MUTATION RATES, POPULATION SIZES AND AMOUNTS OF ELECTROPHORETIC VARIATION OF ENZYME LOCI IN NATURAL POPULATIONS

E. ZOUROS

Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

Manuscript received January 31, 1978
Revised copy received September 12, 1978

ABSTRACT

A method is presented for estimating relative mutation rates or relative effective population sizes, under the hypothesis of adaptively neutral allelic variation. This method was applied to seven surveys of electrophoretic variation. It was observed that electrophoretic mutation rates so obtained follow the gamma distribution and, in Drosophila, are positively correlated with the molecular weights of the enzyme subunits. The variance in mutation rate is larger under the step-wise model of electrophoretic mutation than under the infinite-alleles model. Rates for the most variable loci may exceed rates for less variable loci by a factor of 500. For completely invariant loci, this factor may be as high as 4×10^4, an observation suggesting that these loci are subject to purifying selection. In contrast to mutation rates, effective population sizes may vary at the most by a factor of ten. These results support the hypothesis that differences in the amount of electrophoretic variability among polymorphic loci may reflect differences in the rate by which electrophoretically detectable variation is generated in populations.

IN this paper, I examine some recent hypotheses relating to the amounts of electrophoretic variation observed in natural populations. These hypotheses can be summarized as follows:

(1) GILLESPIE and KOJIMA (1968) were the first to observe a correlation between enzyme function and amount of electrophoretic variation. GILLESPIE and LANGLEY (1974) have reviewed the relevant observations and formulated a hypothesis suggesting that the amount of genetic variation exhibited by an enzyme is determined by its substrate specificity. Enzymes capable of utilizing a family of substrates are more variable than enzymes that utilize a unique substrate. JOHNSON (1974a) has proposed that, in addition to substrate specificity, the ability of an enzyme to regulate the flux along a pathway is important in determining the amount of genetic variability. A multiplicity of forms is favored for regulatory enzymes, but a single, most efficient form is established for nonregulatory enzymes. I have stated elsewhere (ZOUROS 1976a) that, although the two hypotheses are different at the biochemical level, they are identical at the population genetics level, in the sense that both postulate balancing

selection as the force responsible for the amounts of genetic variability retained in natural populations.

Regardless of the validity of these explanations, the observation that, within broad taxonomic units, some enzyme loci always appear to be less variable than others, presents a difficulty for the neutral hypothesis of genetic variability. Until recently, all the statistical methods developed for the testing of this hypothesis predict a certain distribution of locus heterozygosity or of some other statistic. When two or more such distributions are obtained, nothing in the theory of neutral variation would predict that certain enzymes should always occupy the same tail of the distribution. Echoing this difficulty of the hypothesis of neutrality, EWENS and FELDMAN (1976) stated, after reviewing the statistical evidence for and against the theory, that "it is possible that the most useful approaches involving gene frequencies will not involve theoretical populations genetics but rather general arguments and statistical tests relating to gene frequencies and locus function."

However, enzymes that have similar metabolic functions may show comparable amounts of electrophoretic variability, not because they are subject to the same types of selection, but rather because the parameters that determine separation of variants on an electrophoretic gel (JOHNSON 1977) are more similar among such enzymes than they are among enzymes with different functional properties. This similarity in electrophoretic parameters among enzymes may be the result either of phylogenetic relatedness or of selection for a convergent optimum in three-dimensional configuration (ZOUROS 1975). A related observation is that quaternary structure is directly related to enzyme heterozygosity (ZOUROS 1976b; WARD 1977; HARRIS, HOPKINSON and EDWARDS 1977).

This argument accounts for the fact that certain families of enzymes show different amounts of electrophoretic variability, without implicating balancing selection. It does so by postulating different rates of "electrophoretic mutation" for loci coding for these families of enzymes. By "electrophoretic mutation rate," I mean the probability that an amino acid substitution will result in a change of mobility detectable under a given regime of electrophoretic conditions. The question then becomes, how different must these rates be so that observed differences in amounts of variation could be explained by the hypothesis of neutrality? In principle, it would be possible to determine from information about the physicochemical properties of an enzyme molecule the range within which its electrophoretic mutation rate must fall. It is even possible to obtain an estimate of this rate by direct observation (MUKAI and COCKERHAM 1977). There is no reason to expect that the two rates will be the same, but the hypothesis of neutrality may be placed in serious doubt if it had to assume differences among mutation rates that are much larger than the ones predicted from the physico-chemical properties of the enzymes or the ones observed by direct experimentation.

(2) The hypothesis of variable mutation rate has recently received attention (NEI, CHAKRABORTY and FUERST 1976; FUERST, CHAKRABORTY and NEI 1977). These authors were not so much interested in providing an alternative to the Gillespie-Langley or Johnson hypotheses as they were in improving the statistical

methodology for the testing of neutrality. An interesting suggestion made by NEI, FUERST and CHAKRABORTY (1976) and NEI, CHAKRABORTY and FUERST (1976) is that the mutation rate per locus is distributed as a gamma distribution with coefficient of variation approximately one.

(3) KOEHN and EANES (1977) observed a strong positive correlation between molecular weights of enzyme subunits and amounts of electrophoretic variation in Drosophila species. NEI, FUERST and CHAKRABORTY (1978) have extended this observation to a much wider collection of organisms. These observations provide further support for the hypothesis that the amount of variation of an enzyme is primarily determined by its physico-chemical properties, rather than by the environmental diversity it encounters.

I here present an algebraic method for the estimation of the difference in relative magnitudes of electrophoretic mutation rates among enzyme loci. The same method may be used to estimate differences in effective sizes of populations. It utilizes estimates of $M = 4Nu$, where N is the effective population size and u is the mutation rate. M can be estimated from locus heterozygosities or from observed number of alleles and sample sizes, according to EWENS' (1972) theory of sampling of selectively neutral alleles. The method is applied to seven surveys of electrophoretic variability obtained from the literature. The results offer some insights into the points stated above. Specifically, it will be shown that for conservative loci the estimated mutation rates are very low compared to those of polymorphic loci. This is interpreted to mean that electrophoretically invariant loci are under the influence of purifying selection. Among polymorphic loci (including those that are characterized as such on the basis of the 1% heterozygosity criterion), the differences among estimated mutation rates do not appear to be very large. Thus, differences in heterozygosities among such enzymes may be easily attributed to differences in electrophoretic mutation rates. It will also be shown that the mutation rates so obtained follow the gamma distribution and that there is a good correlation between mutation rates and molecular weights of enzyme subunits.

MATERIALS AND METHODS

Method: Let us assume that in an electrophoretic study l protein loci were surveyed for variability in k populations. The populations may belong to the same species or closely related species. The only requirement of the method is that the populations are taxonomically close enough so that mutation rates at homologous loci can be considered equal. Under the hypothesis of neutrality and according to the infinite-alleles model, the effective number of alleles (n), at a given locus (j), in a given population (i), is given by

$$n_{ij} = 1 + 4N_i u_j \ , \tag{1}$$

where N is the effective population size and u the mutation rate. An estimate of n is obtained from

$$\hat{n}_{ij} = 1/\sum_m p^2_{ijm} \tag{2}$$

where p_{ijm} is the observed frequency of the m^{th} allele of the j^{th} locus in the i^{th} population. If we define

$$a_{ij} = \hat{n}_{ij} - 1 \text{ and } a_{ib} = \hat{n}_{ib} - 1 \ , \tag{3}$$

then it follows from (1) that the ratio a_{ij}/a_{ib} provides an estimate of u_j/u_b, where u_b is the mutation rate of locus b ($1 \leq j, b \leq l$).

For two loci ($j = 1$, $b = 2$) studied in k populations there will be k estimates of u_1/u_2. If we plot a_{i1} *versus* a_{i2} for $i = 1$ to k we obtain k points in a two-dimensional space. The slope of a straight line passing through the origin gives a single estimate of u_1/u_2. The line is determined by the following property: it minimizes the summation of the squares of the distances of the k points from itself. It must be noted that in this case neither variable (a_{i1} or a_{i2}) is treated as dependent or independent, and the line minimizes distances rather than residuals. In that respect it is not a regression line.

