

## Patterns of Molecular Variation. I. Interspecific Comparisons of Electromorphs in the *Drosophila mulleri* Complex

R. H. Richardson<sup>1</sup> and P. E. Smouse<sup>2</sup>

Received 12 Aug. 1975—Final 6 Nov. 1975

---

*The average mobility of electromorphs at an enzyme locus in a single population was defined as the weighted average mobility of the electromorphs in that population, where the electromorph frequencies are the weights. A derivative distance measure was defined whose taxonomic utility was determined in the *Drosophila mulleri* species complex. Most of the variation in this metric was at the interspecific level, primarily among clusters of sibling species. The electromorphs of some loci were equally and regularly spaced, while those of other loci were less regular in their spacing. Overall, these minor perturbations from regular spacing did not noticeably detract from the taxonomic utility of average mobility, and cluster analysis yielded the same taxonomic relationships as more conventional nonmolecular treatments. On the other hand, electromorph spacing may be related to functional constraints on the enzyme molecules. Some possible implications of the results for the modes of selection during evolution of the different enzymes are discussed.*

---

**KEY WORDS:** electromorphs; distance; allozymes; *Drosophila* taxonomy; selection; net charge; ladder model.

### INTRODUCTION

After almost a decade of extensive electrophoretic analysis of protein variation in natural populations, many of the initial biochemical and genetic assumptions concerning "allozymes" are being challenged. One of the most commonly made assumptions in electrophoretic studies is the idea that

---

Supported by NSF Grant 22770, AEC Contract AT-(40-1)-4023, and NIH Career Development Award GM 47350-05 to R.H.R.

<sup>1</sup> Department of Zoology, The University of Texas at Austin, Austin, Texas.

<sup>2</sup> Department of Human Genetics, University of Michigan, Ann Arbor, Michigan.

a single band ("allozyme"), if inherited in Mendelian fashion, represents a single allelic form of the protein. It follows that two individuals or two populations which have a common electrophoretic band share a common allele. In some cases, the homology is thought to cross species lines. There has always been a realization that this assumption might not be "quite" true, but the fiction is so taxonomically convenient that the common practice is to operate on the presumption that "if it migrates the same it is genetically the same."

A number of recent studies have challenged the wisdom of this assumption. Johnson (1975a) has pointed out that most amino acid substitutions should lead to no detectable charge or conformational changes under standard electrophoretic conditions. Based on simple models of charge changes, King (1973) has estimated that only 30% of amino acid changes will be electrophoretically detectable and that only 0.6% will involve a double charge change. The demonstration by Singh *et al.* (1974, 1975) that considerable molecular variability exists within single electrophoretic classes suggests that electrophoresis may indeed uncover only the tip of the variation "iceberg." At the very least, it is time to operationally lay aside the oversimplification of assuming a one-to-one correspondence between electromorph and single allele, unless direct evidence to this effect is available. "Electromorphs" are electrophoretically detectable molecular phenotypes, and contrasts among them are phenotypic contrasts only. The heritability of electromorphs is less than unity, but otherwise is unknown, even though Mendelian segregation of different classes may be observed. King and Ohta (1975) viewed electrophoretically detectable genetic differences as simply ambiguous phenotypic features of the various hidden genotypes. In sympathy to this point of view, we use their term "electromorph" in preference to the term "allozyme."

From a taxonomic viewpoint, the ideal analysis should be based on functionally relevant characteristics. Although some deeper probes into biochemical properties of electrophoretic variants have been made (e.g., Ainsley and Kitto, 1975; Miller *et al.*, 1975; Day *et al.*, 1974; Scandalios *et al.*, 1972), the techniques available do not yet permit this sort of large-scale screening so necessary for taxonomic work. Although new screening procedures are being developed, it appears that for some time yet most of the population data will continue to be derived by electrophoresis.

There are two recent developments which suggest to us, however, that electrophoretic mobility may contain untapped genetic information. First, Bulmer (1971) drew attention to multiple-allele loci, where electrophoretic classes with intermediate mobility tend to be more frequent than those with either lesser or greater mobility. This unimodal distribution suggests that the essential features of the electromorph frequency array may be conveyed by its

average position and its dispersion along the scale of electrophoretic mobility. Second, King (1973) and Nei and Chakraborty (1973) have pointed out that electrophoretic mobility should be evolutionarily conservative, because of the aforementioned tendency for amino acid substitutions to go undetected under standard electrophoretic assay conditions. The observed unimodality of the frequency distribution is thus hypothesized to be due to internal heterogeneity of the electromorph bands, particularly for the modal bands. This pair of observations suggests that one should expect a monotonic increase in average electrophoretic separation with phyletic distance and that populations may diverge by simply shifting the whole mobility distribution along the mobility scale. By using the "average mobility" of a population as a measure of position, and scaling relative to the dispersion within the population, we should be able to gauge the *degree* of difference between populations, even where the difference involves nonoverlapping sets of electromorphs.

Closer examination of the spacing of electromorphs may also be informative. Most of the current theory of electrophoretic evolution assumes a ladder model, i.e., that single amino acid substitutions lead to unitary changes in electrophoretic mobility. Johnson (1974) has pointed out that regular increments are perhaps too much to hope for under a range of assay conditions, and has further shown (Johnson, 1975a) how conformational changes may mimic some of the patterns expected under the ladder model. Since electrophoretic mobility is a function of charge, conformation, molecular weight, and polymeric complexity, there is no guarantee that any regular pattern will emerge. If one were to restrict attention to a narrow taxonomic assemblage, however, it should be possible to avoid the more profound molecular alterations characterizing larger taxonomic groups (e.g., large changes in molecular weight and/or polymeric complexity).

