Genetic Divergence Among Inbred Strains of Mice

William R. Atchley
Laboratory of Genetics, University of Wisconsin
Madison, Wisconsin 53706

Continued efforts to produce more and more refined estimates of genetic divergence among groups of organisms has had a rather spectacular effect on both the growth of theory in population genetics and the production of new and more sophisticated analytical techniques. Unfortunately, few groups of organisms exist whose evolutionary history is sufficiently well known to be used to evaluate these theoretical or analytical developments.

Some of the inbred strains of the house mouse, Mus domesticus, provide excellent material to study genetic divergence since the actual sequence of branching and the relative times of separation for the various taxa are well-known. Extensive data on genetic divergence exists at various levels of biological organization. Indeed, over 700 genetic loci have been described from about 250 distinct inbred lines, and linkage group affinities are known for over 400 of these loci. Extensive biochemical differences at individual loci exist between strains and substrains so that genetic divergence can be measured unequivocally (Green, 1981).

Walter Fitch and I have been using genetic data from inbred mouse strains to explore a number of questions about genetic differentiation. Two specific questions will be examined today. 1) Does genetic divergence among inbred strains for single gene loci accurately reflect the known genealogy? 2) Is the magnitude of genetic divergence among inbred strains accurately predicted by the classical model which states that divergence in inbred lines a function of the initial heterozygosity in the founding populations, contamination from outcrossing, and mutation?

METHODS AND MATERIALS

Strains of mice. The ten strains of mice examined (together with their coat color) were A/HeJ (albino), AKR/J (albino), BALB/cJ (albino), CBA/J (agouti), C3H/HeJ (agouti), C57BL/6J (black), C57BR/cdJ (brown), C58/J (black), DBA/1J (grey), and DBA/2J (grey). Coat colors are given to assess the probability of interstrain contamination. The known genetical relations among these inbred strains were described by Staats (1980) and Festing (1979) and are shown in fig. 1.

---

Loci examined: Staats (1980) gives inbred strain distributions for 158 loci. Loci from this list were included in this analysis if they met the condition that data were available for at least 8 of the 10 strains examined here. This resulted in ninety-seven loci being included in these analyses, considerably more loci than have been examined in previous studies of divergence among inbred strains (e.g., Taylor 1972) or wild populations.

All data were for Jackson Laboratory strains except that, where data on Jackson strains were missing, other substrain data were used provided that all substrains possessed the same allele at that locus. Twenty-three loci are unvaried among these ten strains, eight contain only unique variants and 66 are cladistically informative (at least two non-unique alleles) (Appendix A). Of the 97 loci, 62 are proteins, 33 are immunological, and two are unassigned. All 97 loci were used in all parts of this study. The number of alleles at these loci in these strains averages 2.01 with three loci (Hba, H-2, Igh-1) having five alleles, three loci (Aox-1, Igh-2, Mls) having four alleles, nine loci having three alleles, 59 loci having only two alleles, and the remaining 23 having only one. In twelve other strains not discussed here, three of these loci (those already showing five alleles) possess two or more alleles not present here, while eight other loci possess one extra allele. One of the latter loci, Igh-2, already has four alleles in this study. Of the 15 loci possessing more than two alleles, all but four of these loci are immunological. Of the 87 loci whose chromosomal location we know, all are autosomal.

Data from the cladistically informative loci are over 96% complete with only 23 of 660 data points missing. Data for forty-seven of the 66 loci are complete. One strain (C58) is missing seven data points, three strains are missing four and the rest are missing two or less. AKR, C57BL, and DBA/2 have complete data.

For the cladistically non-informative loci, 23 loci are homoallelic, seven have two alleles one of which is unique to one strain, and one locus (Ly, delta aminolevulinate dehydratase) has three alleles, two of which are each unique to its strain. These 31 loci are similar to the other 66 loci except for more missing data (10%) and containing a larger fraction of loci coding for proteins.

Estimating Heterozygosity and Divergence. Divergence between two inbred strains may arise through any of four factors: 1) at loci that were polymorphic in the ancestral population, different alleles may have been fixed in the two inbred lines; 2) some polymorphic loci may not yet have gone to fixation in one or the other of the inbred lines (residual heterozygosity); 3) new alleles created by mutation may have been fixed; and 4) new alleles, introduced by mishaps from another line, may have been fixed (contamination).

