GENETICS OF HEMOGLOBIN IN THE DEER MOUSE, *PEROMYSCUS MANICULATUS*. II. MULTIPLE ALLELES AT REGULATORY LOCI

LEE R. G. SNYDER

Department of Biological Sciences, Stanford University, Stanford, California 94305

AND

Department of Biology, University of California, Riverside, California 92521[1]

Manuscript received April 21, 1977
Revised copy received February 22, 1978

ABSTRACT

Deer mice are polymorphic for electrophoretic hemoglobin phenotypes showing one, two, or three bands. Within the multibanded phenotypes, there is considerable variation in the hemoglobin partitioning, defined as the fraction of total hemoglobin made up by the secondary and tertiary bands. In subspecies *sonoriensis*, for example, hemoglobin partitionings range from 0.03 to 0.38. The inheritance of partitioning values is under remarkably strict genetic control. The genetic variation is additive and the narrow heritability is close to 1.0. The inheritance data can be modeled in precise detail by postulating multiple-allele polymorphisms at globin regulatory loci. Comparison of simulated *versus* actual inheritance data demonstrates that the so-called null structural alleles actually produce functional globins.—The genetic controls in Peromyscus may be analogous to those in primates. Unfortunately, the molecular mechanisms effecting the regulation are unknown. Different subspecies of *P. maniculatus* show strikingly different arrays of partitioning values, but the role of natural selection in maintaining the quantitative polymorphisms remains obscure.

ELECTROPHORETIC surveys of natural populations have provided a wealth of information on qualitative variation involving structural locus polymorphisms, but relatively little information on quantitative variation involving regulatory loci. That lack of information is unfortunate, because variation at regulatory loci is undoubtedly important in evolution. BRITTEN and DAVIDSON (1971) have argued that major events in evolution require the addition of novel patterns of gene regulation or the reorganization of existing regulatory patterns. WILSON and his colleagues have contended that in taxa such as the mammals, the predominant evolutionary changes have involved genetic differences accruing at the regulatory, rather than the structural, loci (WILSON, MAXON and SARICH 1974; WILSON, CARLSON and WHITE 1977; KING and WILSON 1975; PRAGER and WILSON 1975).

[1] Present address.

Complex hemoglobin polymorphisms in *P. maniculatus* (SNYDER 1978) provide a case where electrophoretic surveys can be used to draw conclusions about regulatory polymorphisms. Within a general class of qualitative phenotypes, there is an almost continuous range of quantitative variation in the relative proportions of hemoglobin components. Understanding the source of the quantitative variation is critical to evaluating the possible effects of natural selection on the polymorphic hemoglobin loci. Because selection acts on the phenotypic level, the response to selection will depend upon the extent to which the phenotypic variation stems from underlying genetic variation.

The genetic analysis reported here demonstrates that the quantitative variation in hemoglobin phenotypes represents additive genetic variation, with minimal environmental effects. The quantitative inheritance data can be modeled quite precisely in terms of multiple alleles at four globin regulatory loci. If the models are correct, then the so-called null structural alleles (RASMUSSEN, JENSEN and KOEHN 1968; JENSEN 1969) actually produce functional globin subunits. Finally, population surveys show that four subspecies of deer mice are strikingly differentiated in their quantitative phenotypes and hence must be polymorphic for different arrays of regulatory elements.

MATERIALS AND METHODS

The deer mice used in the inheritance studies were collected from four general areas and represented four subspecies: *Peromyscus maniculatus bairdii* from the great plains of eastern South Dakota, *P. m. rufinus* from the Rocky Mountains of central Colorado, *P. m. sonoriensis* from the Sierra Nevada and White Mountains of eastern California, and *P. m. gambelii* from the Central Valley of California. Detailed information on the trapping sites is given by SNYDER (1977).

Polyacrylamide gel electrophoresis was performed on fresh hemolysates, using the discontinuous buffer system of DAVIS (1964), with minor modifications as described in the accompanying paper (SNYDER 1978). The tube gels were scanned on a Gilford model 240 spectrophotometer. The relative proportions of the different hemoglobin components were determined by measuring areas on the gel scans with a Keuffel and Esser compensating polar planimeter.

A series of preliminary tests were performed to gauge the accuracy of the gel-scanning techniques. Optimal resolution of the hemoglobin bands was obtained by using 1 mg of hemoglobin per gel tube and then scanning the gel at 485 nm. Ideally one would use much lower concentrations of hemoglobin and then scan at the intensely absorbing heme wavelength of 416 nm. That procedure was not possible because the deer mouse hemoglobins partially dissociated into dimers when electrophoresed at low concentrations. However, absorbance spectra obtained from isolated hemoglobin components from several strains of deer mice indicated that all the hemoglobins had the same relative absorbances at 416 and 485 nm. Hence, the wavelength 485 nm could be used to determine the relative amounts of hemoglobin.

The linearity of the densitometry was tested by running the same blood samples at different hemoglobin concentrations and then comparing the measured proportions of the multiple hemoglobin bands. The proportions for any one sample were effectively constant over a two-fold range of hemoglobin concentrations. The combined data from all blood samples revealed no substantial deviations from linearity.

Finally, corrections were made for a slight but consistent overlap between hemoglobin bands. Scans of single-banded hemoglobin phenotypes showed a long, faint leading (anodal) edge, apparently due to a slow but continuous dissociation of hemoglobin tetramers into dimers. The

leading edge constituted only about 6% of the major band (hemoglobin component Hb4), but in multiple-banded phenotypes it augmented considerably the density of the fainter, more anodal hemoglobin components (Hb3 and Hb2). To correct for the overlap, the Hb4 peak on each multibanded phenotype was matched in density with the Hb4 peak of a single-banded phenotype. The two gel scans were superimposed and the leading edge of the Hb4 peak on the single-banded scan was traced onto the multi-banded scan, thus revealing the amount of overlap between Hb4 and Hb3. A similar matching procedure between double-banded and triple-banded scans was used to correct for the overlap between Hb3 and Hb2.

A more extensive description of the gel-scanning and correction procedures is given by SNYDER (1977).

RESULTS

The qualitative genetics of hemoglobin and the molecular structures of the various electrophoretic components are the topic of the accompanying paper (SNYDER 1978). For ease of understanding, the phenotypes, genotypes, and molecular compositions will be reviewed briefly here. Next, models will be developed to account for the quantitative variation in hemoglobin phenotypes. Finally, the predictions of the models will be compared to actual data on quantitative inheritance.

Structural-locus polymorphisms

The deer mice used in the quantitative inheritance studies showed four qualitatively different phenotypes involving three electrophoretically distinct hemoglobin components. Phenotype P1 has a single hemoglobin component, Hb4 ($\alpha_2\beta_2$). Phenotype P2 has Hb4 plus a "weak" second component, Hb3. Depending on genotype, Hb3 consists of $\alpha_2\delta_2$ tetramers and/or $\gamma_2\beta_2$ tetramers. P3, the "strong" double-banded phenotype, shows approximately twice the amount of Hb3 as does P2. P4 is triple banded; the faint third band, Hb2, consists of $\gamma_2\delta_2$ tetramers.

The four phenotypes are generated by alleles at four globin structural loci. Loci *Hba* and *Hbb* are monomorphic, with alleles Hba^1 and Hbb^1 coding for α and β globin subunits, respectively. Loci *Hbc* and *Hbd* are each polymorphic for two alleles. Hbc^1 and Hbd^1 code for subunits γ and δ, respectively. Hbc^0 and Hbd^0 are so-called null alleles; either they produce no functional globins or they code for subunits indistinguishable from α and β, respectively. One of the purposes of this paper is to demonstrate that the "null" alleles do in fact produce functional globin subunits.