The argument can be extended to l loci, in which case there will be $k(l-1)$ estimates of u_j/u_b. The simple case for $l = 3$ is graphed in Figure 1. In the figure, the quantity a_{ij} (j=1,2,3; $i = 1$ to k) is measured along the axis A_j. For any given i, there corresponds a point U_i. Let OX be a straight line that passes through the origin, and let d_i be the distance of point U_i from the line. Line OX has the property that $\sum_{i=1}^{k} d_i^2$ is minimum. Let $\vec{x}$ be the unit vector of line OX and let x_1, x_2, x_3 be its coordinates. The coordinates are related to mutation rates as follows:

$$u_1/x_1 = u_2/x_2 = u_3/x_3 \ . \tag{4}$$

It is convenient to define v as

$$v_j = u_j / \sum_{j=1}^{1} u_j \ . \tag{5}$$

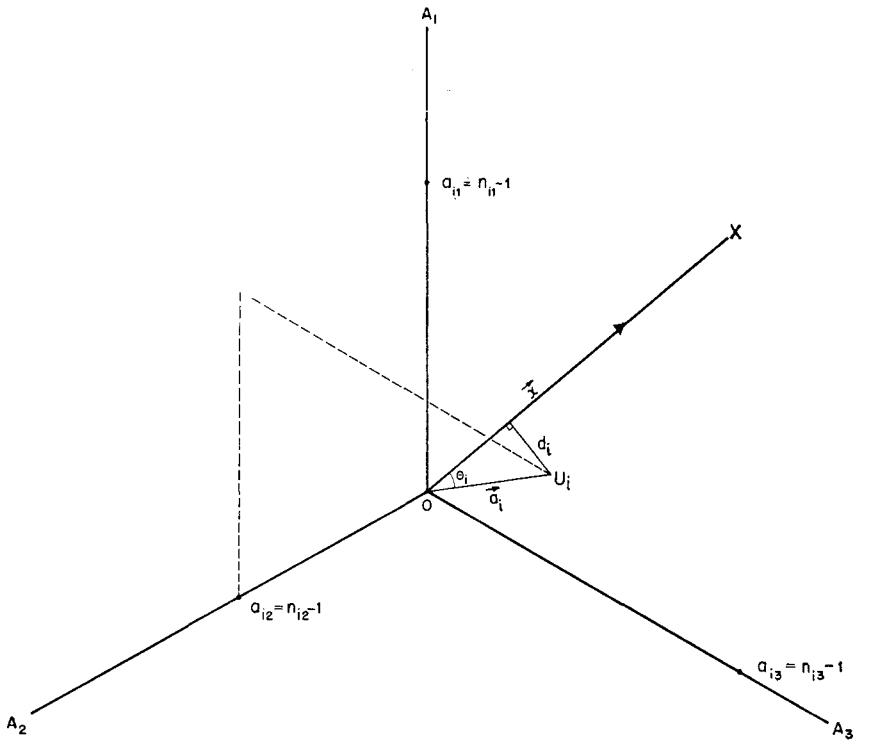

FIGURE 1.—A presentation of the method for the estimation of relative mutation rates in a three-dimensional space.

It follows from (4) that an estimate of v_j can be obtained as:

$$\hat{v}_j = x_j / \sum_{j=1}^{l} x_j \ . \tag{6}$$

It also follows from (5) that $\sum_{j=1}^{l} \hat{v}_j = 1$.

The algebra for the calculation of $\hat{v}_j$ is given in the APPENDIX. Because the $\hat{v}_j$ vector is calculated from electrophoretic data, it estimates relative electrophoretic mutation rates rather than true mutation rates. It will be referred to in the text as the vector of "relative electrophoretic mutation rate," or REMR for short.

The method outlined above can be modified slightly to apply to the step-wise model of electrophoretic mutation. Under this model, the equivalent of (1) is

$$n_{ij} = (1 + 8N_i u_j)^{1/2} \ , \tag{7}$$

so that the equivalent of (3) becomes

$$a_{ij} = \hat{n}^2_{ij} - 1 \ . \tag{8}$$

The rest of the method remains unchanged.

The same method can be used to estimate ratios of effective sizes of populations. If we define

$$a_{cj} = \hat{n}_{cj} - 1 \ , \tag{9}$$

where $\hat{n}_{cj}$ is the effective number of alleles at locus j in population c, then it follows from (1), (2) and (9) that the ratio a_{ij}/a_{cj} provides an estimate of N_i/N_c. The quantities a_{ij} ($i = 1$ to k, $j = 1$ to l) are now plotted in a space of k dimensions; there will be l points on each dimension.

The outlined method provides estimates of relative mutation rates or effective population sizes based on observed rates or effective population sizes based on observed single-locus heterozygosities. It is known from the works of STEWART (1976) and LI and NEI (1975) that single-locus heterozygosities have a large stochastic variance. As a result, the estimated mutation rates may be associated with a large error. The method attempts to minimize this error for a given number of independent estimates of locus heterozygosities. The accuracy of the method depends largely on the number of observations. Relative mutation rates become more reliable with increasing numbers of populations for which single-locus heterozygosities are available. Similarly, effective population size estimates are more accurate when more loci are studied in each population.

The use of heterozygosities for the estimation of relative mutation rates requires that the expectation of

$$\mathrm{a} = \frac{1}{\Sigma p^2_i} - 1 \tag{10}$$

be a linear function of M. It can be shown that this is not strictly true but that $E[a]$ is approximately linear with M, and that the approximation becomes better as M becomes larger.

If we let $j = \Sigma p_i^2$ then we have, for the infinite-alleles model (*e.g.*, LI and NEI 1975),

$$E[j] = \frac{1}{1+M} \text{ and } V[j] = \frac{2M}{(1+M)^2(2+M)(3+M)} \tag{11}$$

An approximation of $E[a]$ can be obtained by using the Taylor expansion:

$$E[a] = E\left[\frac{1}{j}\right] - 1 \approx \frac{1}{E[j]} + \frac{1}{E[j]^3} V[j] - 1 \approx M \left(1 + \frac{2(1+M)}{(2+M)(3+M)}\right) \tag{12}$$

The deviation from linearity can be obtained by examining the term $2(1 + M)/(2 + M)(3 +$

M). This term is 1/3 at $M = 0$ and, at $M = 1$, has a maximum of 0.343 at $M = \sqrt{2} - 1$, and tends to zero as M increases infinitely.

For the step-wise model (MORAN, 1976)

$$E[j] = \frac{1}{(1+2M)^{1/2}} \text{ and } V[j] \approx \frac{M}{3+11.25\,M+13\,M^2+1.7\,M^3} \qquad (13)$$

so that

$$E[a] \simeq \frac{1}{E[j]^2} + \frac{3}{E[j]^4} V[j] - 1 \simeq 2M\left(1 + \frac{3(1+2M)^2}{2(3+11.25M+13M^2+1.7M^3)}\right) \qquad (14)$$

The second term in parenthesis is 0.5 at $M = 0$, 0.466 at $M = 1$, has a maximum of 0.508 at $M = 0.17$, and tends to zero as M goes to infinity. In Figure 2, $E[a]$ is plotted against M. The relationship between $E[a]$ and M is graphed only for low values of M, since it is in this interval of M where the nonlinearity is more pronounced. It can be seen that in both models the deviation from linearity is minor. In the presence of other disturbing factors, such as bottle-neck effect and the drifting variance of heterozygosity, this nonlinear effect is probably negligible.

Materials: The method described above was applied to seven electrophoretic studies. They

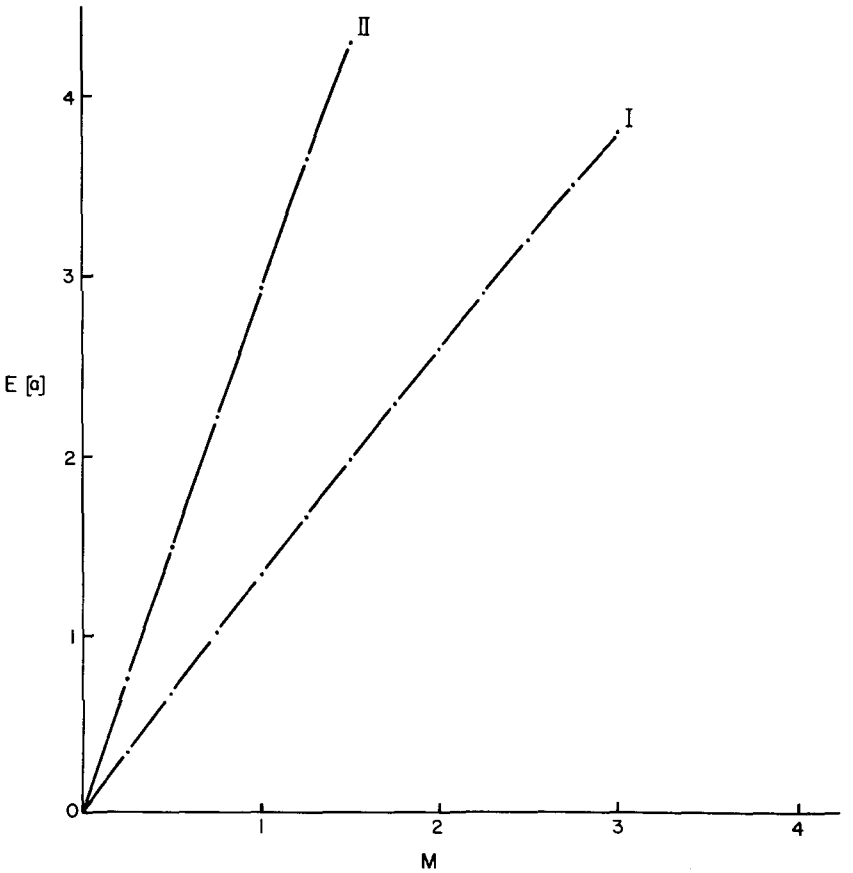

FIGURE 2.—The relationship between $E[a]$ and M, where $M = 4Nu$ and $a = (1/\Sigma p_i^2) - 1$ for the infinite-alleles model (line I), or $a = (1/\Sigma p_i^2) - 1$ for the step-wise model (line II).

are listed in Table 1. In each of these studies a number of protein loci were surveyed in more than one taxon. In one study (NEVO *et al.* 1974), the taxa represent different karyotypes that may belong to different subspecies within the same species. In the other studies, the taxa represent different subspecies or closely related species. In all cases, the taxa are very closely related, so that the assumption of equal mutation rate for homologous loci is justified.