The equation for divergence describing the fraction of loci that is different is

\[ d_{ij} = c_{ij} + 2m_y_{ij} (1 - c_{ij}) \]  

(1)

in which \( d_{ij} \) = the expected fraction of allelic differences per locus between (inbred) strains i and j over the loci examined, \( h = the heterozygosity at these loci in the ancestral wild population, c_{ij} = the correction factor for h as a result of any genetic relationship between i and j (i.e., one minus the coefficient of kinship where \( c_{ij} = 0.75 \) for full sibs), \( m = the rate of fixation \)
of alleles per year in these strains, and \( y_{ij} \) = the number of years since strains \( i \) and \( j \) were separated. For these equations, \( c_{ij} \) is known and \( y_{ij} \) is reasonably estimated.

The first term on the right of equation 1 is the divergence arising from the reduction of polymorphism present in the population at the time the strains were separated (initial plus residual heterozygosity), while the second term is the divergence accumulated by the fixation of newly arising or introduced variants since that time by the other two factors. The term in parentheses corrects for the inability to observe new differences in positions that were already different at the time the two lineages separated.

Genetic relationships among these ten strains give the values of \( c_{ij} \). For these strains, a value of \( c_{ij} = 1.0 \) is probably valid only for the 21 comparisons of C57BL, C57BR and C58 versus the other seven lineages. It is unknown how long ago these lineages were separated from each other.

C57BL and C57BR were derived from a cross of a single Lathrop stock male and female in 1921 so their gametes are related as full sibs (\( c_{ij} = 0.75 \)) and 25% of the alleles are identical by descent. The C58 strain arose from the same male as C57 but a different female, i.e., C58 is related to the two C57 strains as a half sib (\( c_{ij} = 0.875 \)) and 12.5% of the parentally derived alleles are identical by descent.

CBA and C3H both arose in 1920 from a single cross of the Bagg and DBA lines. DBA had been inbred since 1909 and the Bagg line would later be inbred to produce BALB/c. CBA and C3H, therefore, should be identical by descent for all alleles arising from the DBA line and related as full sibs for those alleles from the Bagg line (\( c_{ij} = 0.375 \)). Strains DBA/1 and DBA/2 were created in 1929 by separation of the already inbred DBA stock; thus, all their alleles should be identical by descent (\( c_{ij} = 0 \)) assuming that inbreeding was complete after the 20 years of inbreeding.

Of the 45 pairwise comparisons described here, the numbers involving \( c_{ij} = 0.875, 0.75, 0.375 \) and zero are 2, 1, 1 and 1, respectively.

**RESULTS**

**Phylogenetic relationships among strains:** A phylogeny obtained by parsimony analysis of these 10 strains for the 97 loci by the method of Fitch (1971) is given in Figure 2. Other phylogenetic methods, including UPGMA (Sokal and Sneath, 1973), EVOLVES (Fitch and Margoliash, 1967) neighborliness (Fitch and Margoliash, 1981) and distance Wagner (Farris, 1972), all give the same topology (but not necessarily the same branch lengths) except for the C57 - C58 group where no method should be expected to separate the half-sib derived C58 from the full-sib derived C57 strains. The AKR strain, whose genealogy is not well known, is shown to be allied to the DBA-CBA-C3H strains.

Since these methods all assume a bifurcating structure, the CBA/C3H strains are made the sister group of one of the hybridizing lines. Moreover, the methods all choose the inbred DBA lineage as the sister group and this should be expected since CBA and C3H should share more alleles with descendants of their inbred parent than with descendants of their outbred parent.
This level of congruence in results obtained using different phylogenetic algorithms is unusual since analyses of other biological data by different methods often give quite divergent results. This faithful reproduction of the genealogy of the strains as we know it and the congruence of the results among different methods argue strongly for the robustness of these data to reproduce the actual phylogenetic history for these 10 strains.

There is a total of 169 gene substitutions required to explain the divergence but those in the earlier branches are more properly interpreted as changing originally polymorphic loci to distinct monomorphic loci.

Amount of divergence between strains: There are extensive differences among these 10 inbred strains of mice for these 97 loci. Percentage difference ranges from 14% between DBA/1 and DBA/2 to 54% between DBA/2 and C57BL (Table 1). The mean divergence over all strains and loci (and standard deviation) is 0.40 (± 0.10) while that between the C57 - C58 group and all others is 0.47 (±0.04).