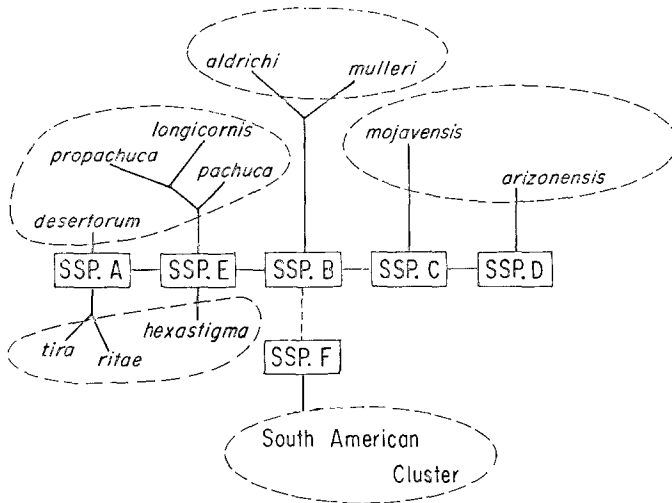
The purpose of this article is to explore the patterns of electrophoretic variation among populations of the *Drosophila mulleri* complex. This complex has been subjected to intensive taxonomic analysis (Sturtevant, 1942; Patterson and Mainland, 1944; Wasserman, 1962), and probable phylogenies have been determined on cytological, morphological, and reproductive criteria. We are able to define several "phyletic clusters" among the species of the complex, and we thus have a range of taxonomic separation available for study. We have reported certain other aspects of this study elsewhere (Richardson *et al.*, 1975), but here we are particularly interested in addressing two questions: (1) Is the average electrophoretic mobility of a population a reliable measure of taxonomic affinity? (2) Are the increments between electromorphs regularly spaced, and, if not, how does this affect the taxonomic analysis? As a subsidiary issue, we shall attempt to indicate means by which the electrophoretic patterns may relate to the interpretation of genetic variation patterns and modes of molecular evolution.

## MATERIALS AND METHODS

### Biological Materials

The data are from 59 populations, drawn from 11 species of the *Drosophila mulleri* complex. We are ignoring the South American members of the group, for which our samples are limited. We are likewise ignoring *D. wheeleri*, *D. spenceri*, and an unnamed *D. mojavenis*-like species from North America. The 11 remaining species may be conveniently grouped into four "phyletic clusters" for analysis (Fig. 1), based on karyotypic and morphological information (Wasserman, 1962; Richardson, unpublished).

Individuals in different clusters are readily distinguished, but individuals from different species of the same cluster (particularly females) are easily confused on morphological criteria. *Drosophila hexastigma* is easily identified by a characteristic pigmentation pattern, but even this exceptional case may be due to a single recessive mutant (W. P. Spencer, reported by Stone, 1955). The male genitalia resemble those of *D. tira* and *D. ritae*. Questionable individuals were identified by hybridization and/or progeny testing. Most of the 59 populations reported here represented wild-caught flies, but a few



**Fig. 1.** Phylogenetic relationships among 11 species determined by polytene chromosome analyses (Wasserman, 1962). Note that *Drosophila longicornis* was changed from Wasserman's figure after he detected an error in the earlier phylogeny (personal communication). Dashed ovals enclose sets of species morphologically very similar, except for *Drosophila hexastigma*. (See text for further explanation.)

are represented by mass cultures, each initiated with 10–30 wild inseminated females.

The sequence of splits shown in Fig. 1 was postulated on the basis of polytene banding patterns. The putative ancestral types (A–F) differ by one or more inversions. Morphological affinities are considered to outweigh the karyotypic evidence for *D. hexastigma* (placed in the *longicornis* cluster). Relations within and between the *tira* and *longicornis* clusters are still somewhat tenuous, and they are both karyotypically and morphologically much closer than is either to the more distinct *mojavensis* or *mulleri* clusters. Although Wasserman (1962) monophyletically differentiates these clusters on inversions, morphologically and ecologically one might consider the combination of ancestral types (A–E) to be a single polytypic species. The species clusters thereby represent polyphyletic derivatives from this ancestral species which was chromosomally polymorphic.

### Electrophoretic Assay

We have followed standard horizontal starch gel electrophoretic assay procedures for all systems (Johnson *et al.*, 1966, 1968; Kojima *et al.*, 1970). The loci reported here are malate dehydrogenase (*Mdh*), octanol dehydrogenase (*Odh*), alcohol dehydrogenase (*Adh*), glutamic-oxaloacetic transaminase (*Got*), aldehyde oxidase (*Ao*), phosphoglucumutase (*Pgm*), and esterase C (*Est*), all of which exhibit three or more electromorphs within the complex. We have also examined the electrophoretic patterns for acid phosphatase, alkaline phosphatase, leucineaminopeptidase, maleic enzyme,  $\alpha$ -glycerophosphate dehydrogenase, esterase D, esterase E, hexokinase, fumarase, glucose-6-phosphate dehydrogenase, 6-phosphoglucose dehydrogenase, phosphoglucose isomerase, aldolase, xanthine dehydrogenase, and isocitrate dehydrogenase. We excluded loci for which we have incomplete data for any species. This situation occurs when staining assay for a species fails. Also, monomorphic and dimorphic loci were excluded.

We have standardized the mobility “name” of each electromorph by dividing its migration on the gel by that of a reference morph run in multiple replicates on the same gel. Differences between morphs with very similar mobilities were confirmed in “side-by-side” comparisons on a single gel. Although the mobility ratio of a morph to the reference varied over a range of about 0.03–0.20 among gels with different lots of starch, the relative spacing between morphs on the same gel was insensitive. Each relative mobility name was determined from an average of more than 100 measurements distributed among 20 or more gels involving several lots of starch. No standard error for a name exceeded  $10^{-1}$ . All standard errors for relative spacing between adjacent morphs were less than  $10^{-3}$ . For routine scoring of

morphs, they were categorized with less than 1% of them being ambiguous. These were omitted, so that probably there were no misidentifications into adjacent mobility classes. Some of the ambiguities may have represented rare morphs which were not isolated in culture and named, although most ambiguities clearly reflect technical problems, such as smudges or streaks.

We should mention at this point that, while the mathematical procedures which follow may be used with any definition of mobility, careful attention to the measure itself may be fairly important. For systems where multiple bands appear in a "homozygote," only one "primary" band was used for each type of "homozygote," and only the corresponding two bands were used for the heterozygotes. (Because of incomplete detection of alleles, "homozygotes" may be heterozygous for two alleles producing different proteins, but one electromorph.) Wherever possible, the reference morph was obtained from standard laboratory stock. Johnson (1975*b*) suggested that use of "off-the-shelf" proteins is preferable for the same purpose. In addition, he brackets the relative mobilities on a gel by simultaneously using two proteins with very different molecular weights. This is a superior procedure, but we found it unnecessary for our surveys.

### Quantitative Analysis

Consider a genetic locus with  $K$  electromorphs ( $G_1, \dots, G_K$ ). We do not know whether an electromorph represents a single allele in any particular case, although one suspects internal heterogeneity in general. We are, however, primarily interested here in mobility as a phenotype—a "property" or "characterization" of the class of alleles coding for proteins within an electromorph. Therefore, for our purposes, that which migrates the same is the same. The average mobility of the  $i$ th population is estimated to be

$$\tilde{\mu}_i = a_1 \tilde{P}_{1i} + \dots + a_K \tilde{P}_{Ki} = \underline{A}' \tilde{\underline{P}}_i \quad (1)$$

where the  $\tilde{P}$ s are the observed frequencies of the various electromorphs and the  $a$ s are the relative mobility values described in the previous section.