Quantitative variation in hemoglobin partitioning

Structural-locus heterozygotes (phenotype P2, genotypes $c^1c^0d^0d^0$ or $c^0c^0d^1d^0$) show a relatively weak Hb3 band, due to globin production by a single c^1 or d^1 structural allele. Structural-locus homozygotes (phenotype P3, genotypes $c^1c^1d^0d^0$ or $c^0c^0d^1d^1$) have a relatively dense Hb3 band, due to globin production by two c^1 or d^1 structural alleles. The double heterozygote (phenotype P4, genotype $c^1c^0d^1d^0$) has a relatively dense Hb3 band ($\alpha_2\delta_2$ plus $\gamma_2\beta_2$ tetramers) plus a faint Hb2 band ($\gamma_2\delta_2$ tetramers). However, within those genotypic classes there is a broad range of variation in the relative densities of the hemoglobin components.

The quantitative variation is conveniently described in terms of the *hemoglobin partitioning*, defined as the fraction of total hemoglobin comprised by Hb3 plus Hb2. The term derives from the fact that deer mice have the same average hemoglobin concentration (approximately 13.4 gm Hb/100 ml blood) regardles of phenotype, but depending upon genotype that total is partitioned into one, two, or three electrophoretic components. Within a subspecies, the structural-locus heterozygotes and homozygotes can usually be distinguished on the basis of partitioning values (Figure 1). For instance, in subspecies *sonoriensis*, breeding tests have indicated that structural-locus heterozygotes have partitionings ranging from 3 to 19%, while partitionings for structural-locus homozygotes generally range from 21 to 38%. However, there is some phenotypic overlap between the genotypic classes, because one *sonoriensis* individual with a partitioning of 18% proved to be a structural-locus homozygote.

The ranges of partitionings and the cut-off points between structural-locus heterozygotes and homozygotes varied markedly between subspecies. In subspecies *rufinus*, the partitionings ranged from 8 to 15% for P2 heterozygotes and 17 to 25% for P3 homozygotes. In subspecies *bairdii* and *gambelii*, the wild-caught

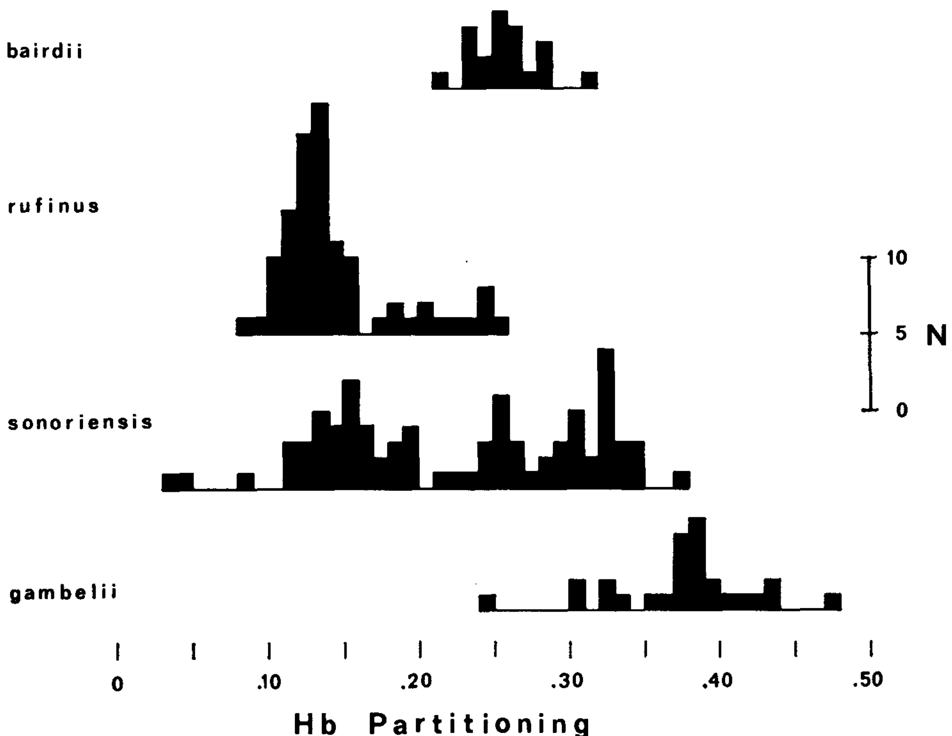

FIGURE 1.—Hemoglobin partitioning values in four subspecies of *P. maniculatus*. The data are for P2 and P3 partitionings only; the P1 phenotypes, which comprise 60% of *rufinus* and 13% of *sonoriensis* phenotypes, and the P4 phenotypes, which comprise 2–4% of the phenotypes in both subspecies, are not shown.

mice were all P3 homozygotes, with partitioning ranges of 21 to 31% and 24 to 47%, respectively. The average P3 partitionings were compared using t-tests between pairs of subspecies having the most similar ranges of partitionings. All three comparisons (*rufinus-bairdii*, *bairdii-sonoriensis* and *sonoriensis-gambelii*) revealed highly significant differences between the means ($P < 0.001$). The mean P2 partitioning values for *rufinus* and *sonoriensis* were also significantly different ($P < 0.001$).

Models of quantitative inheritance

A simple explanation for the quantitative variation is that the amount of globin produced by each structural locus is determined by a closely linked regulatory locus polymorphic for multiple alleles. In the models that follow, the globin production specified by each regulatory allele is assumed to be constant, with no feedback or interactions between loci. Hence, the regulatory alleles have a strictly additive effect.

A. *Active-alleles model*: The first model assumes that structural alleles c^0 and d^0 are active alleles producing α- and β-globin subunits, respectively. Consider a P2 × P2 mating in which the parental regulatory genotypes are $a^1_1a^1_2b^1_1b^1_2c^1_1c^0_1d^0_1d^0_2$ and $a^1_3a^1_4b^1_3b^1_4c^1_3c^0_3d^0_3d^0_4$. The superscripts denote the type of structural allele associated with each regulatory allele, and the subscripts designate individual regulators, each specifying a characteristic amount of globin production. (Hereafter the symbols will be used interchangeably to denote both the genotypes and the relative amounts of globin production.) In a P2 phenotype, the total α production is controlled by three regulators (two a^1 regulators plus the c^0 regulator), the γ production is controlled by one regulator (c^1), and the β production is controlled by four regulators (two b^1 plus two d^0 regulators). (The deer mouse genome also includes two *Hbe*^{0} structural alleles, which contribute about 4% of the total β-type production (SNYDER 1978). For the sake of simplicity, in the quantitative models presented here, the e^0 β-globin production is implicitly included as part of the b^1 β-globin production.) It is assumed that the globin subunits associate randomly to form heterodimers and then tetramers. Hence the partitioning value of the first P2 parent is given by:

$$X_1 = \frac{[\text{Hb3}]}{[\text{Hb4}] + [\text{Hb3}]} = \frac{[\gamma_2\beta_2]}{[\alpha_2\beta_2] + [\gamma_2\beta_2]}$$

$$= \frac{(c^1_1)(b^1_1+b^1_2+d^0_1+d^0_2)}{[(a^1_1+a^1_2+c^0_1)+c^1_1](b^1_1+b^1_2+d^0_1+d^0_2)}$$

$$= \frac{c^1_1}{a^1_1+a^1_2+c^0_1+c^1_1} \quad . \tag{1}$$

Similarly, the partitioning of the second P2 parent is

$$X_2 = \frac{c^1_3}{a^1_3+a^1_4+c^0_3+c^1_3} \quad . \tag{2}$$

In effect, the α and γ subunits are partitioning the total β pool into two hemoglobin components, Hb3 and Hb4, and X will depend only on the relative concentrations of α and γ subunits. Likewise, the partitioning of a P3 offspring from the X_1 and X_2 parents is given by, say:

$$Y = \frac{c^1_1 + c^1_3}{a^1_1 + a^1_3 + c^1_1 + c^1_3} . \tag{3}$$

It is assumed that the *Hba* and *Hbc* gene complexes segregate independently (see DISCUSSION). Therefore, the P3 sibs from a particular P2 × P2 mating will necessarily be identical for the two c^1 regulators, but they will be heterogeneous for various combinations of the four parental a^1 regulators.