Equation (1) can be rearranged to give,

\[ d_{ij} = c_{ij} h (1 - 2m_{ij}) + 2m_{ij} \]  

The estimate of divergence, \( d_{ij} \), for the five pairs of mice strains having \( c_{ij} < 1.0 \) are plotted in figure 3. The \( d_{ij} \) values for the line in figure 3 are adjusted to a divergence time of 58 years, a value determined by assuming all the genetic data were, on average, described in the year 1979, the date of the Festing reference. The divergence time for C57BL, C57BR and C58 (which arose in 1921) is 58 years while that for the C3H and CBA strains (which arose in 1920), is 59 years. The divergence value in Table 1 of the latter pair is standardized to a time scale of 58 years by multiplying 0.20 by 58/59. The two DBA strains originated in 1929 and hence had only 50 years of divergence. Accordingly, their divergence in table 1 is multiplied by 58/50.

The least squares regression equation (with standard errors for slope and intercept) of divergence onto one minus the coefficient of kinship is

\[ d_{ij} = 0.165 (±0.029) + 0.079 (±0.043)(c_{ij}) \]

The intercept is significant with lower and upper 95% confidence limits of 0.072 - 0.257. The regression slope is not significantly different from zero with a probability associated with \( b < 0 \) between 0.1 and 0.2. The intercept of the line, determined largely by the DBA/1 - DBA/2 comparison, is the divergence after 58 years, and provides a fixation rate of 0.165/116 = 1.4 x 10^{-3} fixations per locus per year. Substituting the upper and lower 95% confidence limits values for the intercept we obtain fixation rates ranging from 6.2 x 10^{-4} to 2.2 x 10^{-3} fixations per locus per year.

Heterozygosity per locus. Our estimate of the heterozygosity for the ancestral mouse population from which these stocks arose (obtained by dividing the slope of the line in fig. 3 by one minus the intercept) equals 0.09 in this case. The relatively large standard error of the slope makes the estimate of the original heterozygosity uncertain. However, this value is the same as the 0.09 estimated for 4 populations of Mus musculus musculus (Selander et al., 1969). Estimates for wild populations of M. M. domesticus (Selander et al., 1969, Rice et al., 1980) range from 0.06 to 0.09 to 0.11 for M. m.
brevirostris (Selander et al., 1969). If a new population were created by mixing equal numbers of each of these ten strains, its heterozygosity would be 0.33.

**Approximate age of ancestral populations.** It is not known when the ancestors of these inbred lines were derived from wild populations. The original stock is assumed to have been derived from the pet mouse trade, possibly from a single region in England (Ferris et al., 1983). Using mitochondrial DNA, Ferris et al. (1982) suggested that the "old" inbred strains of mice (including those strains examined here) were derived from a single female of *Mus domesticus*, which is the house mouse of Western Europe and the Mediterranean region. The mtDNA data further indicate that the mtDNA type found in these old strains occurs in only 0.04 of wild *M. domesticus* surveyed.

We do not know when the Bagg, Lathrop, DBA, Detwiler and Cold Spring Harbor stocks that gave rise to these inbred lines were separated from each other nor do we know their level of inbreeding. Probably, they did not all become separated at the same time. Nevertheless, an average time of separation can be estimated from the block of 7 x 3 divergence values in the upper right half of table 1 which represent the C57 - C58 group against the others and which average 0.47 (σ = 0.042) or five times the original heterozygosity. Substituting into equation 1, we obtain

\[
0.47 = 0.09 + 2y \times (1.4 \times 10^{-3}) \times (1 - 0.09).
\]

Thus, y, the time of divergence of these strains is approximately 150 years prior to the Festing 1979 reference or about 1830. This is a reasonable estimate in view of the history of mouse stocks and the mtDNA evidence.

**Other results.** None of the results reported above are materially changed by examining the protein loci separately from the immunological loci except that the overall rate of divergence is greater for the latter and the confidence intervals are greater for each. We are unable to detect any non-randomness with respect to the chromosomes on which the divergence occurred between pairs of strains when corrected for the number of loci and alleles.

**DISCUSSION**

**Divergence and Heterozygosity:** As stated earlier, the divergence seen among these strains may have its source among any of four factors, not including systematic biases which will be discussed presently. The first of these factors is initial heterozygosity. Our value of 0.09 has a high standard deviation (0.04) but it is consistent with estimates from other mice populations on both sides of the Atlantic Ocean. Moreover, it is six standard deviations below the mean heterozygosity of a pool of these ten strains. Hence, one has difficulty accepting the proposition that the observed divergence is the simple consequence of the segregation of original heterozygosity.