The variation among populations is determined by computing

$$\Delta^2(T) = \sum_{i=1}^I (\tilde{\mu}_i - \bar{\mu})^2 / \tilde{\sigma}^2 \quad (2)$$

where  $\bar{\mu}$  is the overall average mobility for the ( $I = 59$ ) populations and  $\tilde{\sigma}^2$  is the average within population variance in mobility. It is easily shown that  $\Delta^2(T)$  can be subdivided to yield

$$\Delta^2(T) = \Delta^2(P) + \Delta^2(S) + \Delta^2(C) \quad (3)$$

where  $\Delta^2(P)$  is the variation among populations (within species),  $\Delta^2(S)$  is the variation among species (within clusters), and  $\Delta^2(C)$  is the variation among

phyletic clusters of species. We may also compute the difference between two populations ( $i$  and  $i'$ ) as a distance measure

$$D^2(i, i') = (\tilde{\mu}_i - \tilde{\mu}_{i'})^2 / 2 \tilde{\sigma}^2 \quad (4)$$

and a corresponding measure for the difference of a pair of species. The derivations of these various measures and the relationships among them are developed in the Appendix. We have shifted from a multinomial distribution with  $K$  classes into an approximation of a standardized univariate normal distribution. This translation contains almost as much information as the original distribution, minimizes sample vagaries of rare alleles, more accurately reflects the actual genotypic to phenotypic relationship when considering electromorphs (especially for several species), and opens up a considerable body of variation analysis patterned after traditional analysis of variation techniques.

The variation measures described above are all defined for a single locus but may all be extended to multiple loci in straightforward fashion. This should allow us to describe the mobility differences among populations for a battery of loci. As a general proposition, we should allow for linkage disequilibria among loci (within populations), but if loci segregate independently the variation analysis simplifies to the sum of separate analyses for the various loci. The work of Langley *et al.* (1974), Mukai *et al.* (1971), Kojima *et al.* (1970), and Charlesworth and Charlesworth (1973) suggests that observed linkage disequilibria are generally small, except for closely linked loci or loci closely associated with inversion polymorphisms. The linkage relations of the loci reported here have not been worked out as yet, but we have evidence for no strong linkage disequilibria (Makela, 1975; Makela and Richardson, unpublished). We ignore this potential complication. In addition, the inversions distinguishing different species are essentially monomorphic within species, so that no serious complications are likely to arise from this source. We have therefore added distance and variation measures across loci; the results should serve as a very good approximation to reality.

## RESULTS AND CONCLUSIONS

### Variation Analysis

The partitioned  $\Delta^2$  analysis is presented in Table I, where each entry is standardized as a decimal fraction of the total variation for the locus (column) in question. The total variation for each locus is presented at the bottom of the table, both in absolute terms and as a fraction of the variation for all seven loci. Several points are evident from the table: (1) The several loci vary considerably in their total dispersion, ranging from 54.23 for *Odh* to 4361.86 for

Table I.  $\Delta^2$  Partition by Locus

Source of variation	Degrees of freedom	Loci (K)										Total
		Mdh (4)	Odh (11)	Adh (7)	Pgm (7)	Got (5)	Est (13)	Ao (10)				
Among species	10	0.9687	0.9129	0.8870	0.9264	0.9940	0.9910	0.8686			0.9670	
Among clusters	3	0.9655	0.6108	0.7337	0.7949	0.5110	0.9747	0.1815			0.7940	
Within clusters	7	0.0032	0.3021	0.1533	0.1315	0.4830	0.0163	0.6871			0.1730	
<i>long-prop-pach-desr</i>	3	0.0008	0.2052	0.1127	0.1286	0.0011	0.0010	0.4143			0.0447	
<i>hexa-ritae-tira</i>	2	0.0022	0.0614	0.0000	0.0000	0.0018	0.0009	0.0878			0.0043	
<i>mull-aldr</i>	1	0.0001	0.0021	0.0000	0.0028	0.4801	0.0009	0.0057			0.1064	
<i>moja-artiz</i>	1	0.0001	0.0335	0.0406	0.0001	0.0000	0.0135	0.1793			0.0176	
Within species	48	0.0313	0.0871	0.1130	0.0736	0.0060	0.0090	0.1314			0.0330	
<i>mojavensis</i>	15	0.0020	0.0206	0.1083	0.0014	0.0000	0.0046	0.0675			0.0169	
<i>arizonensis</i>	5	0.0016	0.0306	0.0001	0.0675	0.0000	0.0023	0.0000			0.0096	
<i>mulleri</i>	5	0.0000	0.0040	0.0000	0.0000	0.0027	0.0008	0.0302			0.0021	
<i>aldrichi</i>	8	0.0163	0.0001	0.0000	0.0000	0.0002	0.0003	0.0076			0.0008	
<i>longicornis</i>	4	0.0005	0.0027	0.0000	0.0001	0.0001	0.0001	0.0141			0.0005	
<i>propachuca</i>	6	0.0107	0.0174	0.0046	0.0000	0.0000	0.0007	0.0110			0.0016	
<i>pachuca</i>	2	0.0002	0.0025	0.0000	0.0046	0.0030	0.0001	0.0010			0.0014	
<i>tira</i>	3	0.0000	0.0092	0.0000	0.0000	0.0000	0.0001	0.0000			0.0001	
Among populations	58	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000			1.0000	
Total $\Delta^2$		221.68	54.23	995.78	1107.40	1986.00	4361.86	318.05			9045.00	
Fraction of total variation		0.0245	0.0060	0.1101	0.1124	0.2196	0.4822	0.0352			1.0000	



*Est*. (2) In spite of this fact, all loci show a preponderance of variation among species (86.86–99.40%). (3) The percentage of the “among-species” component which separates phyletic clusters varies from 18.15% for *Ao* to 97.47% for *Est*. (4) Individual loci show large variation within particular clusters (*Odh* in the *longicornis* cluster, *Got* in the *mulleri* cluster, and *Ao* in the *longicornis* and *mojavensis* clusters). These loci are especially effective in discriminating among sibling species. (5) In addition, certain loci show large amounts of variation among populations (within species), but as this is the subject of a later report we shall not discuss it further here. For the total of seven loci reported, 79.4% of the variation is among clusters, 17.3% is among species (within clusters), and 3.3% is among populations (within species).