In most cases, locus heterozygosities were provided by the authors. In such cases, the effective number of alleles was estimated from

$$n_{ij} = \frac{1}{1-h_{ij}} , \qquad (15)$$

where h_{ij} is the observed heterozygosity of locus j in taxon i. In the cases where only electromorph frequencies were provided, the effective number of alleles was estimated by using equation (2). In a number of studies, the original authors have surveyed several populations from each taxon. In such cases the electromorph frequency for the taxon was obtained as the mean population frequency weighted by population size.

A computational problem arises when a locus is completely monomorphic in all taxa of a given group. For such loci, the method generates relative mutation rates equal to zero. To avoid this problem, the heterozygosity of an invariant locus in one taxon was given the arbitrary value of 0.001. This is the amount of heterozygosity that one would have observed if 2,500 genomes were examined for each taxon, and one variant allele was observed. The taxon that was arbitrarily assigned a heterozygosity of 0.001 was chosen at random. When more than one locus was monomorphic in the entire group, each was assigned a heterozygosity of 0.001 in one taxon, but this taxon was different for each locus.

TABLE 1

Electrophoretic data utilized in the study

Group identification number	Group	Source	Number of taxa in group	Number of loci analyzed in this study	Mean no. of genomes per locus per taxon*
1	the *Drosophila willistoni* species group	AYALA *et al.* (1974)	6 (4 species, 2 semi-species)	28	320
2	Hawaiian species of Drosophila	AYALA (1975)	7 species	31	37
3	The *Drosophila bipectinata* species group	YANG *et al.* (1972)	6 species	19	192
4	the *Drosophila pseudoobscura* group	PRAKASH (1977)	3 species	43	273
5	Species of the genus Menedia (fishes)	JOHNSON (1975)	5 species	19	476
6	Species of the genus Mus (mice)	SELANDER and YANG (1969); SELANDER *et al.* (1969)	3 subspecies	36	157
7	*Thomomys talpoides* (pocket gophers)	NEVO *et al.* (1974)	6 karyotypes	31	92

* This number is only approximately true; within a given group the number of genomes sampled may vary considerably from locus to locus or from taxon to taxon.

RESULTS

The willistoni *group of Drosophila*: The method will be first applied to a single set of data mainly because this set represents the most complete and most comprehensive study to date of electrophoretic variability in natural populations. AYALA *et al.* (1974) have provided information about electrophoretic variability at 28 to 32 loci in several populations belonging to four species and two semispecies of the *willistoni* group. The number of genomes surveyed in each species or semispecies varied from 100 to 600. Twenty-eight loci were utilized here. Four loci (given in the footnote of Table 2) were excluded because they were not surveyed in all six taxa. A special feature of this study, which makes it particularly useful for the purpose of this paper, is that no locus was found to be completely monomorphic in all six taxa.

Table 2 gives the relative electrophoretic mutation rates (REMR) for these 28 loci. The first column under "Relative electrophoretic mutation rate" gives the rates derived under the infinite-alleles model (KIMURA and CROW 1964). This model assumes that any mutation event produces a new allele that did not previously exist in the population. The second column gives the relative mutation rates under the step-wise model. This model (OHTA and KIMURA 1973) assumes that an amino acid substitution either does not change the electrophoretic mobility of the enzyme, or it results in an electromorph that is removed by one step (forward or backward) along a ladder of predefined electrophoretic mobilities. JOHNSON (1974b) compared the two models and concluded that they represent the two extremes of a continuum of possibilities that may determine the electrophoretic detection of newly arisen mutants. This observation is particularly useful because it suggests that if one estimates a parameter such as mutation rate or effective population size under both models, one can be reasonably certain that the two estimates define the limits within which the true value of the parameter lies.

The last three columns of Table 2 give relative electrophoretic mutation rates calculated in a different way. If one replaces $E[j]$ in (11) or (13), depending on the model, with observed mean locus homozygosity, $1 - h_j$, one can have an estimate of M_j for each locus. This estimate can be used instead of a_{ij} in (3). This method has the disadvantage that the effective population size is assumed to be the same in all taxa in a given group. Obviously, this might not be so. But, as will be discussed later (Table 7), the application of the method introduced in this paper for the estimation of relative effective population sizes yields numbers that are not very different among taxa of a given group of organisms. Under the assumption of neutrality, one may then assume equal effective population sizes and proceed to calculate relative mutation rates in the way outlined above. For the sake of comparison this was done for the *willistoni* group for the infinite-alleles model. The results are given in the fifth column of Table 2.

The last two columns of Table 2 give the relative electrophoretic mutation rates one obtains when M is not estimated from heterozygosities, but from EWENS' (1972) formula

TABLE 2

Relative electrophoretic mutation rates (REMR) of 28 enzyme loci in the* willistoni *group*

Enzyme locus	Mean heterozygosity (h_j)	Relative electrophoretic mutation rate† 1	2	3	4	5‡
Lap-5	0.401	0.0719	0.0682	0.0726	0.0575	0.0543
Est-2	0.333	0.0673	0.0654	0.0541	0.0692	0.0662
Est-3	0.098	0.0108	0.0067	0.0118	0.0304	0.0280
Acph-1	0.323	0.0610	0.0513	0.0517	0.0588	0.0434
Ald-1	0.206	0.0238	0.0158	0.0281	0.0260	0.0401
Fum	0.208	0.0257	0.0175	0.0285	0.0395	0.0528
G6pdh	0.206	0.0263	0.0183	0.0282	0.0410	0.0516
Odh-1	0.116	0.0129	0.0086	0.0143	0.0432	0.0501
Xdh	0.525	0.1466	0.2101	0.1200	0.0725	0.0947
Ao-1	0.581	0.1359	0.1430	0.1505	0.0893	0.1012
Ao-2	0.590	0.1398	0.1478	0.1560	0.0710	0.0828
Pgm	0.168	0.0255	0.0211	0.0219	0.0274	0.0370
Adk-1	0.585	0.1331	0.1382	0.1530	0.0645	0.0732
Hk-1	0.118	0.0146	0.0102	0.0145	0.0297	0.0225
Hk-2	0.208	0.0329	0.0257	0.0285	0.0400	0.0450
Got	0.060	0.0070	0.0046	0.0070	0.0372	0.0207
Hbdh	0.041	0.0040	0.0023	0.0046	0.0372	0.0230
Est-4	0.137	0.0269	0.0238	0.0172	0.0291	0.0240
Hk-3	0.054	0.0064	0.0040	0.0062	0.0161	0.0066
Ald-2	0.113	0.0129	0.0086	0.0032	0.0172	0.0233
Adh	0.025	0.0012	0.0007	0.0138	0.0076	0.0037
Mdh-2	0.036	0.0017	0.0009	0.0014	0.0135	0.0088
αGpdh	0.014	0.0013	0.0007	0.0020	0.0117	0.0030
Idh	0.012	0.0011	0.0007	0.0015	0.0091	0.0068
Me-1	0.032	0.0030	0.0018	0.0013	0.0123	0.0061
To	0.029	0.0030	0.0017	0.0036	0.0179	0.0068
Tpi-2	0.012	0.0011	0.0007	0.0032	0.0084	0.0061
Adk-2	0.028	0.0027	0.0017	0.0014	0.0224	0.0183

* Four loci (*Est-5*, *Est-6*, *Est-7* and *Me-2*) were omitted because they were not studied by the original authors (AYALA *et al.* 1974) in all six taxa.