The second factor is residual heterozygosity. Since these strains appear to be fixed for one allele at each of these loci in each of these strains (as they should be after more than 50 years of reported full-sib mating for the youngest of these strains), residual heterozygosity cannot be a significant source of any divergence seen. This leads to consideration of the fixation
Fixation rates: The value of $1.4 \times 10^{-3}$ is very high. On the other hand, since these are inbred lines produced by brother x sister matings, the effective population size is only two; hence, a rapid fixation rate is reasonable provided there is a store of genetic diversity on which to operate. Since the original supply of genic diversity is estimated to be only 0.09, most of the currently observed diversity would appear to have arisen since.

Rate of introduction of new variants: Since the effective (and true) population size is only 2, a newly introduced variant, even if deleterious, has nearly one chance in four of being fixed if it is not a gene being monitored for purity of the lineage. The fixation rate (Crow and Kimura, 1970) is $2NuF$, where $u = \text{rate of introduction of new variants}$, $s = \text{selection coefficient}$, and $F$, the probability that a new variant will be fixed, is

$$F = \frac{(1 - e^{-2s})}{(1 - e^{-4Ns})}.$$  

For small values of $Ns < 0.05$, equation 3 reduces to $1/2N$ and the fixation rate = $2Nu/2N = u$, the introduction rate. If one could rule out contamination (or biased data) as the source of the new variants, then $u$ would be the mutation rate. We therefore estimate that the mutation rate by this method could be as great as to be $1.4 \times 10^{-3}$ mutations per locus per year or $5 \times 10^{-4}$ per generation, an estimate several orders of magnitude greater than the conventional estimates in mice (Schlager and Dickie, 1967).

Are the strains significantly contaminated?: It is widely suspected that many of the inbred mouse strains have been contaminated through outbreeding with other mice and, on the surface, contamination might seem the easiest explanation for the high rate of divergence among strains reported here. While we do not wish to discount the possible presence of contamination, we believe that contamination cannot readily account for our observations for several reasons. The first reason is that contamination can not be from mice of these or closely related strains since this would have the effect of reducing differences.

The second reason is that there would need to be several different contaminating mice, each one more than 40 per cent different from the others, because the major groups derived from the Bagg, DBA, and Lathrop stocks are all different from each by that amount.

The third reason is that wild Mus domesicus are not sufficiently divergent to be the source. Richard Sage (personal communication) has provided us with his data on three inbred strains (C57BL/6, BALB/c and DBA) and eight mice from Morocco, Spain, England, Italy, Switzerland and Yugoslavia assayed for 56 protein loci. The probability of two gametes, one from each pool, having different alleles at a given locus is only 0.136.

The fourth reason is that the contaminating mice would have had to introduce their divergent genes in a very special way. As described above, we have attempted to determine the phylogenetic relationships of these ten strains using a number of different algorithms including UPGMA, parsimony, EVOLVES, neighborliness and distance Wagner and, unlike the case of most biological data, all algorithms give the same tree except for their inability to separate the half-sib C58 from the full-sib C57 strains. Moreover, the tree produced by
these methods is identical to the phylogeny as we know it. It is unlikely that contamination could occur in such a widespread fashion and still preserve the phylogenetic information so well that five different methods all correctly obtain the known relationships and completely agree on those that are not known.

The fifth reason is that we have developed a method (Fitch and Atchley, unpublished) to detect hybridization and it clearly identifies the CBA/C3H group as a hybrid of the DBA/1 - DBA/2 and the A - BALB/c groups. It is difficult to imagine the preservation of the hybrid information in the face of great contamination of the strains.

It would be possible to discount both reasons 4 and 5 if all the contamination occurred prior to all known strain separation, but then we are left without a contamination explanation for the divergence since the separation within the Bagg and DBA strains.

The sixth reason has its basis in fig. 2. If one considers the root of the tree to be in 1830 and the tips in 1979 on the basis of the analysis from fig. 3 (which has nothing to do with parsimony), then 1920 would fall at the point shown by the arrow if time is proportional to distance. The three strain separations that occurred in 1920 and 1921 are those nodes within the dashed lines on the tree where distance is proportional to the number of allelic substitutions during descent. The correspondence of the two suggests that the whole process is behaving in a reasonably clock-like manner. Such behavior would not seem to be a very likely result in the presence of serious, multiple contamination problems.

One might also note that contamination would need to be explained by male interlopers since Ferris et al. (1982) have shown, on the basis of mitochondrial DNA restriction analysis, that all these strains arise from a single ancestral female.

Biases in the data: An alternative basis for explaining such divergence among strains includes several potential systematic biases in the data. There are at least eight.