The overall conclusion to be drawn from Table I is that mobility *per se* is a useful taxonomic metric. Furthermore, since our clusters are based on morphological similarity, they correspond to the sibling species pairs studied by Hubby and Throckmorton (1968), and Ayala *et al.* (1974). Our results support their conclusion that sibling species are nevertheless much more distinct than are populations within a species, but less different than non-sibling species. The utility of mobility as a taxonomic measure is apparent.

The clear-cut results of Table I were obtained by establishing a set of phyletic clusters on independent criteria. It remains to be shown whether these same clusters would be derived from the mobility measures alone. Table II presents a complete listing of pairwise distances between species, in terms of both  $D^2$  and  $\delta^2$  (see Appendix). Using the matrix of unweighted pairwise distance measures ( $\delta^2$ ), we may construct a taxonomic network to represent the relationships among the species. A number of procedures are available for this purpose; the one most in keeping with our treatment of mobility is the hierarchical splitting procedure of Edwards and Cavalli-Sforza (1965), where the variation between groups is maximized at each split and the within-groups variation is minimized. Given this splitting sequence, we can construct a “minimum string” network (Thompson, 1973). The resulting network is shown in Fig. 2, and yields precisely the same four clusters shown in Fig. 1. Clearly then, the utility of the mobility measures does not depend on knowing the clusters beforehand.

Before passing on to other matters, a comment is in order relative to mobility as a taxonomic metric. For a locus with  $K$  electromorphs, there are  $(K-1)$  pieces of information, and these are all intercorrelated, because of multinomial sampling. The values of  $K$  encountered in this study range from 4 for *Mdh* to 13 for *Est*. For each locus, we have extracted a single piece of information (*average mobility*) and have ignored the rest. We mentioned earlier that we have ignored the dimorphic loci for the analyses reported here. A dimorphic locus yields only one piece of information (identical with average mobility), and there is nothing to be gained by using mobility

Table II. Pairwise Species Differences

	<i>mojaensis</i>	<i>arizonensis</i>	<i>mulleri</i>	<i>aldrichi</i>	<i>desertorum</i>	<i>pachuca</i>	<i>propachuca</i>	<i>longicornis</i>	<i>hexastigma</i>	<i>riteae</i>	<i>tira</i>
<i>Mojaensis</i>	—	158.48 <sup>a</sup>	2702.70	1854.66	435.45	950.41	1906.43	1573.70	334.44	368.71	1337.96
<i>Arizonensis</i>	36.32 <sup>b</sup>	—	2298.06	1677.72	553.72	1165.97	1825.78	1693.25	478.40	470.16	1499.01
<i>Mulleri</i>	619.37	766.02	—	962.42	238.22	308.76	646.29	714.77	502.03	579.06	1565.50
<i>Aldrichi</i>	321.99	466.03	267.34	—	271.37	396.67	843.70	894.83	278.24	308.46	873.50
<i>Desertorum</i>	462.66	646.01	277.92	301.52	—	111.42	101.07	128.10	118.96	198.50	239.81
<i>Pachuca</i>	376.20	582.98	154.38	176.30	148.00	—	223.67	182.62	195.03	314.61	563.41
<i>Propachuca</i>	391.50	565.12	200.04	214.27	115.50	106.51	—	62.13	118.00	159.22	404.13
<i>Longicornis</i>	413.10	620.86	262.08	278.39	153.70	97.40	21.30	—	87.41	152.81	313.37
<i>Hexastigma</i>	355.34	558.13	585.70	309.16	237.90	260.04	134.86	104.89	—	27.92	11.78
<i>Riteae</i>	391.75	548.52	675.57	342.73	397.00	419.48	181.96	183.37	55.84	—	24.19
<i>Tira</i>	418.11	624.59	652.29	315.43	299.70	328.66	158.76	141.02	14.72	30.24	—

<sup>a</sup>  $D^2$ , as defined in equation (A10).<sup>b</sup>  $\delta^2$ , as defined in equation (A11).

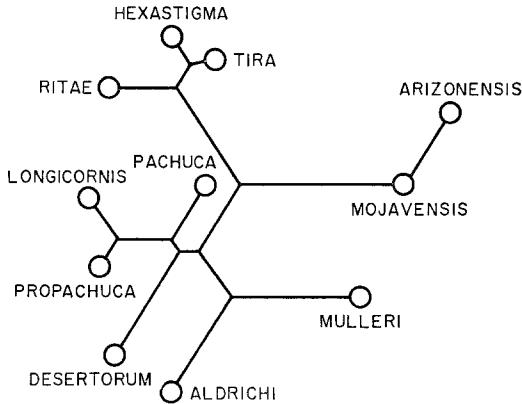


Fig. 2. Minimum string network of species based on average electrophoretic mobility measurements combined over seven loci [*Est-C*, *Ao*, *Adh*, *Pgm*, *Mdh*, *Mdh* (NAD), *Got*].

for such a locus. Since our objective is to evaluate the utility of this dimensional reduction, we have excluded dimorphic loci from consideration. In principle, however, there is nothing to prevent one from computing the mobility measures for such loci; for usual taxonomic purposes, one would generally do so. Average mobility works as well as it does for two reasons: (1) The measure is a linear combination of the mobilities of individual electromorphs, and the scale used is a natural one, rather than being contrived from the analysis itself, as is usual in discriminant analysis. Populations do appear to shift mobility distributions up and down the electrophoretic scale, as will be shown in the next section. (2) The measure is relatively insensitive to sampling variation, since the central classes predominate. The high- and low-mobility electromorphs are generally rare and are likely to be missed in small samples. Since the estimated weights for the mobility values of particular morphs are their estimated frequencies, and the extremal classes are almost inevitably rare, their presence or absence in the sample makes very little difference in the values of  $\tilde{\mu}_i$  and  $\tilde{\sigma}^2$ .

