L/J, the probable explanation is that the strains were still segregating for a few chromosomal segments at the time of subline separation. Thus, any differences detected between such pairs of strains are likely to fall into one of a few small clusters of linked genes. This situation provides leads to the linkage of such differences. Newly detected differences between such substrains should be tested for linkage with other known differences between the same strain pairs.

The amount of information any single locus contributes to relationships among strains depends on

the number of alleles and the frequency of each allele. A locus with only one allele obviously contributes nothing. At the other extreme, a highly polymorphic locus, at which every strain has a private allele, similarly contributes no information on strain relationships. Thus a maximally informative locus would necessarily be intermediate in degree of polymorphism.

The eigenvector method of analysis used here may not be optimal. However, the method does have certain desirable features. First, it makes use of all the information in the similarity matrix. Second, it provides a unique solution. It seems likely that some information is lost in going from Table I to the similarity matrix, since it is not possible to reconstruct Table I from the similarity matrix. If data were available on additional loci, the analysis would be more powerful. Partial strain distributions exist for many other loci. Various attempts to make use of such incomplete information have all failed to yield reliable results. Thus, it is quite important that all strains be compared with respect to the same loci. An additional complication that might have prejudiced the results is linkage. If two loci are closely linked, they do not provide independent information about strain relationships. None of the loci used in this study are known to be closely linked, although two of these have not yet been fully tested with the others.

It is hoped that the information presented here will be helpful to investigators who contemplate genetic experiments utilizing inbred strains. When the object is to detect variation, a strain survey should include such relatively distinctive strains as SWR/J, AU/SsJ, one of the C57- strains, PL/J, P/J, RF/J, SJL/J, SM/J, and DBA/2J. If the object is to cross two or more lines to establish a genetically heterogeneous population for a selection experi-

ment, reference to Tables I and II should permit a rational choice of progenitor strains. In other situations, knowledge of strain relationships should allow more cogent interpretations of differences. For example, a large difference between two closely related strains is likely to be due to at most a few loci.

Summary

Using data on the strain distribution of alleles at 16 polymorphic loci, the genetic relationships among 27 inbred strains, whose pedigree relationships are largely unknown, has been studied. The 27 strains were ranked according to their distinctiveness. The positions occupied by the strains relative to each other is represented in a two dimensional plot. The utility of this information and certain limitations of the data are discussed.

Literature Cited

1. CINADER, B. S., S. DUBISKI, and A. C. WARDLAW. Distribution, inheritance, and properties of an antigen, MuBl, and its relation to hemolytic complement. *J. Exp. Med.* 120:897-924. 1964.
2. MACARTHUR, J. W. Genetics of body size and related characters. I. Selecting small and large races of the laboratory mouse. *Am. Nat.* 78:142-157. 1944.
3. RODERICK, T. H., F. H. RUDDLE, V. M. CHAPMAN, and T. B. SHOWS. Biochemical polymorphisms in feral and inbred mice (*Mus musculus*). *Biochem. Genet.* 5:457-466. 1971.
4. SEAL, H. L. *Multivariate Statistical Analysis for Biologists*. Methuen, London. 1964.
5. SIDMAN, R. L. and M. C. GREEN. Retinal degeneration in the mouse: location of the *rd* locus in linkage group XVII. *J. Hered.* 56:23-29. 1965.
6. SNELL, G. D. and M. CHERRY. Loci determining cell surface alloantigens. In Proc. Symp. on RNA Viruses and Host Genome in Oncogenesis. In press.
7. STAATS, J. Standardized nomenclature for inbred strains of mice: fourth listing. *Cancer Res.* 28:391-420. 1968.
8. TAYLOR, B. A. Strain distribution and linkage tests of 7,12-dimethylbenzanthracene (DMBA) inflammatory response in mice. *Life Sci.* 10:1127-1134. 1971.