The P2 phenotype can also arise from structural-allele polymorphism at the *Hbd* rather than the *Hbc* locus. In a mouse of genotype $a^1_5 a^1_6 b^1_5 b^1_6 c^0_5 c^0_6 d^1_5 d^0_5$, the β and δ subunits partition the α-globin pool, and the partitioning value is given by:

$$\begin{aligned} X_3 &= \frac{\alpha_2\delta_2}{[\alpha_2\beta_2] + [\alpha_2\delta_2]} \\ &= \frac{(d^1_5)(a^1_5 + a^1_6 + c^0_5 + c^0_6)}{[(b^1_5 + b^1_6 + d^0_5) + d^1_5](a^1_5 + a^1_6 + c^0_5 + c^0_6)} \\ &= \frac{d^1_5}{b^1_5 + b^1_6 + d^0_5 + d^1_5} . \end{aligned} \tag{4}$$

In a P2 × P2 mating involving the X_1 and X_3 parents, one-fourth of the offspring will inherit both the c^1 and the d^1 structural alleles and hence will show the triple-banded P4 phenotype. The expression for the P4 partitioning (symbolized Z) is slightly more complicated because of the random association of all four globin types. A typical P4 offspring partitioning would be:

$$\begin{aligned} Z &= \frac{[\mathrm{Hb3}] + [\mathrm{Hb2}]}{[\mathrm{Hb4}] + [\mathrm{Hb3}] + [\mathrm{Hb2}]} = \frac{[\alpha_2\delta_2] + [\gamma_2\beta_2] + [\gamma_2\delta_2]}{[\alpha_2\beta_2] + [\alpha_2\delta_2] + [\gamma_2\beta_2] + [\gamma_2\delta_2]} \\ &= \frac{(a^1_1 + a^1_5 + c^0_5)(d^1_5) + (c^1_1)(b^1_1 + b^1_5 + d^0_1) + (c^1_1)(d^1_5)}{(a^1_1 + a^1_5 + c^0_5)(b^1_1 + b^1_5 + d^0_1) + (a^1_1 + a^1_5 + c^0_5)(d^1_5) + (c^1_1)(b^1_1 + b^1_5 + d^0_1) + (c^1_1)(d^1_5)} , \end{aligned} \tag{5}$$

where a^1_1; b^1_1, c^1_1, and d^0_1 are the alleles contributed by the X_1 parent, and a^1_5, b^1_5, c^0_5; and d^1_5 are from the X_3 parent. Again, it is assumed that the *Hbb* and *Hbd* gene complexes segregate independently. Therefore, the P4 sibs from a given mating will necessarily be identical for alleles c^1_1, c^0_5, d^1_5, and d^0_1, but will be heterogeneous for four combinations of the other regulatory alleles, a^1_1, a^1_5, b^1_1, and b^1_5.

B. *Null-alleles model*: The second model assumes that the c^0 and d^0 alleles

produce no functional globin. Hence, the production of α and β globins is controlled solely by the a^1 and b^1 regulators, and the c^0 and d^0 regulators do not enter into the equations. The corresponding partitioning formulae are:

$$X_1 = \frac{(c_1^1)(b_1^1+b_2^1)}{[(a_1^1+a_2^1)+c_1^1](b_1^1+b_2^1)} = \frac{c_1^1}{a_1^1+a_2^1+c_1^1} \tag{6}$$

$$X_3 = \frac{(d_5^1)(a_5^1+a_6^1)}{[(b_5^1+b_6^1)+d_5^1](a_5^1+a_6^1)} = \frac{d_5^1}{b_5^1+b_6^1+d_5^1} \tag{7}$$

$$Y = \frac{c_1^1+c_3^1}{a_1^1+a_3^1+c_1^1+c_3^1} \tag{8}$$

$$Z = \frac{(a_1^1+a_5^1)(d_5^1)+(c_1^1)(b_1^1+b_5^1)+(c_1^1)(d_5^1)}{(a_1^1+a_5^1)(b_1^1+b_5^1)+(a_1^1+a_5^1)(d_5^1)+(c_1^1)(b_1^1+b_5^1)+(c_1^1)(d_5^1)} \,. \tag{9}$$

Numerical simulations of quantitative inheritance

The object of the quantitative models is to predict offspring partitionings as functions of the parental partitionings Given the large number of variables in equations (1)–(9) and the complexities of genetic transmission, it is impossible to derive explicit functions for Y and Z in terms of the X_i. One could predict the offspring partitionings if the parental regulatory allele values could be determined directly, but that too is impossible. One can, however, estimate the arrays of regulatory alleles present in various populations and then derive approximate expectations for the quantitative inheritance.

The model arrays used in this study were based on the following assumptions: (1) there are 15 or more distinct alleles at each regulatory locus, (2) there is no linkage disequilibrium between regulatory and structural alleles within each of the four *Hb* gene complexes, (3) the globin production values of the alleles differ in approximately step-wise fashion, (4) the distribution of allele frequencies over allele production values is quasi-normal, and (5) the allele frequency distributions are similar for the *Hba* and *Hbb* loci and for the *Hbc* and *Hbd* loci. Under those assumptions, the mean and variance of the *Hbc* and *Hbd* regulators can be estimated from the distribution of P2 phenotypes observed in natural populations. Given quasi-normal distributions for both the allele values and P2 partitionings, then a mouse with the mean partitioning value will most likely possess the high-frequency, mean-value regulatory alleles. Hence the values of the mean alleles can be estimated by substitution in equation (1):

$$\bar{X} = \frac{\bar{c}^1}{2\bar{a}^1+\bar{c}^0+\bar{c}^1} \,,$$

where $\bar{X}$ is the mean P2 partitioning value observed in a wild-caught sample and

$\bar{a}^1, \bar{c}^0$ and $\bar{c}^1$ are the high-frequency, mean-value regulatory alleles. Note that $\bar{c}^0 = \bar{c}^1 = \bar{c}$, because there is no linkage disequilibrium between regulatory and structural alleles, and that $\bar{a}^1$ can be arbitrarily set to equal $(\frac{1}{2} - \bar{c})$. By substition and rearrangement,

$$\bar{c} = \bar{X} \ . \tag{10}$$

From population samples of subspecies *sonoriensis*, $\bar{c}$ was estimated to be 0.16. Next, the variance in allele values was adjusted to fit the observed variance in partitionings. Figure 2 shows the actual P2 partitionings in a population of *sonoriensis*, compared with simulated partitionings generated by arrays of 21 alleles at each of the four regulatory loci. The *Hbc* and *Hbd* arrays had a mean value of 0.16 and a standard deviation of 0.03. The *Hba* and *Hbb* allele values were derived arbitrarily by adding the constant 0.18 to the corresponding *Hbc* or *Hbd* allele value, yielding a distribution with a mean of 0.34, but with the same standard deviation. The fit between observed and simulated partitionings was acceptably close, so that those arrays of alleles could then be used for simulations of quantitative inheritance.