### Incremental Mobility Shifts

The frequency histograms for the electromorphs of *Ao* are shown in Fig. 3. It is noteworthy that the increments between adjacent electromorphs are equal. When variation occurs within a species, only adjacent morphs are encountered, much as expected from standard ladder models with unit

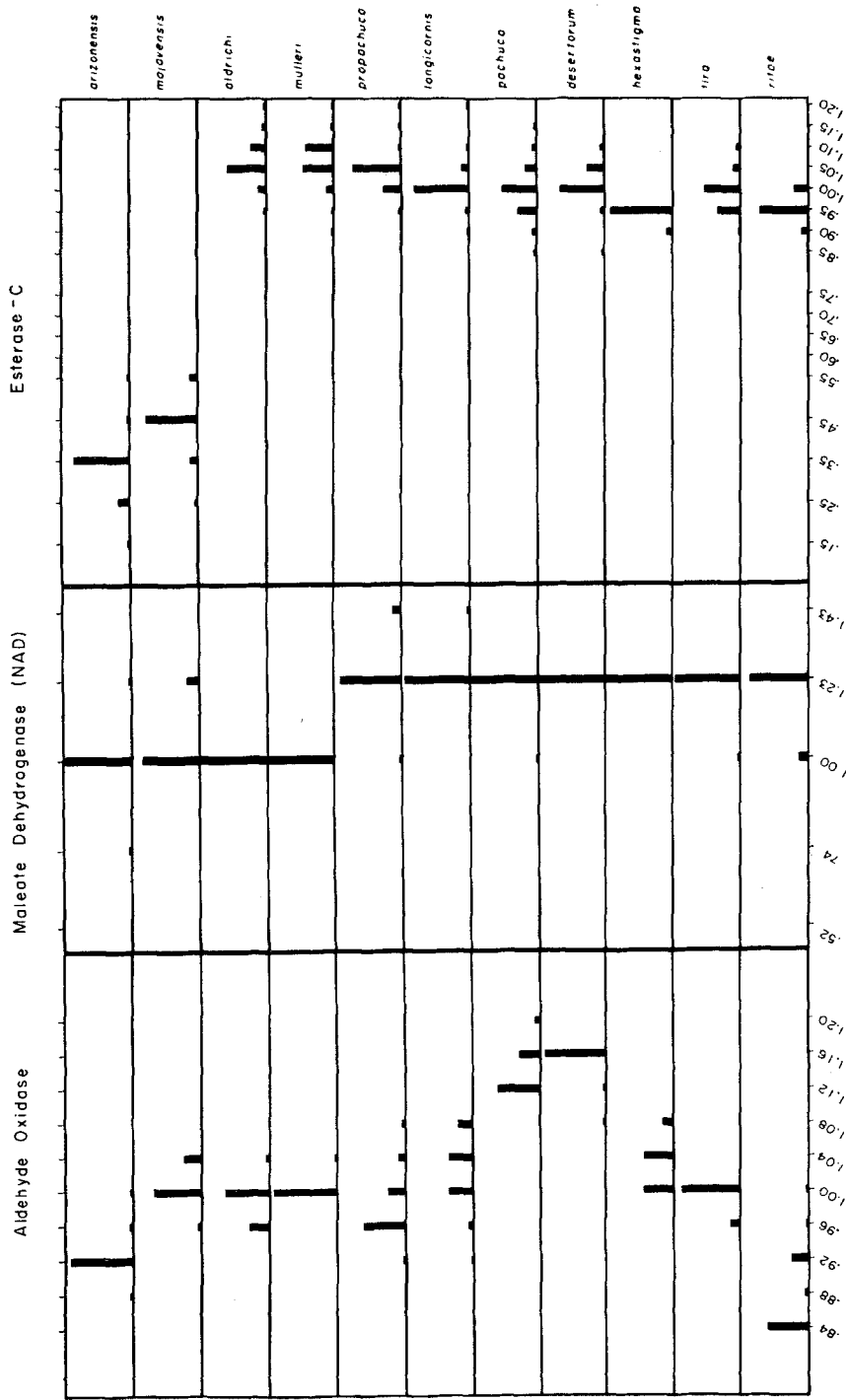


Fig. 3. Frequencies and average mobilities of electromorphs for *Al*, *Mdh*, and *Est*. Electromorph classes known in closely related species are indicated for *Mdh* and *Est*.