† The REMR's in each column were calculated as follows: (1) M_{ij} estimated as n_{ij}-1 (infinite-alleles model). (2) M_{ij} estimated as $(n^2_{ij}\text{-}1)/2$ (step-wise model). (3) M_j estimated as $(1/(1\text{-}h_j))\text{-}1$ (infinite-alleles model). (4) M_{ij} estimated from EWENS (1972) (infinite-alleles model); the whole species treated as one population. (5) M_{ij} estimated from EWENS (1972) (infinite-alleles model); one population considered from each species.

‡ The populations used for the estimation of REMR are as follows: Tucupita of *D. willistoni*, Caripito of *D. tropicalis*, Tucupita of *D. equinoxialis*, Barinitas of *D. paulistorum* (Amazonian), El Dorado of *D. paulistorum* (Orinocan), and Barinitas of *D. nebulosa*. For *D. nebulosa*, Barinitas was selected over Guri because one locus, *ODH*, was not assayed in the latter population.

$$E[k] = \frac{M}{M} + \frac{M}{M+1} + \ldots + \frac{M}{M+2n-1}\ , \qquad (16)$$

where k is the observed number of alleles in a sample of $2n$ genomes. Relative electrophoretic mutation rates were calculated in two different ways. In one case (column 6 of Table 2), all populations of the same species were pooled together. $2n$ was set equal to the number of genomes examined in each species

(summation over all populations) and $E[k]$ was set equal to the observed number of alleles in the whole species. In the second case (last column of Table 2), only one population from each species (or subspecies for *D. paulistorum*) was considered. This was the population with the largest sample of genomes. The names of these populations are given in the footnote of Table 2. The value of M was obtained by Newton's reiteration in a digital computer for most of the combinations of k and n that appear in the original data (AYALA *et al.* 1974). For sample sizes that were very close to the ones used in the reiteration, M was obtained by linear approximation using the values of M corresponding to the nearest lower and nearest higher n, and same k. This approximation introduces a negligible error in the estimation of relative mutation rates.

Several points of interest emerge from Table 2: (1) Relative electrophoretic mutation rates in columns 3 and 4 were estimated from locus heterozygosities and are expected to be highly correlated with mean locus-heterozygosities. It may be observed, however, that the rankings of mean heterozygosities and REMR are not identical. In the *willistoni* group, the locus with the higher heterozygosity (when averaged over all species and semispecies) is *Ao-2*; yet, it turns out that the locus with the highest electrophoretic mutation rate is *Xdh*. This locus ranks fourth in terms of average heterozygosity. As expected, when M is calculated from average locus homozygosities (column 5), the resulting REMR's rank identically to mean heterozygosities.

The fact that the method produces results that can be different from the one inferred from a simple inspection of heterozygosities suggests that it may have some intrinsic value in the estimation of relative mutation rates. It is worth noting, in this respect, that there is good reason to believe that the assignment of the highest electrophoretic mutation rate to *Xdh* is not fortuitous. This is the locus at which SINGH, LEWONTIN and FELTON (1976) discovered 37 allelic classes within 146 independently extracted genomes of *Drosophila pseudoobscura*. It is also the enzyme with the second largest subunit molecular weight known in Drosophila (KOEHN and EANES 1977).

(2) When sets one and two of relative electrophoretic mutation rates are compared, it is observed that the step-wise model generates more heterogeneity among mutation rates than does the infinite-alleles model. Under the latter model, the ratio between the highest (*Xdh*) and lowest (*Idh*) mutation rates is 135.7. For the step-wise model this value is more than twice as large (318.3). This observation is quantified in Table 3, where the gamma distribution is fitted to REMR's under both models. It can be seen that the scale parameter $B = s^2/\bar{x}$ (where $\bar{x} = 1/l$ is the same for both models) is always larger in the step-wise model. This means that for the hypothesis of neutrality the step-wise model is more restrictive. This is so because if the systematic differences in the amount of heterozygosity among loci had to be explained on the basis of variable mutation rates alone, then the stepwise model would require that these differences be much larger than the ones required under the infinite-alleles model.

(3) The last three columns of Table 2 must be compared to the third column, because all refer to the infinite-alleles model. It can be seen that the differences

TABLE 3

Fit of the gamma distribution to the distribution of relative electrophoretic mutation rates

Group	B/C	Relative electrophoretic mutation rate interval <0.01	0.01–0.02	0.02–0.03	0.03–0.05	0.05–0.07	0.07–0.10	>0.10	χ^2/P
1	0.0601	11	4	5	1	2	1	4	4.505
	0.5942	10.16	4.31	2.94	3.84	2.29	2.02	3.02	0.479
	0.0843	14	4	3	0	3	0	4	6.380
	0.4239	12.37	3.67	2.41	3.08	1.90	1.76	2.80	0.271
2	0.0203	9	3	2	10	3	4	0	11.230
	1.5904	5.27	6.79	5.64	7.19	3.47	1.92	0.71	0.047
	0.0256	11	2	4	7	3	3	1	6.500
	1.2598	6.73	6.48	5.02	6.39	3.25	2.08	1.05	0.260
3	0.1005	8	2	1	1	3	1	3	3.165
	0.5235	6.19	2.41	1.69	2.36	1.58	1.58	3.19	0.675
	0.1685	10	1	1	3	1	0	3	2.900
	0.3124	8.66	1.94	1.27	1.69	1.14	1.14	3.15	0.715
4	0.0754	28	2	1	6	2	2	2	5.206
	0.3085	24.94	5.07	3.05	3.65	2.11	1.81	2.36	0.391
	0.1719	29	2	4	2	4	0	2	10.768
	0.1353	31.00	2.79	1.68	2.06	1.29	1.20	2.97	0.056
5	0.1817	10	1	1	2	2	1	2	1.860
	0.2897	9.01	1.88	1.22	1.59	1.08	1.10	3.12	0.868
	0.2326	10	2	0	2	3	0	2	7.672
	0.2262	10.15	1.63	1.03	1.37	0.89	0.93	3.00	0.175
6	0.0604	21	3	1	3	2	2	4	5.394
	0.4601	16.88	5.26	3.31	4.07	2.34	1.94	2.20	0.370
	0.0661	22	2	2	2	2	1	5	7.666
	0.4199	17.57	5.00	3.13	3.85	2.23	1.87	2.34	0.176
7	0.0828	17	4	2	1	2	0	5	5.696
	0.3998	14.82	4.00	2.54	3.22	1.95	1.77	2.70	0.337
	0.1165	19	3	1	2	1	0	5	4.246
	0.2769	17.11	3.25	2.01	2.57	1.61	1.49	2.94	0.515
8*	0.0454	9	3	2	6	2	0	4	6.132
	0.8447	6.92	4.39	3.28	4.39	2.65	2.21	2.16	0.293
	0.0720	12	1	4	3	2	0	4	5.845
	0.5345	9.72	3.72	2.55	3.33	2.08	1.87	2.73	0.322
9†	0.0139	3	6	5	7	4	3	0	2.199
	2.5677	2.02	5.40	5.91	8.54	3.92	1.79	0.42	0.821

B, the first number in the pairs of second column and *C*, the second number, are the estimated parameters of the gamma distribution. The numbers in the body of the table are the observed and expected numbers of loci with relative electrophoretic mutation rates included within the specified interval. The last column gives the results of the χ^2 test for the fitting of the observed and expected distributions (d.f. 5). Within each group, the first set of numbers refers to the infinite-alleles model, the second set of numbers to the step-wise model.

* *Willistoni* and *Hawaiian* Drosophila groups pooled.

† *Willistoni* group; parameter *M* estimated according to EWENS (1972).

between columns 1 and 3 are minor. This observation is in agreement with the result, to be discussed later, that under the hypothesis of neutrality effective population sizes seem not to vary much among taxa. Likewise, there is not much difference between columns 4 and 5. In both cases, the parameter *M* was estimated according to EWENS (1972). But these two columns differ quite significantly

from either 1 or 3, in that they produce a more even distribution of REMR's. The ratio between the highest (*Xdh*) and lowest (*Adh*) REMR in column 4 is only 12, and the variance among REMR's of column 1 is four times as large as the variance among REMR's of column 4. The significance of this observation and the bearing it has on the hypothesis of neutrality will be discussed later.

(4) The suggestion by NEI, FUERST and CHAKRABORTY (1976) that mutation rates may be distributed according to a gamma distribution with a coefficient of variation close to one, was based on observed rates of amino acid substitutions during evolution in 19 polypeptides. One may ask whether electrophoretic mutation rates, calculated under the assumption of neutrality, follow the gamma distribution. This is done for the data of Table 2. The two parameters of the gamma probability density function were estimated using the method of matching of moments (HASTINGS and PEACOCK 1975). The scale parameter, B, was estimated as $B = s^2/\bar{x}$ and the shape parameter, C, as $C = (\bar{x}/s)^2$, where $\bar{x}$ is the sample mean and s^2 is the sample variance (unadjusted). The function was then integrated numerically. The expected and observed distributions are given in Figure 3, and the test of the goodness of fit is given in Table 3. It can be seen from both Figure 3 and Table 3 that the fit is quite good. But the fit to the gamma distribution with coefficient of variation one is less satisfactory (Table 4). The deviation is not significant for the infinite-alleles model, but it is for the stepwise model.