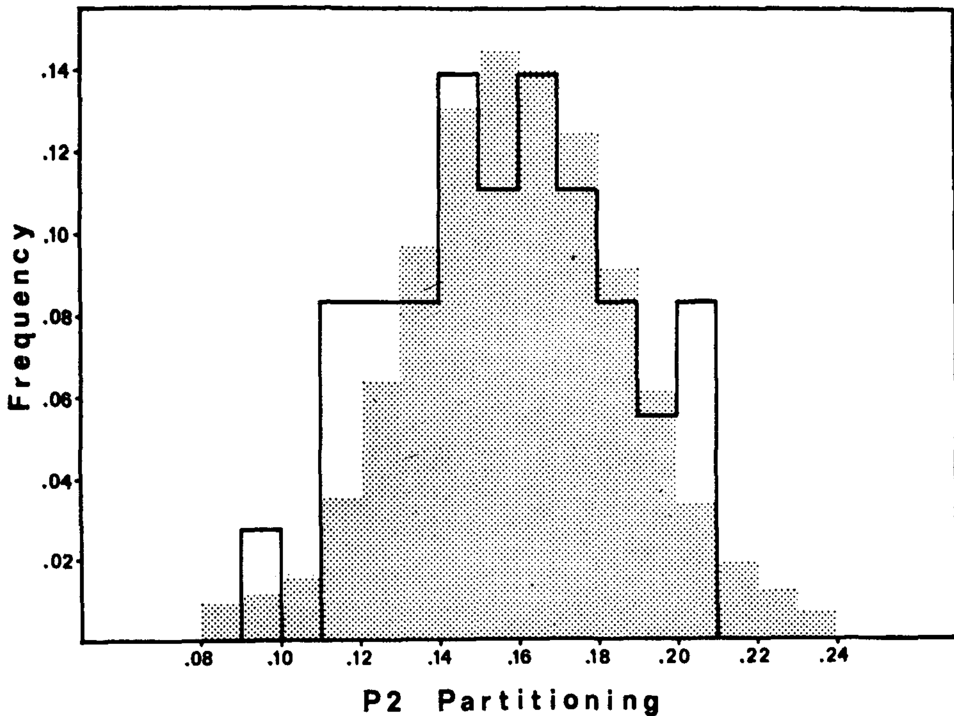

FIGURE 2.—Simulated *versus* observed partitionings of P2 phenotypes in *P. m. sonoriensis*. Solid histogram: observed partitionings in a sample of 35 wild-caught adults. Stippled histogram: simulated partitionings in a sample of 10,000 individuals.

A typical simulation program utilized the allele frequency distributions and formulae (1) through (9) to generate expected values of Y and Z, given $(X_1 + X_2)$ or $(X_1 + X_3)$. The program chose allele values for the three P2 parents at random from the "population" arrays, then performed the two types of "matings" and selected single P3 and P4 offspring. The Y and Z values were accumulated in registers labeled $(X_1 + X_2)$ or $(X_1 + X_3)$. After, say, 10,000 matings the contents of each $(X_1 + X_2)$ or $(X_1 + X_3)$ register were processed to obtain the mean and standard deviation of the Y or Z values. The standard deviation of the partitionings in each parental register was multiplied by the appropriate value of Student's t-statistic, to establish approximate 95% confidence limits for individual values of Y or Z. The end result of the simulation was an expected mean and approximate confidence limits for Y or Z, conditional upon the sum of the parental partitionings. Hence, the simulation results could be compared directly with the data from individual laboratory crosses.

It should be emphasized that the simulations were not simple exercises in curve-fitting. The regulatory allele values were estimated for each subspecies from the array of P2 partitionings observed in sample of wild-caught mice, and the simulations depended only upon those values. No attempt was made to adjust the allele arrays to achieve a closer fit to the actual inheritance data.

Parent-offspring correlation for P2 × P1 crosses

The first test of the quantitative models involved crosses of double-banded P2 mice to single-banded P1 mice. Half the progeny of such a cross will show the P2 phenotype, having received the c^1 or d^1 structural allele from the P2 parent. If the c^1 and d^1 regulators are predominant in determining the partitioning, as is suggested by the partitioning formulae, then there should be a close correlation between the partitioning values of a P2 parent and its P2 offspring. The inheritance data are shown in Figure 3, where each data point represents a P2 parent compared with a single P2 offspring chosen at random from its progeny. The matings involved noninbred pairs from subspecies *rufinus* and *sonoriensis*, plus several *gambelii* × *sonoriensis* hybrid pairs. The correspondence between parent and offspring was very close over the entire range of partitionings. The intraclass correlation coefficient, expressing the ratio of between-pair variance to the total variance, was 0.986. In other words, only 1.4% of the total phenotypic variance in the sample could be ascribed to the joint effects of measurement error, environmental variables, and/or residual genetic differences. The close correlation was not due simply to maternal effects, because most of the P2 parents were male.

The relative magnitudes of measurement error and possible environmental effects were evaluated by re-testing P2 parents and groups of P2 offspring at three-month intervals, and by replicating each step in the measurement procedures. The nested analyses of variance are given by SNYDER (1977). The conclusion was that approximately 50% of the variance between P2 relatives was due to measurement error during electrophoresis and planimetry. Importantly, the quantitative models indicate that the remaining variance between parents and offspring can be attributed to predictable genetic differences. For instance,

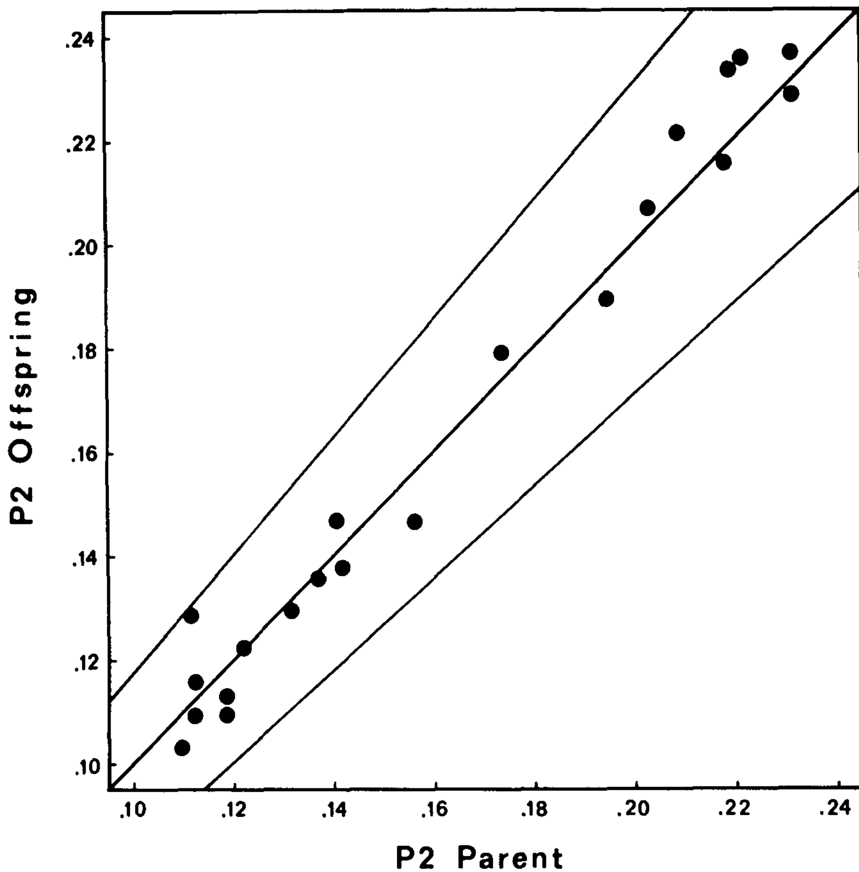

FIGURE 3.—Correlation of partitioning values between P2 parents and P2 offspring from P2 × P1 crosses. Each point is based on a single P2 offspring chosen at random from the progeny. The partitioning of each mouse was determined from three replicate gels, and each gel scan was measured three times. Parent and offspring were electrophoresed concurrently, to further reduce the experimental error. The solid lines are the expected mean and approximate 95% confidence limits for individual data points, generated by numerical simulations of the active-alleles model.

if the P2 parent is an *Hbc*-locus heterozygote and if the c^0 allele is active, then the parent and offspring will share only the c^1 regulator and one a^1 regulator; they will carry different c^0 regulators and second a^1 regulators. The expected variance due solely to segregation of c^0 and a^1 regulators (or of d^0 and b^1 regulators, in *Hbd*-locus heterozygotes) is indicated by the solid lines in Figure 3, which represent the expected mean bracketed by approximate 95% confidence limits for individual data points. Because all the data points fall within the confidence limits, there is no need to postulate environmental factors to explain the residual variance.