Bias 1: Loci that do not vary may go unreported. Unvaried loci may not be reported because of their apparent inutility. The extent to which this might occur is unknown but, if it exists, it would cause our estimate of divergence to be too high. However, it does not seem likely that there are three hundred unreported unvaried loci, the number required to reduce the average divergence from 0.40 to 0.10.

Bias 2: Breeders may consciously or unconsciously select for mice that are healthy, docile, have large litters, mature early, and/or have long reproductive periods. However, it does not appear a priori that this practice would have any systematic effect on which alleles would be fixed. We will, however, propose in the Discussion a way that it might effect the estimated initial heterozygosity.

Bias 3: Missing data may affect the calculations. The effect would depend upon whether there was more or less divergence in those loci that have missing data than in those that do not. Since there is no reason to believe loci that are missing data are different from those that are not, there is no a
Priori directional bias. Moreover, since the amount of missing data is small, its effects can not alter greatly the total amount of divergence.

Bias 4: Inbreeding may have occurred in the stocks prior to brother x sister mating. Previous inbreeding would cause the allelic difference at the time of instituting sib mating in the C57/C58 and in the CBA/C3H group to be overestimated. This, in turn, would lead to underestimation of the amount of divergence since that time and, as a result, give an estimation of the rate of fixation that is too low.

Bias 5: Loci monitored as diagnostic for a strain will not be permitted to change. Hence, the estimated rate of fixation would be too low.

Bias 6: The analytical methods may have failed to detect all the genic diversity present. This would cause the estimation of the amount and, hence, the rate of fixation to be too low.

Bias 7: We do not know exactly when the alleles at various loci in a strain were determined. The majority of genotypes were obtained from Festing (1979) so that, to the extent that average date of fixation is prior to 1979, the estimation of the fixation rate is too low.

Bias 8: Equation 3 used to calculate the divergence rate errs to the extent that it assumes that all deleterious variants have an $s$ value between zero and 0.025. For larger values of $s$, the fixation rate must underestimate the rate at which new variants are introduced.

Only bias 1 would exaggerate the rate of divergence and, hence, the fixation rate. However, it does seem unlikely that there are three hundred examined but unreported invariant loci, the number required to make the estimated fixation rate of $1.4 \times 10^{-3}$ comparable to previous estimates of the mutation rate. Because the evidence appears to be against both bias and contamination explaining these data, the fixation rate may be the mutation rate and, thus, there may be a high mutation rate in these strains (but see below).

Evidence against a high mutation rate: Johnson et al. (1981) have been studying mutagenesis in crosses between DBA/2J and C57BL/6J. Examining 20 loci (all among the 97 we examined) in 3848 progeny of such crosses (76,960 observations) where no mutagen was employed, they found no new mutants. Thus, their mutation rate is two orders of magnitude below our fixation rate. This is particularly telling since this analysis was on inbred strains of mice.

Other puzzles with these data: In addition to the puzzle of a high fixation rate, there are two additional puzzles.

The second puzzle is that there are few published reports of new mutants (other than those for which new strains are selected) arising within these strains. Surely if mutations are occurring at a rate greater than $10^{-3}$ per locus per generation, many new mutations might well have been observed in the generation of their initial occurrence. The major opportunity to see such new mutations would appear to be within the laboratories of commercial breeders. Absence of such reports may mean they were not looked for, not seen, discounted, or not deemed of sufficient importance to report. This puzzle is not addressed further.
The third puzzle is, why do we see only two different alleles at so many loci? Of the 145 substitutions of one non-unique allele for another, 71 are parallel (or back) substitutions of some allele from among the other 74. At 52 of the 66 cladistically informative loci, there are only two alleles but these 52 loci require 113 of the 169 total substitutions observed and only 41 of these 113 separate the three major ancestral stocks. How can there be so many (at least 113 - 52 = 61) new mutations being fixed in 52 positions without additional new variants arising? Why aren't there three or more known variants at many of these loci rather than only two? The occurrence across these strains of only two variants at the vast majority of the loci might suggest the retention of original heterozygosity from the ancestral population were it not for the five-fold increase in heterozygosity and the clear requirement for the fixation of many new mutants. Another view of this puzzle is that, with an average of only two alleles per locus, the maximum heterozygosity is only 0.5, achieved when both alleles have the same frequency. The C57 - C58 group has an average divergence from the other strains of 0.47, which is not significantly different from 0.5. This result implies they are nearly maximally diverged from the others given the allelic composition of these ten strains.

Contamination does not readily explain this property of the data.