(5) KOEHN *and* EANES (1977) have obtained from the literature subunit

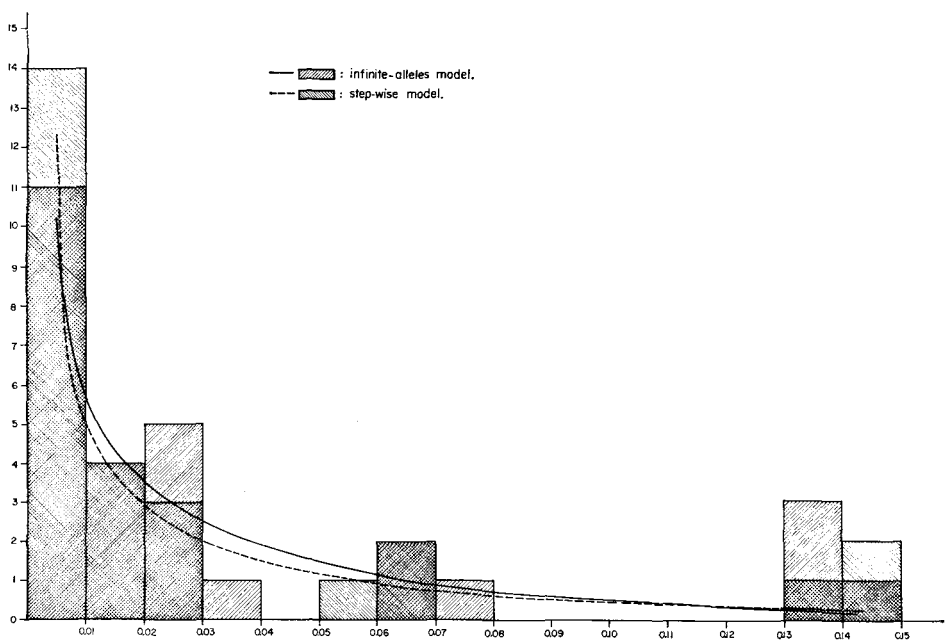

FIGURE 3.—The expected gamma distributions and the obtained distributions (histograms) of relative electrophoretic mutation rates in *Drosophila willistoni*.

TABLE 4

Fit of the gamma distribution with coefficient of variation one to the distribution of relative electrophoretic mutation rates of the willistoni *group*

REMR range	Expected	Observed under the infinite-alleles model	Observed under the step-wise model
<0.10	6.84	11	14
0.01–0.02	5.17	4	4
0.02–0.03	3.90	5	3
0.03–0.05	5.18	1	0
0.05–0.07	2.96	2	3
0.07–0.10	2.25	1	0
>0.1	1.70	4	4
		$\chi^2 = 10.596$	$\chi^2 = 18.510$
		d.f. = 6	d.f. = 6
		$P = 0.102$	$P = 0.005$

molecular weights for eleven Drosophila enzymes. They observed that these weights were strongly correlated with various measurements of enzyme polymorphism. In Figure 4, the relative electrophoretic mutation rates of the *D. willistoni* group (columns 3 and 4 of Table 2) are plotted against subunit molecular weight. There are ten enzymes involved [alkaline phosphatase, for which KOEHN and EANES (1977) provided the molecular weight, was not studied by AYALA *et al.* (1974)]. Two loci with aldolase activity and two loci with acetaldehyde oxidase activity were reported by AYALA *et al.* (1974). For these enzymes the average electrophoretic mutation rate was used for the graph, as well as for the statistical test of Table 5.

It can be seen from Figure 4 that there is a strong positive correlation between mutation rate, derived under the hypothesis of neutrality, and subunit molecular weight. The safest way to quantify this relationship is to obtain an index of correlation using a nonparametric test. Kendall's rank correlation coefficient is 0.69 (Table 5), and the probability that it is not different from zero is less than 0.01.

TABLE 5

Kendall's rank correlation coefficient between relative electrophoretic mutation rates and subunit molecular weights in four groups of Drosophila

Group	n	τ	P
D. willistoni	10	0.689	<0.01
Hawaiian	11	0.527	0.01
D. bipectinata	8	0.071	n.s. *
D. pseudoobscura	10	0.444	0.073†

n is the number of loci involved in the test. These are the enzyme loci for which KOEHN and EANES (1977) list subunit molecular weights. When more than one locus of same enzymatic function was surveyed, the average relative electrophoretic mutation rate was used for the test.

* n.s. = nonsignificant.

† From normal approximation.

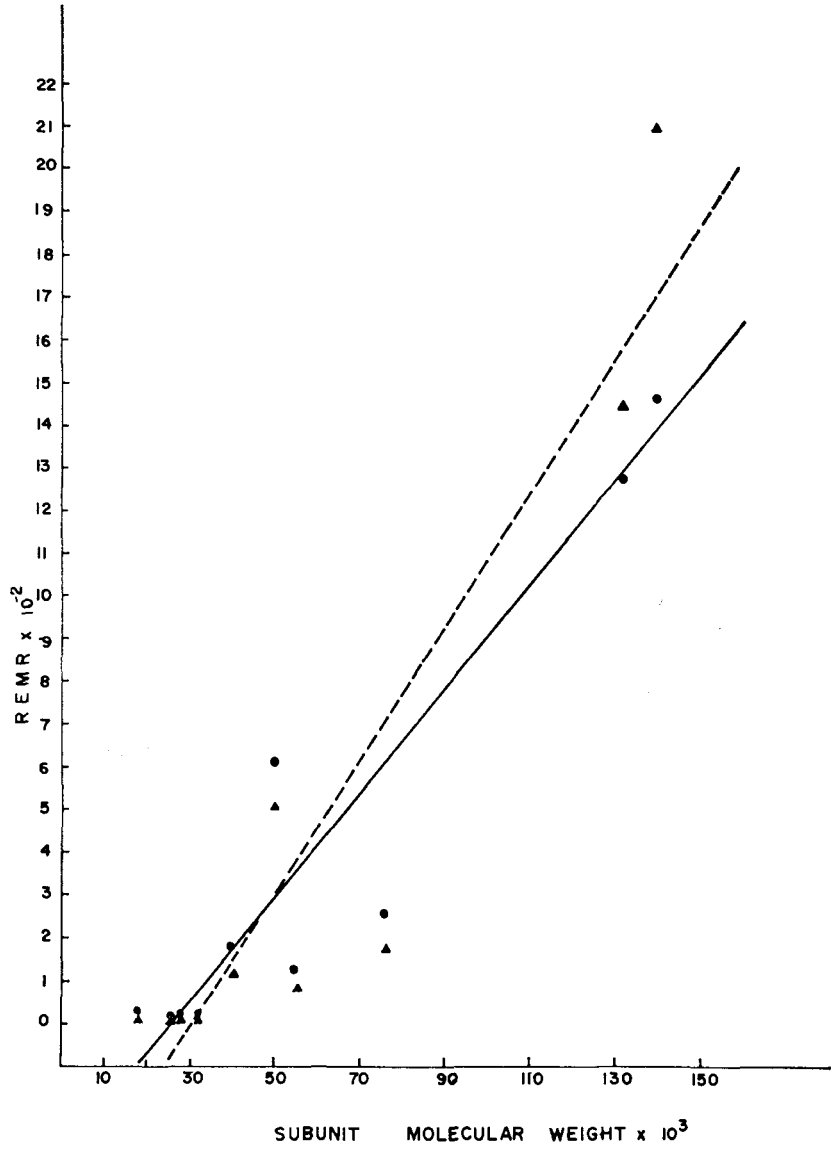

FIGURE 4.—Regression of relative electrophoretic mutation rate against molecular weight of enzyme subunit in *D. willistoni*. Solid line and dots: the infinite-alleles model; broken line and triangles: the step-wise model.

The application of the method to other electrophoretic studies: The analysis described for the *willistoni* group was extended to six more studies. These studies involve organisms that are evolutionarily quite distantly related. Table 1 provides information about these studies. As in the *D. willistoni* group, no locus was completely monomorphic in all six species of the *Drosophila bipectinata* group. Such loci existed, however, in the remaining six groups: there was one

such locus in the Hawaiian species of Drosophila, 17 in the *Drosophila pseudoobscura* species-subgroup, six in species of the genus Menidia (fish), 19 in three subspecies of Mus (house mouse), and eight among karyotype-subspecies of *Thomomys talpoides* (pocket gophers). In all these cases, a heterozygosity value of 0.001 was entered in one taxon group, as discussed above.