Partitioning values of P3 and P4 offspring

In P2 × P2 matings of parents with similar genotypes (either $c^1c^0d^0d^0$ or $c^0c^0d^1d^0$), one-fourth of the offspring will show the P3 phenotype (corresponding to genotypes $c^1c^1d^0d^0$ or $c^0c^0d^1d^1$). Offspring partitioning values were obtained from 15 such matings in which the parents were randomly chosen, noninbred pairs from the same subspecies. Figure 4 shows the data for ten pairs of *P. m. sonoriensis*, along with the expected mean and confidence limits from simulations based on equations (1)–(3). The offspring partitioning (Y) was not a simple linear function of ($X_1 + X_2$); rather, the expected function was slightly curvilinear. Two data points fell outside the approximate confidence limits, but that is not surprising because those intervals are based simply on additive genetic effects; they do not compensate for the measurement error in the sums,

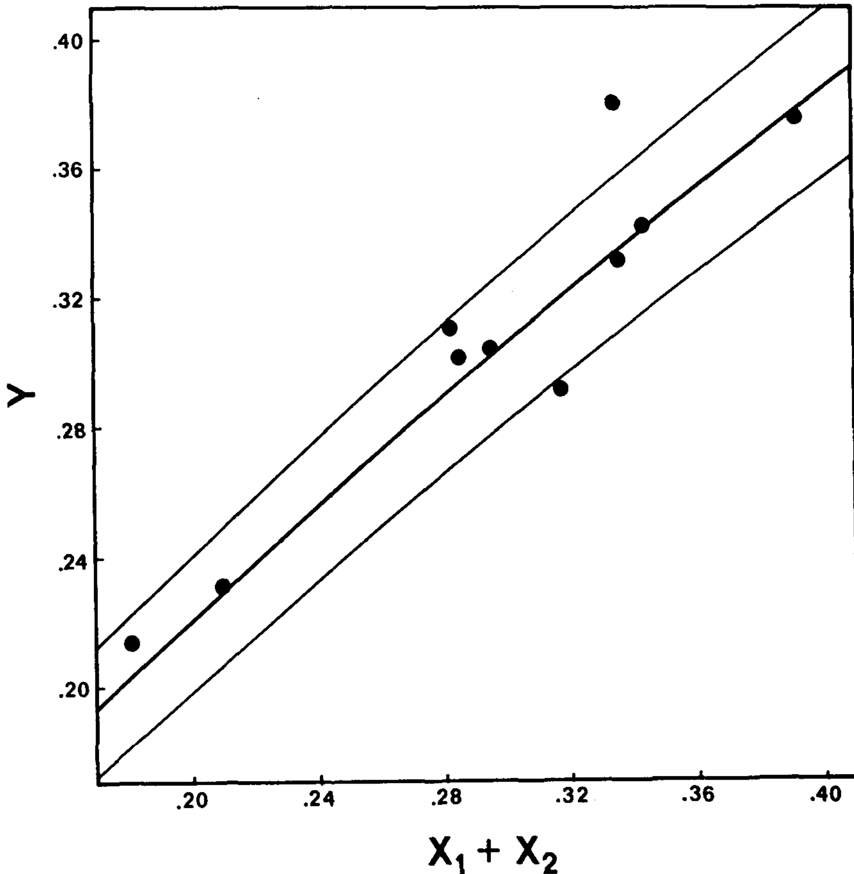

FIGURE 4.—Partitioning values for P3 offspring (Y) *versus* the sum of P2 parental partitionings ($X_1 + X_2$) in *P. m. sonoriensis*. Each data point is based on a single P3 offspring chosen at random from the P3 progeny. The solid lines represent the expected mean and approximate 95% confidence limits for individual data points, based on the active-alleles model.

$(X_1 + X_2)$. Therefore the fit between inheritance data and simulated expectations was once again remarkably close.

Offspring partitioning data were also obtained from eight P2 × P2 matings within subspecies *rufinus*. In five pairs, the parents were heterozygous for the same structural locus and yielded P3 offspring; in three pairs, the parents were heterozygous for unlike structural loci and produced triple-banded, P4 progeny. (Genotypically, the latter matings were of type $c^1c^0d^0d^0 \times c^0c^0d^1d^0$, and one-fourth of the progeny were double heterozygotes, $c^1c^0d^1d^0$, with phenotype P4). Inheritance data on single offspring from those eight crosses were compared against expectations for P3 or P4 phenotypes, based on arrays of alleles estimated for *rufinus* using equation (10), with $\bar{X} = 0.13$. The plots (not shown) indicated that the data matched the expectations quite closely, with all eight data points falling well within the confidence limits.

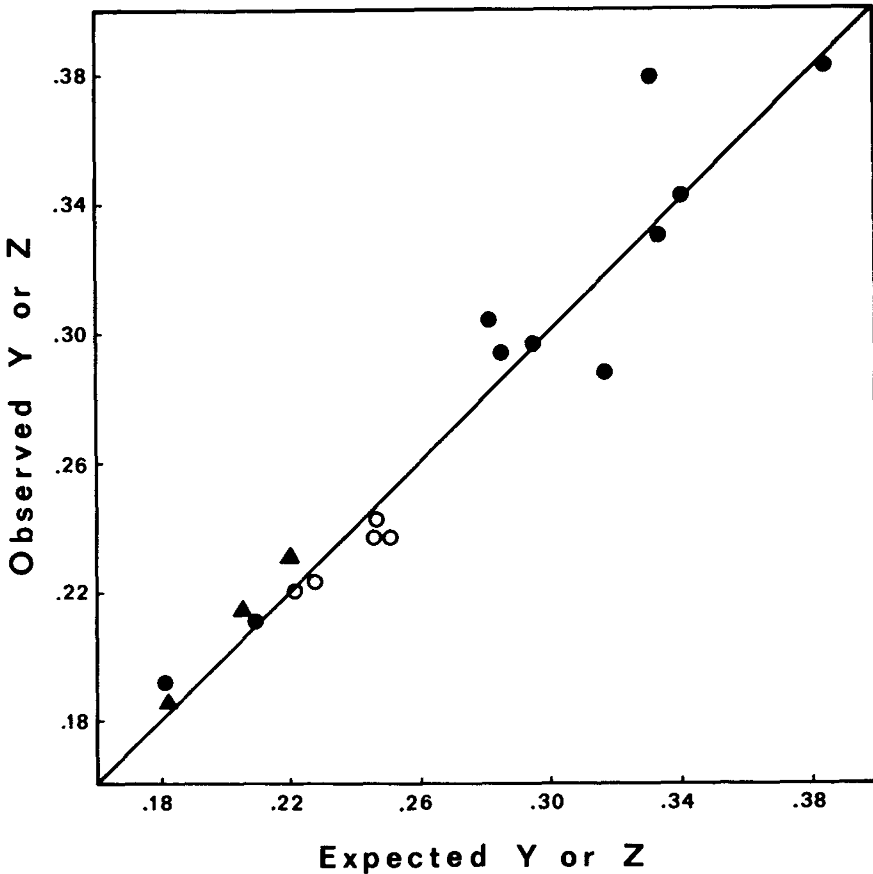

FIGURE 5.—Observed partitionings of P3 and P4 offspring *versus* expectations from the active-alleles model. Closed circles: P3 offspring from *sonoriensis*; open circles: P3 offspring from *rufinus*; closed triangles: P4 offspring from *rufinus*.