Reconciliation: An Alternative Hypothesis: We have presented evidence against all the readily conceivable explanations of these data. The importance of inbred mice and the degree of divergence is so great as to demand that an explanation be found. It is the primary purpose of this paper to pose, once more, the problem. It is a secondary purpose to put forward, tentatively, a hypothesis that, if true, might resolve two of the puzzles described.

This hypothesis is developed from a suggestion that there is direct selection for heterozygosity in the creation of the inbred lines which, through linkage, even effects loci that do not themselves have recessive deleterious alleles. When developing the A strain, Strong (1978) deliberately chose the most vigorous male and female of each litter to create the next generation because he was concerned that inbreeding depression would wipe out the lineage before it became homozygous at all loci. This regimen is almost certainly selecting the most heterozygous animals for further breeding. However, fearing that his might still not be enough, Strong started several such lines in hopes that at least one lineage might survive the inbreeding crisis.

The question then arises, how heterozygous might the first pair of siblings be? Between any two siblings selected at random, one expects them to carry three-quarters of the heterozygosity of their parents, but the two most heterozygous mice in the litter might contain a larger fraction of it. Then the question becomes, what was their parents heterozygosity? If we assume that each had the population average of 0.09, then the total fraction of heterozygous loci in their combined pool could be as large as 0.18. This is still considerably below what would be needed to explain the many divergences in excess of 0.4. But, if even this procedure of choosing the most vigorous offspring should prove wanting, crossing two such parallel inbred lines might restore vigor, albeit at the cost of slowing the march to total homozygosity. The effect, however, would be to expand once more the effective fraction of loci that were initially heterozygous.
This hypothesis asserts, then, that the divergence we see today was determined very early, largely as the result of initial selection for heterozygosity. Note that if originally there were only two major alleles at most of these loci, this hypothesis would also explain why there are only two alleles present at so many of the loci examined. It also avoids the need for a high mutation rate to explain the diversity. A major difficulty with it is that it is not clear how many of the original strains had such a methodological origin nor how the hypothesis can be tested. Moreover, it cannot account for the divergence between DBA/1 and DBA/2.

Conclusion: The data document a diversity among inbred strains of mice much too great to be accounted for by the usual simple genetic mechanisms. Moreover, we present evidence against the commonly expressed belief that this diversity is the result of contamination of the stocks. An alternative mechanism, i.e., selection for heterozygosity, could explain much of the observed divergence for several strains. But we are less concerned that one of the alternatives be proved correct than we are that an explanation for the documented divergence be sought that does not rely on the too facile explanation of contamination.

Literature Cited

Green, M. C. Genetic Variants and Strains of the Laboratory Mouse (Gustav Fischer Verlag, New York, 1981).
Schlager, G. and M. M. Dickie, Genetics, 57, 319 (1967).

This research was supported by grants from the National Science Foundation (DEB-7814197 {WMF} and DEB-8109904 {WRA}).
The 66 cladistically informative loci examined are: 

**Acf-1** (albumin conformation factor); **Ah-1** (aryl hydrocarbon responsiveness); **Akp-1** (alkaline phosphatase); **Aox-1** (Aldehyde oxidase); **Bgl-e, Bgl-s, Bgl-t** (B-galactosidases); **Car-2** (carbonic anhydrase); **Cdms** (cadmium resistance); **Ce-2** (liver catalase); **Ea-5, Ea-7** (erythrocyte cell surface alloantigens); **Ema** (electrophoretic mobility and agglutinability of red blood cells); **Es-1, Es-3, Es-10** (esterases); **Fv-1, Fv-2** (Friend leukemia virus susceptibility); **Gdc-1** (glycerolphosphate dehydrogenase); **Glk** (galactokinase); **Gpd-1** (glucose-6-phosphate dehydrogenase); **Gpi-1** (glucose phosphate isomerase); **Gus-r, Gus-s** (B-glucuronidase); **Gv-1** (gross virus antigen); **H-1, H-2, H-7, H-13** (histocompatibility complex); **Hba, Hbb** (hemoglobins alpha and beta); **Hc** (hemolytic complement); **Idh-1** (isocitrate dehydrogenase); **If-1** (NDV induced circulating interferon); **Igh-1, Igh-2, Ighsrc, Igk-V** (immunoglobulins); **Lap-1** (leucine arylamino peptidase); **Lv-1** (liver proteins); **Lyt-1, Lyt-2, Lyt-3, Lyr-4** (T-lymphocyte cell alloantigens); **Lyb-2** (B-lymphocyte alloantigen); **Map-1** (mannosidase sialylation); **Mls** (minor mixed lymphocyte stimulating locus); **Mod-1** (malate enzyme); **Mup-1** (major urinary protein); **Pep-3** (peptidase); **Pkg-2** (phosphoglycerate kinase); **Pgm-1** (phosphoglucomutase); **Pre-1, Pre-2** (prealbumins); **Qa-2, Qa-3** (a lymphocyte antigen); **Qed-1** (a lymphocyte antigen); **rd** (retinal degeneration); **Rnr** (rennin regulation); **Sas-1** (serum antigenic substance); **Sep-1** (serum protein); **Slp** (serotonin plasma level); **Ss** (serum serological, controls variability in a complement component); **Syp-1** (seminal vesicle protein); **Tam-1** (tosylarginine methyl esterase); **Tla** (thymus leukemia antigen).