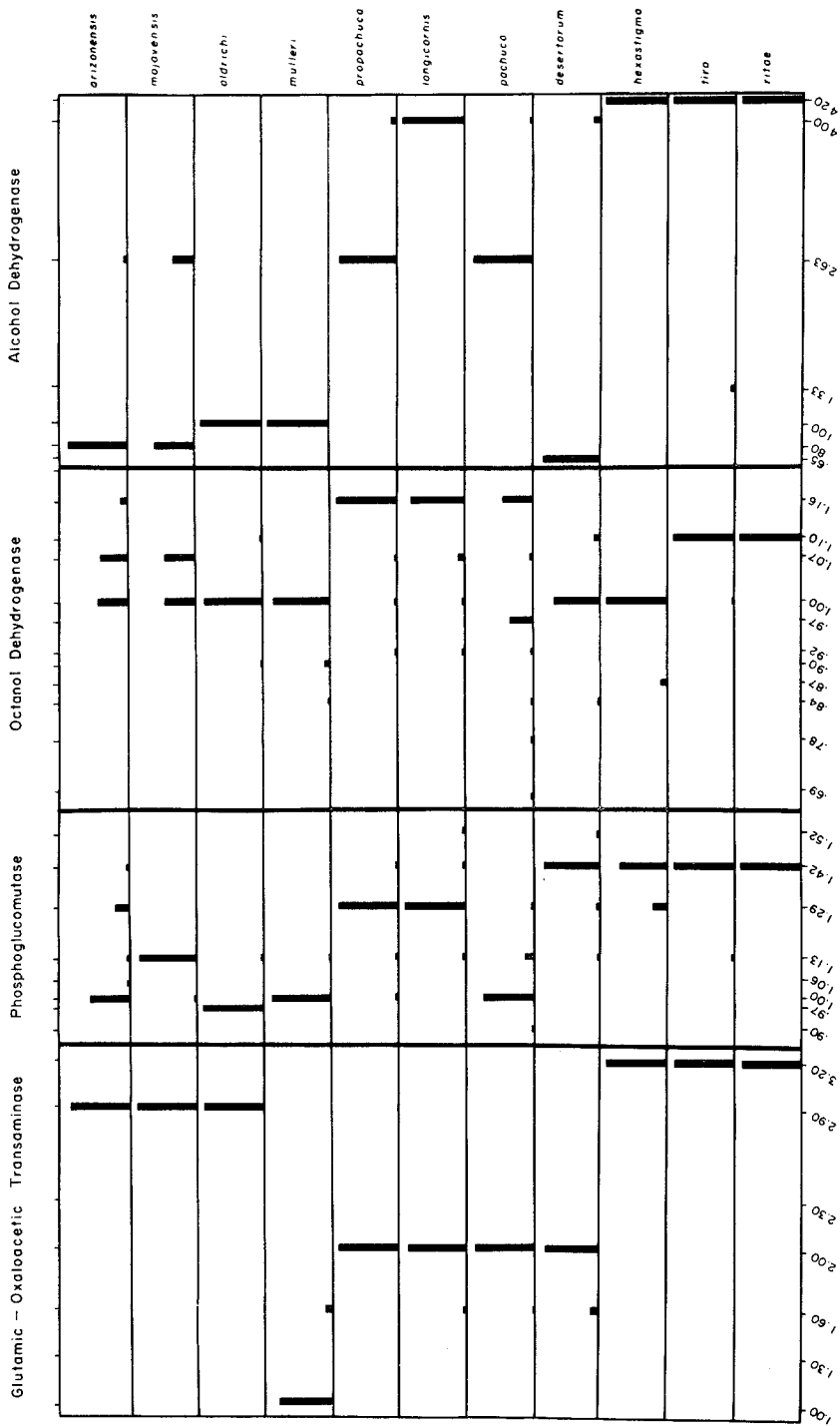


Fig. 4. Frequencies and average mobilities of electromorphs for *Goi*, *Pgm*, *Odh*, and *Adh*. An electromorph class found in a closely related species is indicated for *Goi*.

charge changes separating the bands (King, 1973; Nei and Chakraborty, 1973). In addition, species within a cluster share the same electromorphs or exhibit adjacent morphs in the sequence. The same sort of ladder pattern is exhibited by *Mdh* (Fig. 3), and while conformational alterations cannot be ruled out (Johnson, 1975a) we resist invoking them for the present. A slightly different pattern is evident for the *Est* locus (Fig. 3), which shows two separate ladders, one of five rungs and one of eight rungs. Within each ladder, species have mobility distributions which cascade along the phylogeny. The large separation between the two ladders is accompanied by a change in step size, and these two observations suggest a major structural alteration accompanying the discontinuity. The loci shown in Fig. 3 are the sort which should yield reasonable taxonomic results when average mobility is used as a metric.

The mobility shifts of the other four systems deviate in one respect or another from the ladder model. The *Got* locus tends to be essentially monomorphic within a cluster, although the electromorph for the *tira* cluster is out of register (Fig. 4). Electromorphs are not equally spaced for *Pgm*, but the spacing is quite consistent within a cluster (Fig. 4). In addition, different clusters have different morphs; the mobility distribution seems to cascade along the phylogeny. A similar pattern is observed for *Odh*, but overall separation is less apparent, and more "out-of-register" morphs are encountered (Fig. 4). The situation is least clear for *Adh* (Fig. 4), where each cluster is characterized by a unique set of electromorphs, and no regularity of spacing is evident. In spite of the fact that none of these last four loci meets the ladder ideal, they are all useful for taxonomic analysis, primarily because clusters are more easily distinguished than species (within clusters) or populations (within species). Individual loci introduce irregularities, but when the loci are used in combination, the overall pattern is robust to these minor perturbations.

## DISCUSSION

Day *et al.* (1974) and Ainsley and Kitto (1975) have shown that different electrophoretic forms of ADH (in *D. melanogaster*) are characterized by different kinetic parameters. The electrophoretic differences encountered in the present study are somewhat larger than those encountered within *D. melanogaster*, and the irregularities of spacing suggest that more is involved than simple charge changes. It should prove useful to examine the kinetic properties of this more disparate set of electromorphs. In a more general vein, our results show that there is more variation (and a larger number of electromorphs) among species and clusters than within a single species. When studying kinetic differences among electromorphs, it might be better to cast a wider taxonomic net.

The large shift in the EST electromorphs is also worth a closer look. It is noteworthy that EST is located in the gut (Kambysellis *et al.*, 1968), and may well be a digestive enzyme. The electrophoretic shift coincides with an abrupt change in larval substrates. Fellows and Heed (1972) have shown that the species of the *mojavensis* cluster utilize giant cacti: organ pipe (*Lamprosema coccineum*), agria (*Machaerocereus*), and cina (*Rathbunia*), and we have reared *D. arizonensis* from garambullo (*Myrtillocactus*). Species from other clusters are routinely reared from prickly pear (*Opuntia*). The coincidence is intriguing, to say the least. Some effort to determine the substrate affinities of these electromorphs would seem to be in order.

It is still not clear what we may infer from a long series of equally spaced electromorphs, such as evidenced by *Ao* or *Est*. The regularity of the phenomenon and the unimodality of the within-species frequency distributions are so easily accounted for on the theoretical basis of neutral mutation (King, 1973; Nei and Chakraborty, 1973) that Bulmer's (1971) suggestion of a selective explanation has not been well received. Johnson (1975a) has suggested, however, that much of the electrophoretically expressed variation reflects conformational alterations of the enzyme. Whether one can infer adaptive differences associated with these changes remains to be seen. It is worth noting here that Scandalios *et al.* (1972) found a monotonic relationship between the electrophoretic mobility of maize catalase electromorphs and the heat stability and specific activity of those same electromorphs. While it is not possible to generalize from this single example, we submit that a good deal more work along these lines is in order.

We should also mention the fact that AO consists of two separate polypeptides (Mahler and Cordes, 1966, p. 653), which are under the control of separate loci. One of these gene products is electrophoretically invariant, and is shared by other enzymes (e.g., XDH, XO) (Collins *et al.*, 1971). The invariance of this shared polypeptide suggests that it may be under severe molecular constraints, imposed by the necessities of properly combining with diverse other molecular species. In turn, however, it is possible that the invariant polypeptide imposes constraints on these variable polypeptides. (Only AO is of concern here.) Whether this would lead to regular increments between adjacent electromorphs is not clear, but perhaps the laddering effect really does represent single charge changes between adjacent electromorphs.

Clearly, loci differ in their modes of evolution. Since molecular constraints vary among enzymes (Clarke, 1970), these molecular aspects must be considered if one is to understand the adaptive significance of electrophoretic variation. It is difficult to generalize electrophoretic studies, except by their biochemical properties, because the loci examined constitute a sample from an unknown reference population of genes. As in any study of enzyme electrophoretic variation, our sample derives only from stable soluble

enzymes. Furthermore, enzymes are derived from mRNA which is transcribed from only 1–2% of the DNA in *Drosophila* (based on Lengyel and Penman, 1975). While evidence of the adaptive importance of electrophoretically detectable variation first appeared a decade ago (Kojima and Yarbrough, 1967; Richardson and Johnson, 1967), extrapolating these results to all electrophoretic variation has followed a tortuous path. Molecular diversity alone reveals little about molecular function, and a better measure is needed to relate population variation to adaptation. We submit that mobility patterns of electromorphic variation among populations provide at least some intriguing clues regarding the potential adaptive significance of this molecular variation.