The findings from the *D. willistoni* group are reinforced when all seven groups are examined together. The distributions of relative electrophoretic mutation rates for all six studies follow the same pattern, being fairly similar to that for the *D. willistoni* group given in Figure 2. The fit to gamma distribution is given in Table 3. Both the infinite-alleles and the step-wise models produce equally good fits. The main difference between distributions is confined to the first interval (REMR < 0.01), where the majority of relative electrophoretic mutation rates fall. For the first three groups of Drosophila, this interval includes about one-third of all loci (under the infinite-alleles model), or about half the loci (under the step-wise model). The *D. pseudoobscura* subgroup follows a different distribution. A very high percentage of loci in this group show a very low mutation rate when compared to the few loci with high amounts of variability. In all likelihood, this results from the different kinds of loci included in the surveys. In the *D. pseudoobscura* group, PRAKASH (1977) included eleven nonspecific proteins. Most of these loci were invariant. Loci coding for nonspecific proteins were not studied in the other three groups of Drosophila. Among vertebrates the distribution of relative electrophoretic mutation rates if fairly similar, with 50% to 60% (depending on the model) of the loci showing similar electrophoretic mutation rates. Because the number of taxa in most of these groups is small, one expects that these distributions will be seriously affected by stochastic factors. This is particularly true for the *D. pseudoobscura* and mice groups. This can also be the reason why some distributions are more flat than that obtained when *willistoni* and the Hawaiian Drosophilas are pooled together (entry 8 in Table 3).

It was stated earlier that mutation rates estimated from locus heterozygosities may contain a large error resulting from the drift variance of heterozygosity and that an obvious way to reduce this error is to examine a large number of populations or species. Table 6 gives the results one obtains when the data from the *willistoni* group are pooled with those from the *Hawaiian Drosophila*. This pooling results in a group of 13 taxa each surveyed for the same 26 loci. Within the genus Drosophila, the two groups are distantly related, yet mutation rates of homologous loci could still be considered to be the same. More importantly, the two sets of data come from the same laboratory, so that differences in heterozygosity resulting from the electrophoretic techniques used must be minimal. The differences between columns 1 and 2 probably reflect errors in the estimation of relative mutation rates due to such factors as sampling variance of heterozygosity and bottle-neck effect rather than true differences among mutation rates of homologous loci in the two groups. If mutation rates are assumed to be the same in the two groups, then column 3 provides the best estimate of their relative magnitude. The fit of these mutation rates to gamma distribution is given in Table 3.

TABLE 6

Comparison of REMR's (infinite-alleles model) of 26 loci surveyed in willistoni *and Hawaiian Drosophila groups*

Locus	Classification according to function	Relative electrophoretic mutation rates 1	2	3
Lap-5	MS	0.0779	0.0363	0.0652
Acph-1	MS	0.0650	0.0625	0.0650
Ald-1	NR	0.0258	0.0629	0.0378
Fum	NR	0.0283	0.0021	0.0201
G6pdh	R	0.0285	0.0968	0.0486
Odh-1	MS	0.0142	0.0103	0.0129
Xdh	R	0.1602	0.0788	0.1351
Ao-1	MS	0.1472	0.1006	0.1333
Ao-2	MS	0.1519	0.0410	0.1190
Pgm	R	0.0285	0.0453	0.0331
Adk-1	R	0.1449	0.0883	0.1181
Hk-1	R	0.0162	0.0545	0.0375
Hk-2	R	0.0356	0.0544	0.0413
Got	R	0.0074	0.0142	0.0091
Hbkh	R	0.0042	0.0145	0.0072
Est-4	MS	0.0277	0.0305	0.0283
Hk-3	R	0.0066	0.0127	0.0083
Ald-2	NR	0.0029	0.0785	0.0339
Adh	MS	0.0137	0.0058	0.0026
Mdh-2	NR	0.0012	0.0001	0.0012
αGpdh	NR	0.0018	0.0033	0.0018
Idh	NR	0.0014	0.0411	0.0127
Me-1	R	0.0012	0.0398	0.0140
To	MS	0.0032	0.0018	0.0026
Tpi-2	NR	0.0031	0.0015	0.0014
Adk-2	R	0.0012	0.0225	0.0096

1: *D. willistoni* group treated alone; 2: *Hawaiian* group treated alone; 3: *D. willistoni* and *Hawaiian* treated as one group of 13 taxa. MS: multiple substrate; R: regulatory; NR: non-regulatory.

Estimation of relative effective population sizes: Rather than estimating relative electrophoretic mutation rates, the method can be used to estimate ratios of effective population sizes. Table 7 gives the "relative effective population sizes" for the four species and the two semispecies of the *D. willistoni* species-group. The table also gives the mean heterozygosity for each taxon. Again, it may be observed that the ranking of effective population sizes does not match the ranking of heterozygosities. *D. willistoni* ranks fifth in terms of mean heterozygosity, but it ranks first (under the step-wise model) or second (under the infinite-alleles model) in terms of effective population size. *D. nebulosa*, ranking, with *D .p. orinocan*, in first place in terms of heterozygosity, is ranked in fifth place in terms of population size. Again, it may not be fortuitous that the largest value was assigned to *D. willistoni*. This species appears to have a much wider geo-

TABLE 7

Relative effective population sizes for the six taxa of the willistoni *species-group*

Taxon	Mean heterozygosity	Infinite-alleles model	Step-wise model
D. willistoni	0.165	0.19207	0.24979
D. tropicalis	0.138	0.11645	0.09863
D. equinoxialis	0.181	0.15444	0.14481
D. paulistorum Amazonian	0.206	0.18763	0.18029
D. paulistorum Orinocan	0.215	0.19803	0.19840
D. nebulosa	0.214	0.15138	0.12808

graphical distribution, and its populations appear to me much more numerous than is the case for its sibling species (SPASSKY *et al.* 1971).

LEWONTIN (1974) has observed that if one compares effective population sizes derived from mean heterozygosities, one will discover that the sizes of all populations that have been examined for electrophoretic variability fall within a surprisingly narrow range. This is clearly illustrated in Table 7, where it can be seen that the most numerous population exceeds the less numerous one by a factor of only 2.5. Extension to other groups reaffirms this observation. The ratio of the largest effective population size to the smallest varies from 1.7 (Mus) to 6.2 (gophers).

DISCUSSION

The foregoing analysis provides several pieces of information that have a bearing on the controversy regarding the mechanisms that maintain genetic variability in natural populations.

The often-stated view that the difference in the amount of electrophoretic variability between "variable" and "conservative" enzyme loci may reflect differences in mutation rates is quantified. Some of the findings provide support for the theory that allelic enzyme variation is neutral to forces of natural selection. Some other findings lead to the formulation of some rather specific statements, the falsification of which may warrant the rejection of the hypothesis of neutrality.

In the second column of Table 6, 26 loci of Drosophila are classified as loci conding for enzymes with multiple substrates, loci coding for enzymes with regulatory function, and loci with nonregulatory function. This classification was introduced by GILLESPIE and LANGLEY (1974) and JOHNSON (1974), as discussed above. The actual assignment of a locus to one or the other category was based on the same criteria as the ones used by these authors. The dichotomy introduced by GILLESPIE and LANGLEY (1974) will separate enzymes with multiple substrates from all others, while JOHNSON's (1974) hypothesis will require that each of the three categories be compared to each other. The criteria used for the assignment of a locus to one or the other functional category have been questioned (SELANDER 1976). But even if the categorization is valid, the

hypothesis of neutrality would still have no difficulty in explaining the observations. If one examines the relative electrophoretic mutation rates in the fifth column of Table 6, one will find that on the average mutation rates for multiple-substrate enzymes exceed mutation rates of single-substrate enzymes by a factor of 1.7. For the regulatory *versus* nonregulatory enzymes comparison, this factor is 2.7. Obviously, mutation rates do not vary much between functional categories. In contrast, there is a large amount of variability in mutation rate within categories. Given the observation that within a category a given locus consistently exhibits more electrophoretic variability than another locus, one must conclude that most of this variability is real rather than stochastic. The explanation of this observation may lie not with the functional properties of the enzyme and the environmental heterogeneity to which it is exposed, but rather with the probability that a mutational event will cause a change in the protein molecule that is detectable by the screening method in use. It is worth noting that the enormous amount of information about electrophoretic variability now available has been collected with a rather narrow variety of electrophoretic techniques. It must not come as a surprise that, under such more or less invariable techniques, some enzyme loci show more variability than others.