The cladistically non-informative loci examined are: **Ags** (alpha galactosidase); **Amy-1, Amy-2** (amylase); **Apl** (acid phosphatase-kidney); **Apk** (acid phosphatase-liver); **C-3** (complement component-3); **Car-1** (carbonic anhydrase); **Ea-2, Ea-4** (erythrocyte antigen); **Eg** (endoplasmic glucuronidase); **Es-2, Es-6, Es-8, Es-13** (esterase); **Gdr-1, Gdr-2** (G6PD regulator); **Gpt-1** (glutamic-pyruvic transaminase-1); **Got-1, Got-2** (glutamate oxaloacetate transaminase); **Gr-1** (glutathione reductase-1); **Lv** (delta-aminolevulinate dehydratase); **Ly-5** (Lymphocyte antigen-5); **Mph-1** (macrophage antigen-1); **Np-1** (nucleoside phosphorlyase); **Pgd** (6-phosphogluconate dehydrogenase); **Pgm-2** (phosphoglucomutase-2); **Raf** (regulation of alpha-fetoprotein); and **Trf** (transferrin).
Table 1. Minimum pair-wise differences for 97 loci between ten inbred mouse strains. Values below the diagonal are actual differences and those above the diagonal reflect the proportion of differences for those loci known for both members of the pair.

<table>
<thead>
<tr>
<th></th>
<th>C57BL</th>
<th>C57BR</th>
<th>C58</th>
<th>BALB/c</th>
<th>A</th>
<th>AKR</th>
<th>CBA</th>
<th>C3H</th>
<th>DBA/1</th>
<th>DBA/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL</td>
<td></td>
<td>- .23</td>
<td>.27</td>
<td>.41</td>
<td>.41</td>
<td>.51</td>
<td>.51</td>
<td>.51</td>
<td>.52</td>
<td>.54</td>
</tr>
<tr>
<td>C57BR</td>
<td>20</td>
<td>-</td>
<td>.19</td>
<td>.43</td>
<td>.44</td>
<td>.41</td>
<td>.44</td>
<td>.49</td>
<td>.49</td>
<td>.50</td>
</tr>
<tr>
<td>C58</td>
<td>23</td>
<td>15</td>
<td>-</td>
<td>.47</td>
<td>.48</td>
<td>.40</td>
<td>.44</td>
<td>.46</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>BALB/c</td>
<td>38</td>
<td>36</td>
<td>39</td>
<td>-</td>
<td>.24</td>
<td>.50</td>
<td>.46</td>
<td>.41</td>
<td>.37</td>
<td>.45</td>
</tr>
<tr>
<td>A</td>
<td>39</td>
<td>37</td>
<td>40</td>
<td>22</td>
<td>-</td>
<td>.43</td>
<td>.44</td>
<td>.30</td>
<td>.38</td>
<td>.46</td>
</tr>
<tr>
<td>AKR</td>
<td>48</td>
<td>35</td>
<td>34</td>
<td>46</td>
<td>40</td>
<td>-</td>
<td>.31</td>
<td>.34</td>
<td>.35</td>
<td>.39</td>
</tr>
<tr>
<td>CBA</td>
<td>47</td>
<td>36</td>
<td>36</td>
<td>41</td>
<td>40</td>
<td>28</td>
<td>-</td>
<td>.20</td>
<td>.35</td>
<td>.31</td>
</tr>
<tr>
<td>C3H</td>
<td>47</td>
<td>41</td>
<td>38</td>
<td>37</td>
<td>28</td>
<td>31</td>
<td>18</td>
<td>-</td>
<td>.38</td>
<td>.38</td>
</tr>
<tr>
<td>DBA/1</td>
<td>49</td>
<td>41</td>
<td>41</td>
<td>34</td>
<td>35</td>
<td>33</td>
<td>35</td>
<td>-</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>DBA/2</td>
<td>43</td>
<td>36</td>
<td>37</td>
<td>36</td>
<td>37</td>
<td>31</td>
<td>24</td>
<td>30</td>
<td>11</td>
<td>-</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1. Phylogenetic relations of the ten strains of mice examined in this study. The thin lines show periods of brother x sister mating, the box-like lines show periods where random breeding is known (boxes have smooth ends) or is assumed (boxes have jagged ends). Nodes with a (solid) • descend to taxa used to construct Fig. 3.