Active- versus *null-allele models*

The quantitative inheritance data have resolved the question of whether the c^0 and d^0 structural alleles are true null alleles or whether they produce α- and β-type globin subunits, respectively. The combined data on P3 and P4 offspring partitionings are shown in Figures 5 and 6, with observed values plotted against expectations generated by the active- *versus* null-allele models, respectively. The fit to the active-alleles model is quite close, with ten observations falling above the expected values and eight observations falling below. By contrast, with the null-alleles model, the observed values exceed the expectations by an average of 15% and all 18 points fall above the line.

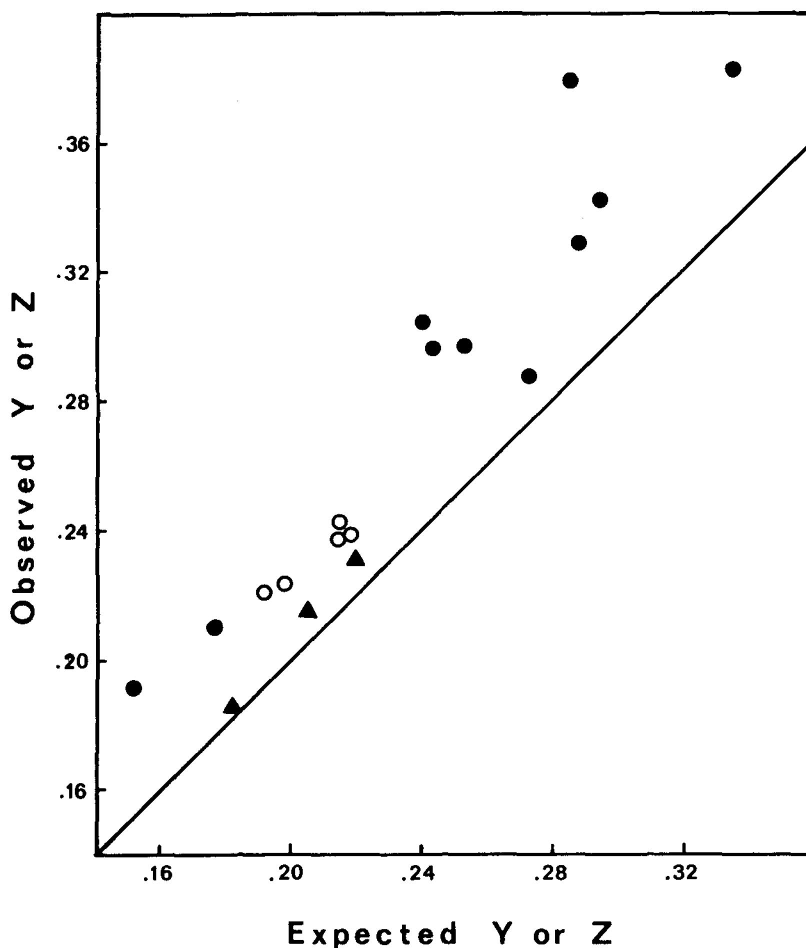

FIGURE 6.—Observed partitionings of P3 and P4 offspring *versus* expectations from the null-alleles model. Symbols are the same as in Figure 5.

Narrow heritability of partitioning values

An important parameter in quantitative inheritance is the narrow heritability, h^2, defined as the proportion of total phenotypic variation that can be attributed to additive genetic variance (FALCONER 1960). The narrow heritability is commonly estimated as the slope of the regression of offspring values on midparental values. If all the variation stems from loci with strictly additive effects, then the regression coefficient and the narrow heritability reach the maximum value of 1.0.

The quantitative inheritance data indicate that the narrow heritability of partitioning values is very high. Unfortunately, it is not possible to calculate an unbiased estimate of heritability from the data *per se*, because the laboratory studies did not include all the types of matings possible in natural populations. (For instance, there are no data on P3 × P2 matings, a common type in *sonoriensis* populations.) However, an expected narrow heritability for a population could be calculated from the same simulations used so successfully to predict the results from single matings. The simulations were based on the same arrays of regulatory alleles used previously, plus various frequencies of the c^0 and d^0 structural alleles. Simulated matings were made at random, and for each mating the mean offspring phenotype was compared to the midparent phenotype. In all cases the offspring-midparent regression coefficient was very close to, and statistically indistinguishable from, 1.0.

Minimum number of regulatory alleles

A number of distinct regulatory alleles have been "genetically isolated" in laboratory strains. Each strain was initiated by mating a wild-caught P2 mouse to one or more wild-caught P1 mice. All the P2 offspring from such matings necessarily carry c^1 or d^1 regulatory alleles that are identical by descent—that is, they are exact replicates of the single c^1 or d^1 regulatory allele carried by the P2 parent. Each strain has been propagated by mating the P2 offspring to unrelated P1 mice, so that in each generation all the P2 mice remain identical-by-descent for the c^1 or d^1 regulator. The strains are useful because they indicate the range of partitioning values produced by a single c^1 or d^1 regulator in combination with randomly chosen a^1, b^1, c^0, and d^0 regulators. Data accumulated over several generations indicate that, in an average sample of identical-by-descent phenotypes, the standard deviation in partitioning value is approximately 3% of the mean. That statistic allows a rough calculation of the minimum number of distinct regulatory alleles present in natural populations.

For instance, in *sonoriensis* populations, the P2 phenotypes are apparently produced by d^1 structural alleles (SNYDER 1978). The partitioning values are continuously distributed between approximately 0.09 and 0.19, with scattered values down to 0.03 (Figure 1 and unpublished data). One can divide that range into nonoverlapping intervals consisting of the mean partitioning value, plus or minus two standard deviations. (The choice of interval $\pm 2s$ is a statistical convention—in a normal distribution, 95.4% of the individual values will fall

within two standard deviations of the mean.) Each interval can be visualized as the space occupied by a distinct allele; that is, it represents the range of partitioning values assumed by a distinct d^1 regulator arrayed on different genetic backgrounds.

That sort of reasoning indicates that the minimum number of distinct d^1 regulatory alleles in *sonoriensis* populations is ten. Obviously that estimate is conservative because it is based on discrete intervals between allele values. In reality the values of regulatory alleles may vary almost continuously, and the actual number of alleles may be considerably larger than ten.

DISCUSSION

There is an increasing number of cases in which genetic polymorphisms are known to affect quantitative variation in the levels of proteins or enzymes. Different allozymes for human red cell acid phosphatase have characteristic effects on maximal enzyme activity (HARRIS 1975). The alcohol dehydrogenase loci in maize have multiple alleles that affect the relative amounts of enzyme synthesized (SCHWARTZ 1976). In house mice, the synthesis of β-galactosidase is affected by what appears to be a regulatory locus closely linked to the structural locus (BREEN, LUSIS and PAIGEN 1977; LI and DANIEL 1976). Also in house mice, the activity levels of renin (WILSON *et al* 1977) and glucokinase (COLEMAN 1977) are affected by alleles at single genetic loci. Alcohol dehydrogenase activity in *Drosophila melanogaster* is apparently affected by modifier loci that are separate from, but loosely linked to, the structural loci (WARD 1975; BARNES and BIRLEY 1975). In larvae of the dipteran *Chironomus tentans*, the presence of a mutant regulatory allele alters the quantities of several hemoglobin isoforms (THOMPSON and PATEL 1972; THOMPSON and HORNING 1973). Human populations in tropical areas historically plagued by malaria are usually polymorphic for various thalassemia alleles that reduce the production of either α or β globins (WEATHERALL and CLEGG 1972).

In several cases it appears that quantitative differences in protein or enzyme levels can be ascribed to variation in the number of structural loci present. There is evidence that human populations are polymorphic for the number of α-globin structural loci, such that some individuals have two rather than one α locus (RUCKNAGEL and WINTER 1974; NUTE 1974; BAINE *et al.* 1976). In pigs, the activity of serum arylesterase may be controlled by five multiple alleles at a single regulatory locus (AUGUSTINSSON and OLSSON 1961); alternatively, since the enzyme activity levels occur in integer multiples of the weakest allele, the genetic control may stem from multiple duplications of the structural locus. Similarly, variation in levels of salivary amylase in bank voles (*Clethrionomys glareola*) is probably due to chromosomes carrying one, two, or three structural loci (NIELSEN 1977). A polymorphism for structural-locus duplications might also explain integer differences in the rate of synthesis of aminolevulinate dehydratase in inbred strains of house mice (COLEMAN 1971).