### APPENDIX

Given a genetic locus with  $K$  electromorphs ( $G_1, \dots, G_K$ ) and a set of relative mobility values  $\underline{A}' = (a_1, \dots, a_K)$ , we define the average mobility of the  $i$ th population to be

$$\mu_i = a_1 P_{1i} + \dots + a_K P_{Ki} = \underline{A}' \underline{P}_i \quad (\text{A1})$$

where the  $P_{Ki}$  are the frequencies of the electromorphs in the  $i$ th population.

The variation in mobility within the  $i$ th population can be obtained as follows. The mobilities of two electromorphs, drawn at random from the population, are correlated to some extent, because of multinomial distribution of the morphs. The frequency variation within the sample is defined by a ( $K \times K$ ) covariance matrix

$$\Sigma_i = \begin{bmatrix} P_{1i}(1-P_{1i}) - P_{1i}P_{2i} & \dots & -P_{1i}P_{Ki} \\ -P_{1i}P_{2i} & P_{2i}(1-P_{2i}) & \dots & -P_{2i}P_{Ki} \\ \vdots & \vdots & \ddots & \vdots \\ -P_{1i}P_{Ki} & P_{2i}P_{Ki} & \dots & P_{Ki}(1-P_{Ki}) \end{bmatrix} \quad (\text{A2})$$

The mobility variation within the  $i$ th population is given by

$$\begin{aligned} \sigma_i^2 &= \underline{A}' \Sigma \underline{A} = \sum_{k=1}^K a_k^2 P_{ki} (1-P_{ki}) \\ &\quad - \sum_{k=1}^{K-1} \sum_{k'=k+1}^K a_k a_{k'} P_{ki} P_{k'i} \\ &= \sum_{k=1}^K a_k P_{ki} (a_k - \mu_i) \end{aligned} \quad (\text{A3})$$



In practice,  $\mu_i$  and  $\sigma_i^2$  must be estimated from sample data. If  $P_{ki}$  is estimated by  $\tilde{P}_{ki} = (X_{ki} \div N_i)$ , the observed frequency ( $X_{ki}$  is the number of individual morphs of the  $k$ th type recovered from a sample of size  $N_i$ ), then we simply estimate  $\mu_i$  and  $\sigma_i^2$  by computing

$$\begin{aligned} \tilde{\mu}_i &= \underline{A}' \tilde{P}_i = a_1 \tilde{P}_{1i} + \dots + a_K \tilde{P}_{Ki} \\ \tilde{\sigma}_i^2 &= \underline{A}' \Sigma A = \sum_{k=1}^K a_k \tilde{P}_{ki} (a_k - \tilde{\mu}_i) \end{aligned} \tag{A4}$$

The taxonomic analysis consists of partitioning the total variation among populations into various components. This total variation is estimated by

$$\Delta^2(T) = \sum_{i=1}^{I=59} (\tilde{\mu}_i - \bar{\mu})^2 / \tilde{\sigma}^2 \tag{A5}$$

where  $\bar{\mu}$  and  $\tilde{\sigma}^2$  are defined as

$$\bar{\mu} = \left[ \sum_{i=1}^I \tilde{\mu}_i \div I \right] \quad \tilde{\sigma}^2 = \frac{\sum_{i=1}^I N_i \tilde{\sigma}_i^2}{\sum_{i=1}^I N_i} \tag{A6}$$

We use  $\tilde{\sigma}^2$  to avoid sampling difficulties with the  $\tilde{\sigma}_i^2$ . By ignoring the differences of within-population variance, we obtain a stable and well-estimated standard of reference ( $\tilde{\sigma}^2$ ). All population differences are scaled relative to this average within-population variance. The total variation among populations  $\Delta^2(T)$  can be simply related to a  $\chi^2$  test with  $(I-1)$  degrees of freedom. Although we shall not be involved with statistical testing, the analogy is useful, and we shall follow it.

The total variation  $\Delta^2(T)$  may be partitioned as follows:

$$\begin{aligned} \Delta^2(T) &= \sum_{i=1}^I (\tilde{\mu}_i - \bar{\mu})^2 / \tilde{\sigma}^2 \\ &= \sum_{i=1}^I (\tilde{\mu}_i - \bar{\mu}_s)^2 / \tilde{\sigma}^2 + \sum_{i=1}^I (\bar{\mu}_s - \bar{\mu}_c)^2 / \tilde{\sigma}^2 + \sum_{i=1}^I (\bar{\mu}_c - \bar{\mu})^2 / \tilde{\sigma}^2 \\ &= \Delta^2(P) + \Delta^2(S) + \Delta^2(C) \end{aligned} \tag{A7}$$

where  $\bar{\mu}_s$  and  $\bar{\mu}_c$  are the unweighted mobility averages for a species and cluster, respectively, and are defined as is  $\tilde{\mu}$  in (A4). The variance measures  $\Delta^2(P)$ ,  $\Delta^2(S)$ , and  $\Delta^2(C)$  represent the dispersions among populations (within species), among species (within clusters), and among clusters, respectively. (They have 48, 7, and 3 degrees of freedom, respectively, for our data set, given the number of species in the various clusters.)

One other situation is of some interest. The difference in average mobility between a pair of populations is measured by the distance metric

$$D^2_{(ii')} = (\tilde{\mu}_i - \tilde{\mu}_{i'})^2 / 2\tilde{\sigma}^2 \tag{A8}$$

which is simply (A5) with  $I = 2$ . The computational aspects of (A7) can become quite cumbersome, particularly when a number of different combinations of populations are investigated. Fortunately, (A8) can be used to accomplish the same objective. First, we construct a matrix of all possible  $[I(I-1)/2]$  pairwise  $D^2$  values. The dispersion among any desired set of  $J$  populations may be computed from the relation

$$\Delta^2(J) = \left(\frac{2}{J}\right) \sum_{i=1}^{J-1} \sum_{i'=i+1}^J D^2(ii') \quad (\text{A9})$$

We may view  $\Delta^2(J)$  as the sum of squared distances of the  $J$  populations from their centroid. The partition of (A7) is then obtained by subtraction.