To explain why some groups of functionally related enzymes show more variability than others, one must make a further assumption. As stated in the beginning, this further assumption is that, in general, functional relation implies evolutionary relatedness. From what we know about the evolution of enzyme genes (OHNO 1970; RIGBY, BURLEIGH and HARTLEY 1974; MARKERT 1975), this assumption is justified.

The hypothesis of neutrality may be in difficulty, however, if the difference between the mutation rates assigned to the most variable and most conservative loci is larger than the maximum allowed from consideration of the physical properties of the enzymes. It can be seen from Table 2 that, when heterozygosities are used for the estimation of parameter M, the ratio factor between extreme mutation rates in the *willistoni* group is 135 for the infinite-alleles model, 318 for the step-wise model. The true figure must lie somewhere in between. These figures do not change significantly in the other groups, provided that completely invariant loci are not taken into consideration. At present, it cannot be decided whether a 100-fold to 500-fold difference in electrophoretic mutation rate can be accounted for by differences in the physical properties of the enzyme molecules and by the inability of the experimental procedure to translate these differences into differences in electrophoretic mobility. When M is estimated from EWENS' (1972) formula, the variance in mutation rate becomes much smaller. This means that, depending on the test used, the hypothesis of neutrality could generate quite different predictions regarding the distribution of mutations rates. According to EWENS' (1972) test, neutrality could account for the observations if it were found that mutation rates varied very little. But a variance in mutation rate of that magnitude would be incompatible with neutrality if locus heterozygosities were used for the testing of that hypothesis. Clearly, this is a weakness of the hypothesis

TABLE 8

Ratios between highest and lowest electrophoretic mutation rates

Group	Infinite-alleles model		Step-wise model	
	Invariant loci excluded	Invariant loci included	Invariant loci excluded	Invariant loci included
1. *D. willistoni*	135	——	318	——
2. Hawaiian Drosophila	87	4,394	160	5,285
3. *D. bipectinata*	277	——	696	——
4. *D. pseudoobscura*	111	11,207	587	40,485
5. Menidia	58	13,178	96	22,912
6. Mus	19	3,150	26	4,347
7. Thomomys	185	11,301	413	29,310

of neutrality and, in particular, of the infinite-alleles model for the generation of electrophoretic variability.

A different picture emerges when invariant loci are taken into account. Then, the difference factor is of the order of 10^4 (Table 8). It seems unlikely that the predisposition to electrophoretic mutability or that the discrimination power of the electrophoretic technique may vary so widely from one enzyme to another. One, then, may have to postulate that strong purifying selection is acting on these invariant loci. Such a type of selection was never denied by neutralists and, indeed, it is the most important element in the revised form of the theory of neutrality (OHTA 1974). It is difficult, however, to attach much significance to these high ratios between mutation rates of variable and invariable loci. The arbitrarily assigned value of $0.001/n$ (where n is the number of taxa in the group) for the mean heterozygosity of invariant loci may be very low. It corresponds to heterozygosity that would have been observed if one had examined 2,500 genomes for each taxon and had discovered one variant. Given that the average number of genomes per taxon in the studies cited here is 200, one concludes that the sample size required for the justification of such a small heterozygosity is 12.5 times more than the one actually utilized. The results of Table 6 would suggest that the hypothesis of neutrality could explain the variation of heterozygosities among loci in species of the *D. willistoni* group and in Hawaiian Drosophilas without the need of purifying selection.

The need for a direct estimation of mutation rates has been emphasized by many authors. MUKAI (*cf.* MUKAI 1970; MUKAI and COCKERHAM 1977) pioneered work in that direction, but it will take some time before direct and sufficiently accurate estimates of mutation rates become available in such numbers as to allow one to evaluate directly the competing hypotheses concerning enzyme variability. The present analysis emphasizes the point that such an evaluation need not await the accumulation of directly estimated absolute mutation rates. An estimation of the relative magnitudes of electrophoretic mutation rates of a number of enzyme loci may take us a long way towards rejecting or accepting the hypothesis of neutrality. More important, we do not need to accurately esti-

mate the difference factor between mutation rates; what we need is its upper limit. For example, if we had found that in Drosophila the electrophoretic mutability of *Xdh*, *Ao* or *Adk* may exceed the electrophoretic mutability of *Adh*, *Mdh*, *αGpdh* or *Idh* by a factor of ten at the most then we might have good grounds to reject the hypothesis of neutrality. The results of MUKAI and COCKERHAM (1977) suggest that electrophoretic mutation rates may, indeed, not be as disparate as the hypothesis of neutrality requires. These authors searched for spontaneous electrophoretic mutation events in five enzymes of *D. melanogaster*: α-glycerol-3-phosphate dehydrogenase, malate dehydrogenase, hexokinase, alcohol dehydrogenase, and α-amylase. Of these, the first two may be classified as nonregulatory, the third as regulatory, and the last two as enzymes with multiple substrates (alcohol dehydrogenase is classified by JOHNSON (1974) as regulatory). They observed one electrophoretic mutation even at each of the following enzyme loci: malate dehydrogenase, hexokinase, and α-amylase. No mutation occurred at the remaining two loci. The number of allele generations tested for each locus was about the same. There is no reason to believe, from these data, that the mutation rates are as different as the hypothesis of neutrality would predict. Yet, since no mutation event was observed in two out of five enzymes and the estimation error for the locus-specific mutation rate is large, the matter is far from being resolved.

Reference was made to LEWONTIN's (1974) observation that the neutrality model requires that the effective sizes of all populations surveyed for electrophoretic variation must fall within a narrow range. Pursuing the same line of argument, one might note that the hypothesis of neutrality makes diametrically opposed predictions about electrophoretic mutation rates and effective population sizes. It predicts variability in mutation rates, but invariability for population size. More specifically, electrophoretic mutation rates between variable and conservative loci must be different by a factor of 500 or more, but effective population sizes must be different by a factor of ten or less. This difficulty with the hypothesis of neutrality had also been noted by AYALA *et al.* (1974). If these comparisons of mutation rates and effective population sizes are valid, one may conclude that, in any survey of electrophoretic variability in natural populations, a much larger proportion of the variability among heterozygosities is due to varying mutation rate than is due to varying effective population size.

The observation that the hypothesis of neutrality generates electrophoretic mutation rates that follow the gamma distribution proves little by itself, unless there is some independent evidence that mutation rates are, indeed, distributed according to gamma distribution. NEI, FUERST and CHAKRABORTY (1976) have provided such evidence. If this is established, then the observation given here lends further support for the hypothesis of neutrality.

It was shown that electrophoretic mutation rates predicted from the model of neutrality correlate well with molecular weights of enzyme subunits. This correlation definitely argues in favor of neutrality. But it must be noted that the relevant information comes only from Drosophila. Even within this genus, two-thirds of the enzymes regularly being scored have been ignored in the test, be-

cause the molecular weights of their subunits remain unknown. In addition, the correlation was very good in only two out of four groups of Drosophila species. In the third group the correlation was close to being significant, but in the fourth group the correlation was clearly not significant (Table 5). HARRIS, HOPKINSON and EDWARDS (1977) have compared heterozygosities and molecular weights in human enzymes and found no correlation. Their sample contained 87 enzymes, a number that is eight times larger than the number of Drosophila enzymes tested. It is unfortunate that the method presented here for the estimation of relative electrophoretic mutation rates cannot be applied to human enzymes, because no complete parallel studies of the same enzymes in several human populations (or races) exist. As a result, it cannot yet be determined whether enzyme molecular weights in humans are correlated with mutation rates predicted from the hypothesis of neutrality.

The method for the estimation of relative mutation rates (or effective population sizes) presented here was applied to electrophoretic variation, but it can also be applied to genetic variation recorded by other means. It seems that we are still a long way from the time when amino acid sequencing will be routinely used to record variability in populations. Until such time arrives, the use of less efficient methods will continue, and it will be prudent to speak of electromorphs, thermoelectromorphs (THROCKMORTON 1977), ph-electromorphs, etc., and of "mutation rates" at the corresponding level of screening. This method may be used to estimate the relative sizes of these mutation rates under the hypothesis of neutrality, provided that the infinite-alleles model or the step-wise model provide reasonable approximations to the process by which such genetic variants are generated in natural populations.

A. C. THOMSON and K. L. M. WELDON offered mathematical advice, and P. E. HERTZ offered valuable comments. My special thanks go to MASATOSHI NEI, who provided many valuable comments; his efforts far surpassed the regular duties of an editor. Support was provided from the National Research Council of Canada.

LITERATURE CITED

AYALA, F. J., 1975 Genetic differentiation during the speciation process. Evolutionary Biology **8**: 1–78.

AYALA, F. J., M. L. TRACEY, L. G. BARR, J. F. MCDONALD and S. PEREZ-SALAS, 1974 Genetic variation in natural populations of five Drosophila species and the hypothesis of the selective neutrality of protein polymorphisms. Genetics **77**: 343–384.