In contrast to the cases just mentioned, the data presented here demonstrate

precise, fine-scale genetic control of a trait that varies continuously, rather than in integral steps. The continuous nature of the variation rules out a simple explanation based on variable numbers of duplicated structural loci. The hemoglobin partitionings do not occur in coarse multiples of 5% or 10% of total hemoglobin. Instead, the phenotypic differences between distinct regulatory alleles are on the order of 1% of total hemoglobin. The range of phenotypes in subspecies *sonoriensis* indicates that there are ten or more distinct alleles controlling the globin production of the *Hbd* structural locus. If the quantitative variation were due simply to multiple duplications, then some chromosomes would have to carry ten or more copies of the *Hbd* structural locus. There is no evidence in any mammal for such extensive duplication of globin loci.

It has been assumed that the quantitative variation stems from multiple regulatory alleles, such that the globin production of an individual structural locus is controlled by the allele at its closely adjacent regulatory locus. Alternatively the variation could stem from a multiplicity of structural alleles *per se*, with each structural allele having an inherent and characteristic rate of globin synthesis. (The formulation and conclusions of the quantitative models remain the same under either assumption). It is quite plausible that the amino acid sequence or corresponding nucleotide sequence of a globin subunit might control its rate of synthesis (HARRIS 1975; LODISH 1976; GOLDBERGER 1974; SMITH 1975). However, in the present case there would have to be arrays of ten or more allelic globins, each with a sequence substitution that affects its rate of synthesis, but not its electrophoretic properties. If nucleotide or amino acid substitutions were frequent enough to account for differences in globin production, then one would expect to see a greater range of electrophoretic mobilities. Instead, with the exception of the relatively rare structural allele reported by MAYBANK and DAWSON (1976), all the globin subunits in *P. maniculatus* populations fall into just two electrophoretic mobility classes (SNYDER 1978).

Theoretically, one could demonstrate the existence of adjacent regulatory and structural loci by detecting recombination between them, but with eucaryotes such fine-structure mapping is usually impossible. In a rare case, CHOVNIK *et al.* (1976) were able to screen several million Drosophila progeny and thereby detect recombination between distinct regulatory and structural components of the rosy locus (coding for xanthine dehydrogenase). Such mapping is obviously impossible with mice.

The Peromyscus data are interesting because there may be analogous genetic control of hemoglobin proportions in primates. The hemoglobin phenotypes in several species of Macaque monkeys are remarkably similar to Peromyscus phenotypes, in that there is an α-globin structural polymorphism coding for two hemoglobin components, and populations are polymorphic for a wide range of partitioning values. For instance, the histogram of partitioning values in *Macaca fascicularis* (NUTE 1974) is strongly bimodal, with peaks at 0.35 and 0.65, and the form of the distribution in almost identical (except for scale) to the simulated

distributions obtained for Peromyscus populations. Unfortunately, there appear to be no inheritance data for hemoglobin partitioning in Macaques. In human populations, individuals heterozygous for abnormal hemoglobins characteristically have "partitioning values" that are less than 0.5 (RUCKNAGEL and WINTER 1974; NUTE 1974). For instance, in humans heterozygous for sickle-cell hemoglobin, the proportion of total hemoglobin comprised by Hb S varies from 0.25 to 0.47. Genetic studies on parents and offspring carrying the sickle cell trait indicate that the proportion of Hb S is under fine-scale control, with a narrow heritability estimated at 0.88–0.94 (NANCE and GROVE 1972).

There are two drawbacks to the genetic models presented here. First, the models make no provision for feedback interactions between the various loci. It is assumed that each regulatory allele allows the production of a characteristic amount of globin, regardless of the amounts of other globins in the cell. Hence, in the simulated phenotypes, there is imbalanced synthesis, and the total amounts of α-type (α and γ) and β-type (β and δ) globins differ by an average of 8%. It is assumed that after random association of α-type and β-type subunits into tetramers, the excess chains of either type are simply degraded. That process is known to occur in certain human β-thalassemias, where the synthesis of α- and β-globins is not coordinated and the excess α chains precipitate and/or are degraded (FESSAS 1963; BANK and O'DONNELL 1969; CLEGG and WEATHERALL 1972). However, routine degradation of 4% of total globin production seems wasteful, and globin synthesis is probably coordinated by complicated feedback interactions (BANK *et al.* 1974; GARRICK, DEMBURE and GARRICK 1975; ORKIN, SWAN and LEDER 1975; CHENG and KAZAZIAN, 1976).

The second drawback is that the models assume independent segregation between pairs of gene complexes (*Hba–Hbc* and *Hbb–Hbd*), which probably arose by gene duplication. Linkage relationships between those pairs cannot be tested directly because the *Hba* and *Hbb* structural loci are both monomorphic. The duplicated pairs may have become separated by chromosomal rearrangement. However, in other mammalian species such as man (WEATHERALL and CLEGG 1976; RUCKNAGEL and WINTER 1974) and *Mus musculus* (STERN, RUSSELL and TURNER 1976; POPP and BAILIFF 1973), the duplicate α-type globin loci are closely linked on one chromosome segment, and the β-type loci are closely linked on a separate chromosome segment. If the duplicate gene complexes are similarly linked in Peromyscus, then the models cannot account for some of the empirical data. Specifically, with linkage there should be no genetic variation among P3 sibs from a P2 × P2 mating. Assume that *Hba* and *Hbc* are closely linked. Then in a P2 × P2 mating involving *Hbc*-locus heterozygotes, all P3 offspring will have identical genotypes, $a_1^1c_1^1/a_2^1c_2^1$, where the $a_1^1c_1^1$ chromosome segment is contributed by one parent and the $a_2^1c_2^1$ segment by the other. If the partitioning value is governed solely by the a^1 and c^1 regulators, as in equation (3), then the partitioning values of the P3 sibs should be identical. Empirically, that is not the case. Therefore, if the duplicate complexes are in

fact linked, then the observed variation among the P3 sibs must be ascribed to environmental effects, to feedback interactions between the α-type and β-type loci, or to other factors not included in the models.

The potential deficiencies of the models should not detract from the data itself. The evidence for fine-scale genetic control of partitioning values is clear and unambiguous. The inheritance data demonstrate that almost all the phenotypic variation in populations stems from genetic variation. Furthermore, because the four subspecies have distinctive distributions of partitioning values, those populations must be genetically differentiated with respect to regulatory alleles. That raises the question of whether the hemoglobin phenotypes are subject to differential selection. It is not yet known whether the various hemoglobin components differ in their biochemical and physiological properties. If the components are functionally distinct, then individuals with different proportions of hemoglobins may differ in their oxygen transport capabilities and in their adaptive fitnesses under oxygen stress. Conversely, the high heritability of partitioning values implies that the partitioning phenotypes are selectively neutral (FALCONER 1960). In simple genetic systems with high heritability for fitness, selection in favor of particular phenotypes acts to reduce the amount of genetic variation (FISHER 1930). If there is some simple relationship between partitioning value and fitness, then it is not clear how the broad range of partitionings can be maintained in natural populations.

A preliminary version of this research was included in a Ph.D. dissertation submitted to the Department of Biological Sciences, Stanford University. I thank MARCY UYENOYAMA for assistance in developing the quantitative models. The research was supported by a National Science Foundation Fellowship to the author, by a Public Health Service Training Grant in Population Biology to Stanford University, and by intramural research funds from the University of California at Riverside. Computer facilities were kindly provided by WARD WATT, MARCUS FELDMAN, and JONATHAN ROUGHGARDEN at Stanford and by CHARLES TAYLOR at UCR.