Figure 2. An example of the phylogeny obtained by parsimony analysis of the 97 loci by the method of Fitch (13). Other methods gave the same topology (but not necessarily the same branch lengths) except for the C57 - C58 group for which all three pairs are expected to have had eight or nine allelic differences in 97 loci at the time of their separation. There is a total of 169 gene substitutions required to explain the divergence but those in the earlier branches are more properly interpreted as changing originally polymorphic loci to different monomorphic loci.

Figure 3. Plot of $d_{ij}$ versus $c_{ij}$ from the equation

$$d_{ij} = c_{ij} h (1 - 2m_{y_{ij}}) + 2m_{y_{ij}}$$

with $y_{ij} = 58$ years. The linear regression gives $d_{ij} = 0.079 (±0.043)(c_{ij}) + 0.165 (±0.029)$. Distances are from Table 1 adjusted to 58 years, the age of the C57 - C58 separation.
Figure 3

Fraction of Loci Different vs. 1 - Coefficient of Kinship

- DBA1/DBA2
- CBA/C3H
- C57BL/C57BR
- C57BR/C58

0.3
0.2
0.1
0
0.25
0.5
0.75
1.0

1 - Coefficient of Kinship
Response to Questions From the Floor

Bill Hohenboken: Was the average number of alleles per locus equal for protein and immune loci?

W. R. Atchley: The 33 immune loci exhibit more diversity. Of the 15 loci with more than 2 alleles among these 10 strains, 11 are related to the immune system.

E. J. Eisen: Does the fact that many of these inbred lines were derived originally because they were susceptible to tumors affect the interpretation of the results?

W. R. Atchley: When these strains were developed, there was considerable interest in tumor phenotypes, allografts, and the like, so some selection was probably practiced in some strains for a few immune system loci. However, it appears that no comparable selection was practiced on the protein loci since there were few, if any, techniques which permitted one to assay protein phenotypes directly. In terms of the effect of interpretation of the results, there is significantly more divergence between strains in the immune loci. However, both protein and immune loci give the same correct branching sequences when tree-building algorithms are employed. Other than the greater divergence in alleles at immune loci, I don't know what other consequences tumor susceptibility might have on these results.

Jack Hill: Would you comment on laboratory mutation rate versus mutation rate in the wild and selection "multiplying" in subsequent generations of favorable mutants in heterozygous conditions?

W. R. Atchley: The best studies on spontaneous and induced mutations involve laboratory animals. The number of individuals involved and the level of experimental control is such that natural population studies of mutation rates are difficult. Hence, I don't know of rigorous data for the comparison you
Gerald Havenstein: The strains that you have used in your study have been selected over the years primarily on single gene traits. Would the same answer with regard to polygenic divergence be obtained if you had used strains that had been selected on polygenic characters, such as body weight and litter size?

W. R. Atchley: Inbred mouse strains probably were selected (intentionally or unintentionally) for polygenic traits during development of the lines. As I stated during my talk, there appears to have been selection for heterozygosity during the initial formation of these strains, and traits such as viability or litter size, may have been involved. Genetic divergence in polygenic traits associated with continuously varying skeletal or life history traits do not give an accurate reflection of the genealogy comparable to that we find with the single gene data.

A. W. Nordskog: Is there anything one could do in terms of genealogical analysis, with, say a chicken population, non-inbred or otherwise, with many years of pedigree information and various types of performance data?

W. R. Atchley: The tree-building algorithms I have described today can be used with a wide variety of types of data. Some of the methods were developed for molecular data but many of the others can analyze a variety of types of data. These types of analyses have been widely applied in evolutionary and systematic biology and there is a rather extensive literature associated with their application. Thus, the types of analyses would depend upon the types of performance data you have, how much divergence there is among the samples, how well-defined the samples are, and the accuracy of the genealogical data.
John Keele: How much of the variability in mandible measurement are you assuming are due to genetics?

W. R. Atchley: In the brief example I gave of divergence between inbred strains in mandible dimensions, all of the divergence is due to genetic causes.