EWENS, W. J., 1972 The sampling theory of selectively neutral alleles. Theoret. Popul. Biol. **3**: 87–112.

EWENS, W. J., and M. W. FELDMAN, 1976 The theoretical assessment of selective neutrality. In *Population Genetics and Ecology*, edited by S. KARLIN and E. NEVO. Academic Press, New York, San Francisco, and London.

FUERST, P. A., R. CHAKRABORTY and M. NEI, 1977 Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetcis **86**: 455–483.

GILLESPIE, J. H. and K. KOJIMA, 1968 The degree of polymorphism in enzymes involved in energy production compared to that in nonspecific enzymes in two *Drosophila ananassae* populations. Proc. Natl. Acad. Sci. U.S. **61**: 582–585.

GILLESPIE, J. H., and C. H. LANGLEY, 1974 A general model to account for enzyme variation in natural populations. Genetics **76**: 837–848.

HARRIS, H., D. A. HOPKINSON and Y. H. EDWARDS, 1977 Polymorphism and the subunit structure of enzymes: a contribution to the neutralist-selectionist controversy. Proc. Natl. Acad. Sci. U.S. **74**: 698–701.

HASTINGS, N. A. J. and J. B. PEACOCK, 1975 *Statistical Distributions.* Wiley, New York and Toronto.

JOHNSON, G. B., 1974a Enzyme polymorphism and metabolism. Science **184**: 28–37. ——, 1974b On the estimation of effective number of alleles from electrophoretic data. Genetics **78**: 771–776. ——, 1977 Assessing electrophoretic similarity: the problem of hidden heterogeneity. Ann. Rev. Ecol. Syst. **8**: 309–328.

JOHNSON, M. S., 1975 Biochemical systematics of the atherinid genus *Menidia.* Copeia **1975**: 662–691.

KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

KOEHN, R. K. and W. F. EANES, 1977 Subunit size and genetic variation of enzymes in natural populations of *Drosophila* Theoret. Popul. Biol. **11**: 330–341.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia Univ. Press, New York and London.

LI, WEN-HSING and M. NEI, 1975 Drift variances of heterozygosity and genetic distance in transient states. Genet. Res. (Camb.) **25**: 229–248.

MARKERT, C. L., 1975 Evolution of a gene. Science **189**: 102–114.

MORAN, P. A. P., 1976 Wandering distributions and the electrophoretic profile. II. Theoret. Popul Biol. **10**: 145–149.

MUKAI, T., 1970 Spontaneous mutation rates of isozyme genes in *Drosophila melanogaster.* Drosophila Inform. Serv. **45**: 99.

MUKAI, T. and C. C. COCKERHAM, 1977 Spontaneous mutation rate of enzyme loci in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. U.S. **74**: 2514–2517.

NEI, M., P. A. FUERST and R. CHAKRABORTY, 1976 Testing the neutral hypothesis by distribution of single locus heterozygosity. Nature **262**: 491–493. ——, 1978 Subunit molecular weight and genetic variability of proteins in natural populations. Proc. Natl. Acad. Sci. U.S. **75**: 3359–3362.

NEI, M., R. CHAKRABORTY and P. A. FUERST, 1976 Infinite allele model with varying mutation rate. Proc. Natl. Acad. Sci. U.S. **73**: 4164–4168.

NEVO, E., Y. J. KIM, C. R. SHAW and C. S. THAELER, JR., 1974 Genetic variation, selection and speciation in *Thomomys talpoides* pocket gophers. Evolution **28**: 1–23.

OHNO, S., 1970 *Evolution by Gene Duplication.* Springer-Verlag, Heidelberg.

OHTA, T., 1974. Mutational pressure as the main cause of molecular evolution and polymorphism. Nature **252**: 351–354.

OHTA, T. and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22**: 201–204.

PRAKASH, S., 1977 Genetic divergence of closely related sibling species *Drosophila pseudoobscura, Drosophila persimilis,* and *Drosophila miranda.* Evolution **31**: 14–23.

RIGBY, P. W. J., B. D. BURLEIGH, JR. and B. S. HARTLEY, 1974 Gene duplication in experimental enzyme evolution. Nature **251**: 200–204.

SELANDER, R. C., 1976 Genic variation in natural populations. In: *Molecular Evolution,* edited by F. J. AYALA. Sinauer Ass., Sunderland, Massachusetts.

SELANDER, R. W., W. G. HUNT and S. Y. YANG, 1969 Protein polymorphism and genetic heterozygosity in two European subspecies of the house mouse. Evolution **23**: 653–667.

SELANDER, R. K. and S. Y. YANG, 1969 Protein polymorphism and genic heterogeneity in a wild population of the house mouse (*Mus musculus*). Genetics **63**: 653–667.

SINGH, R. S., R. C. LEWONTIN and A. A. FELTON, 1976 Genetic heterozygosity within electrophoretic "alleles" of xanthine dehydrogenase in *Drosophila pseudoobscura*. Genetics **64**: 609–629.

SPASSKY, B., R. C. RICHMOND, S. PEREZ-SALAS, O. PAVLOVSKY, C. A. MOURAO, A. S. HUNTER, H. HOENIGSBERG, TH. DOBZHANSKY and F. J. AYALA, 1971 Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. Evolution **25**: 129–143.

STEWART, F. M., 1976 Variability in the amount of heterozygosity maintained by neutral mutations. Theoret. Popul. Biol. **9**: 188–201.

THROCKMORTON, L. H., 1977 Drosophila systematics and biochemical evolution. Ann. Rev. Ecol. Syst. **8**: 235–254.

WARD, R. D., 1977 Relationship between enzyme heterozygosity and quaternary structure. Biochem. Genet. **15**: 123–135.

YANG, S. Y., L. L. WHEELER and I. R. BOCK, 1972 Isozyme variation and phylogenetic relationships in the *Drosophila bipectinata* complex. Univ. Texas Publ. **7213**: 213–236.

ZOUROS, E., 1975. Electrophoretic variation in allelozymes related to function or structure? Nature **254**: 446–448. ——, 1976a The distribution of enzyme and inversion polymorphism over the genome of Drosophila: evidence against balancing selection. Genetics **83**: 169–179. ——, 1976b Hybrid molecules and the superiority of the heterozygote. Nature **262**: 227–229.

Corresponding editor: M. NEI

APPENDIX

Let $a_{ij} = n_{e_{ij}} - 1$, where $n_{e_{ij}}$ is defined in (2). The quantity a_{ij} for $i = 1$ to k is measured along axis A_j (Figure 1). Let U_j be the point determined by a_{ij} for $j = 1$ to l, OX be the straight line that minimizes $D = \sum_{i=1}^{k} d^2$ and $\vec{x}$ $(x_1, \ldots x_l)$ be the unit vector of OX, *i.e.* $\sum_{j=1}^{l} x_j^2 = 1$.

From Figure 1 we have:

$$d_i^2 = |\vec{a_i}|^2 \sin^2\theta_i = |\vec{a_i}|^2 - |\vec{a_i}|^2 \cos^2\theta_i = |\vec{a_i}|^2 - (\vec{a_i}\cdot\vec{x})^2 = \\ = \sum_{j=1}^{l} a_{ij} - \left(\sum_{j=1}^{l} a_{ij}x_j \right)^2 \tag{16}$$

and

$$D = \sum_{i=1}^{k} \left[\sum_{j=1}^{l} a^2{}_{ij} - \left(\sum_{j=1}^{l} a_{ij}x_j \right)^2 \right] . \tag{17}$$

We want to minimize D subject to $\sum_{j=1}^{l} x_j^2 - 1 = 0$.

If λ is a Lagrange multiplier, then the function to be minimized is:

$$D' = \sum_{i=1}^{k} \left[\sum_{j=1}^{l} a^2{}_{ij} - \left(\sum_{j=1}^{l} a_{ij}x_j \right)^2 \right] + \lambda \left(\sum_{j=1}^{l} x_j^2 - 1 \right)$$

After setting the first partial derivatives of D' with respect to x_j and λ equal to zero, we obtain

$$\sum_{i=1}^{k} \left(a_{ij} \sum_{j=1}^{l} a_{ij} x_j \right) = \lambda x_j \tag{18}$$

$$\sum_{j=1}^{l} x_j^2 - 1 = 0 \tag{19}$$

System (18) can be written as

$$[A^T A]\vec{x} = \lambda \vec{x} \tag{20}$$

where

$$A_{kxl} = \begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1l} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{k1} & \cdot & \cdot & \cdot & a_{kl} \end{bmatrix} \tag{21}$$

and

$$\text{and} \qquad ||\vec{x}|| = 1 \tag{22}$$

The eigen vector corresponding to the largest eigen value of (20) provides the solution to system (18).