LITERATURE CITED

AUGUSTINSSON, K.-B. and B. OLSSON, 1961 Genetic control of arylesterase in the pig. Hereditas **47**: 1–22.

BAINE, R. M., D. L. RUCKNAGEL, P. A. DUBLIN, JR. and J. G. ADAMS, III, 1976 Trimodality in the proportion of hemoglobin G Philadelphia in heterozygotes: Evidence for heterogeneity in the number of human alpha chain loci. Proc. Natl. Acad. Sci. U.S. **73**: 3633–3636.

BANK, A. and J. V. O'DONNELL, 1969 Intracellular loss of a free α chains in thalassemia. Nature **222**: 295–296.

BANK, A., R. GAMBINO, R. RAMIREZ, G. MANIATIS, C. NATTA, D. KACIAN, S. SPIEGELMAN and P. A. MARKS, 1974 Regulation of globin synthesis in the thalassemias. Ann. New York Acad. Sci. **241**: 247–252.

BARNES, B. W. and A. J. BIRLEY, 1975 Genetical variation for enzyme activity in a population of *Drosophila melanogaster*. II. Aspects of the inheritance of alcohol dehydrogenase activity in $Adh^{S/S}$ flies. Heredity **35**: 115–119.

BREEN, G. A. M., A. J. LUSIS and K. PAIGEN, 1977 Linkage of genetic determinants for mouse β-galactosidase electrophoresis and activity. Genetics **85**: 73–84.

Britten, R. J. and E. H. Davidson, 1971 Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. **46**: 111–133.

Cheng, T. and H. H. Kazazian, Jr., 1976 Unequal accumulation of α- and β-globin mRNA in erythropoietic mouse spleen. Proc. Natl. Acad. Sci. U.S. **73**: 1811–1815.

Chovnik, A., W. Gelbart, M. McCarron and B. Osmond, 1976 Organization of the rosy locus in *Drosophila melanogaster*: Evidence for a control element adjacent to the xanthine dehydrogenase structural element. Genetics **84**: 233–255.

Clegg, J. B. and D. J. Weatherall, 1972 Hemoglobin synthesis during erythroid maturation in β-thalassemia. Nature New Biol. **240**: 190–192.

Coleman, D. L., 1971 Linkage of genes controlling the rate of synthesis and structure of aminolevulinate dehydratase. Science **173**: 1245–1246. ——, 1977 Genetic control of glucokinase activity in mice. Biochem. Genet. **15**: 297–305.

Davis, B. J., 1964 Disc electrophoresis—II. Method and application to human serum proteins. Ann. New York Acad. Sci. **121**: 404–427.

Falconer, D. S., 1960 *Introduction to Quantitative Genetics.* Ronald Press, New York.

Fessas, P., 1963 Inclusions of hemoglobin in erythroblasts and erythrocytes of thalassemia. Blood **21**: 21–32.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Garrick, L., P. Dembure and M. Garrick, 1975 Interaction between the synthesis of α and β globin. Eur. J. Biochem. **58**: 339–350.

Goldberger, R. F., 1974 Antogenous regulation of gene expression. Science **183**: 810–816.

Harris, H., 1975 *The Principles of Human Biochemical Genetics*, 2nd ed. North-Holland, Amsterdam.

Jensen, J. N., 1969 Blood protein polymorphisms in natural populations of the brush mouse (*Peromyscus boylei*). Ph.D. Dissertation, Arizona State University.

King, M.-C. and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188**: 107–116.

Li, I. C. and W. L. Daniel, 1976 Correlation between structural variation and activity of murine kidney β-galactosidase: Implications for genetic control. Biochem. Genet. **14**: 933–952.

Lodish, H. F., 1976 Translational control of protein synthesis. Ann. Rev. Biochem. **45**: 39–72.

Maybank, K. M. and W. D. Dawson, 1976 Genetic and developmental variation of hemoglobin in the deer mouse, *Peromyscus maniculatus.* Biochem. Genet. **14**: 389–400.

Nance, W. E. and J. Grove, 1972 Genetic determination of phenotype variation in sickle cell trait. Science **177**: 716–717.

Nielsen, J. T., 1977 Variation in the number of genes coding for salivary amylase in the bank vole, *Clethrionomys glareola.* Genetics **85**: 155–169.

Nute, P. E., 1974 Multiple α-globin loci in monkeys, apes, and man. Ann. New York Acad. Sci. **241**: 39–60.

Orkin, S., D. Swan and P. Leder, 1975 Differential expression of α- and β-globin genes during differentiation of cultured erythroleukemic cells. J. Biol. Chem. **250**: 8753–8760.

Popp, R. and E. Bailiff, 1973 Sequence of amino acids in the major and minor β chains of the diffuse hemoglobin from BALB/c mice. Biochim. Biophys. Acta **303**: 61–67.

Prager, E. M. and A. C. Wilson, 1975 Slow evolutionary loss of the potential for interspecific hybridization in birds: A manifestation of slow regulatory evolution. Proc. Natl. Acad. Sci. U.S. **72**: 200–204.

RASMUSSEN, D. I., J. N. JENSEN and R. K. KOEHN, 1968 Hemoglobin polymorphism in the deer mouse, *Peromyscus maniculatus*. Biochem. Genet. **2**: 87–92.

RUCKNAGEL, D. L. and W. P. WINTER, 1974 Duplication of structural genes for hemoglobin α and β chains in man. Ann. New York Acad. Sci. **241**: 80–92.

SCHWARTZ, D., 1976 Regulation of expression of Adh genes in maize. Proc. Natl. Acad. Sci. U.S. **73**: 582–584.

SMITH, D. W. E., 1975 Reticulocyte transfer RNA and hemoglobin synthesis. Science **190**: 529–535.

SNYDER, L. R. G., 1977 Polymorphisms for hemoglobin and 6-phosphogluconate dehydrogenase in the rodent genus, *Peromyscus*. Ph.D. Dissertation, Dept. of Biological Sciences, Stanford University (University Microfilms, Ann Arbor, Michigan). ——, 1978 Genetics of hemoglobin in the deer mouse, *Peromyscus maniculatus*. I. Multiple α- and β-globin structural loci. Genetics **89**: 511–530.

STERN, R., E. RUSSELL and B. TAYLOR, 1976 Strain distribution and linkage relationship of a mouse embryonic hemoglobin variant. Biochem. Genet. **14**: 373–381.

THOMPSON, P. and M. J. HORNING, 1973 Regulatory interactions involving two hemoglobin loci of *Chironomus*. Biochem. Genet. **8**: 309–319.

THOMPSON, P. and G. PATEL, 1972 Compensatory regulation of two closely linked hemoglobin loci in *Chironomus tentans*. Genetics **70**: 275–290.

WARD, R. D., 1975 Alcohol dehydrogenase activity in *Drosophila melanogaster*: a quantitative character. Genet. Res., Camb. **26**: 81–93.

WEATHERALL, D. J. and J. B. CLEGG, 1972 *The Thalassemia Syndromes*, 2nd ed. Blackwell, Oxford. ——, 1976 Molecular genetics of human hemoglobin. Ann. Rev. Genet. **10**: 157–178.

WILSON, A. C., S. S. CARLSON and T. J. WHITE, 1977 Biochemical evolution. Ann. Rev. Biochem. **46**: 573–639.

WILSON, A. C., L. R. MAXSON and V. M. SARICH, 1974 Two types of molecular evolution. Evidence from studies of interspecific hybridization. Proc. Natl. Acad. Sci. U.S. **71**: 2843–2847.

WILSON, C. M., E. G. ERDOS, J. F. DUNN and J. D. WILSON, 1977 Genetic control of renin activity in the submaxillary gland of the mouse. Proc. Natl. Acad. Sci. U.S. **74**: 1185–1189.

Corresponding editor: R. C. LEWONTIN