This device is especially useful in computing pairwise species differences. Given two species ( $S_1$  and  $S_2$ ), with  $J_1$  and  $J_2$  sampled populations, respectively, the squared distance between them is seen to be

$$\begin{aligned} D^2(S_1 \text{ vs. } S_2) &= \Delta^2(J_1 + J_2) - \Delta^2(J_1) - \Delta^2(J_2) \\ &= \left[ \frac{J_1 J_2}{J_1 + J_2} \cdot \frac{(\bar{\mu}_1 - \bar{\mu}_2)^2}{\bar{\sigma}^2} \right] \end{aligned} \quad (\text{A10})$$

where  $\bar{\mu}_1$  and  $\bar{\mu}_2$  are the unweighted mean mobilities for the two species. To compare squared distances among species (unweighted by numbers of sampled populations), we compute

$$\begin{aligned} \delta^2(S_1 \text{ vs. } S_2) &= [(J_1 + J_2)/(J_1 J_2)] \cdot D^2(S_1 \text{ vs. } S_2) \\ &= (\bar{\mu}_1 - \bar{\mu}_2)^2 / \bar{\sigma}^2 \end{aligned} \quad (\text{A11})$$

The net result of these linear transformations is to convert a  $K$ -class multinomial to a univariate (essentially) normal distribution. The distance measures used are standard least-squares procedures.

### ACKNOWLEDGMENTS

We gratefully acknowledge the assistance we received from many people during the conduct of this work. In particular, Martha E. Richardson carried out the electrophoretic analyses and assisted in some of the collections. Drs. Lynn H. Throckmorton, Marvin Wasserman, and M. R. Wheeler assisted in numerous ways from the inception of the first electrophoretic studies in the *repleta* group more than 10 years ago. Assistance in acquiring collections was freely given by Drs. E. Casas-Díaz, V. M. Salceda, and E. Hernández Xolocotzi (Chapingo, Mexico), by Dr. J. S. Johnston (Baylor University, Waco, Texas), and by Mr. Ike Russell and W. R. Johnson, Jr. (Tucson, Arizona). We appreciate the helpful comments from several colleagues during the preparation of the manuscript.

## REFERENCES

- Ainsley, R., and Kitto, G. B. (1975). Selection mechanisms maintaining alcohol dehydrogenase polymorphisms in *Drosophila melanogaster*. In Markert, C. L. (ed.), *Isozymes*, Vol. 2: *Physiological Function*, Academic Press, New York, pp. 733-742.
- Ayala, F. J., Tracey, M. L., Hedgecock, D., and Richmond, R. C. (1974). Genetic differentiation during the speciation process in *Drosophila*. *Evolution* **28**:576.
- Bulmer, M. G. (1971). Protein polymorphism. *Nature* **234**:410.
- Charlesworth, B., and Charlesworth, D. (1973). A study of linkage disequilibrium in populations of *Drosophila melanogaster*. *Genetics* **73**:351.
- Clarke, B. (1970). Selective constraints on amino acid substitutions during the evolution of proteins. *Nature* **228**:159.
- Collins, J. F., Duke, E. J., and Glassman, E. (1971). Multiple molecular forms of xanthine dehydrogenase and related enzymes. IV. The relationship of aldehyde oxidase to xanthine dehydrogenase. *Biochem. Genet.* **5**:1.
- Day, T. H., Hillier, P. C., and Clarke, B. (1974). Properties of genetically polymorphic isozymes of alcohol dehydrogenase in *Drosophila melanogaster*. *Biochem. Genet.* **11**:141.
- Edwards, A. W. F., and Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics* **21**:362.
- Fellows, D. P., and Heed, W. B. (1972). Factors affecting host plant selection in desert-adapted cactiphilic *Drosophila*. *Ecology* **53**:850.
- Hubby, J. L., and Throckmorton, L. H. (1968). Protein differences in *Drosophila*. IV. A study of sibling species. *Am. Naturalist* **102**:193.
- Johnson, F. M., Kanapi, C. G., Richardson, R. H., Wheeler, M. R., and Stone, W. S. (1966). An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. *Proc. Natl. Acad. Sci.* **56**:119.
- Johnson, F. M., Richardson, R. H., and Kambyzellis, M. P. (1968). Isozyme variability in species of the genus *Drosophila*. III. Qualitative comparison of the esterases of *D. aldrichi* and *D. mulleri*. *Biochem. Genet.* **1**:239.
- Johnson, G. B. (1974). On the estimation of the effective number of alleles from electrophoretic data. *Genetics* **78**:771.
- Johnson, G. B. (1975a). Enzyme polymorphism and adaptation. *Stadler Symp.* **7**: in press.
- Johnson, G. B. (1975b). Use of internal standards in electrophoretic surveys of enzyme polymorphism. *Biochem. Genet.* **13**:833.
- Kambyzellis, M. P., Johnson, F. M., and Richardson, R. H. (1968). Isozyme variability in species of the genus *Drosophila*. IV. Distribution of the esterases in the body tissues of *D. aldrichi* and *D. mulleri* adults. *Biochem. Genet.* **1**:249.
- King, J. L. (1973). The probability of electrophoretic identity of proteins as a function of amino acid divergence. *J. Mol. Evol.* **2**:317.
- King, J. L., and Ohta, T. (1975). Polyallelic mutational equilibria. *Genetics* **79**:681.
- Kojima, K., and Yarbrough, K. M. (1967). Frequency dependent selection at the esterase-6 locus in *D. melanogaster*. *Proc. Natl. Acad. Sci.* **57**:645.
- Kojima, K., Gillespie, J., and Tobari, Y. N. (1970). A profile of *Drosophila* species enzymes assayed by electrophoresis. I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other systems. *Biochem. Genet.* **4**:627.
- Langley, C. H., Tobari, Y. N., and Kojima, K. (1974). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**:921.
- Lengyel, J., and Penman, S. (1975). hnRNA size and processing as related to different DNA content in two dipterans: *Drosophila* and *Aedes*. *Cell* **5**:281.
- Mahler, H. R., and Cordes, E. H. (1966). *Biological Chemistry*, 2nd ed., Harper and Row, New York, 1009 pp.
- Makela, M. E. (1975). Theoretical and statistical analysis of mixtures of genetically differentiated gene pools. Ph.D. dissertation, University of Texas at Austin.

- Miller, S., Pearcy, R. W., and Berger, E. (1975). Polymorphism at the  $\alpha$ -glycerophosphate dehydrogenase locus in *Drosophila melanogaster*. I. Properties of adult allozymes. *Biochem. Genet.* **13**:175.
- Mukai, T., Mettler, L. E., and Chigusa, S. (1971). Linkage disequilibrium in a local population of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **68**:1065.
- Nei, M., and Chakraborty, R. (1973). Genetic distance and electrophoretic identity of proteins between taxa. *J. Mol. Evol.* **2**:323.
- Patterson, J. T., and Mainland, G. B. (1944). The *Drosophilidae* of Mexico. *Stud. Genet. Drosophila II: Univ. Tex. Publ.* **4445**:9.
- Richardson, R. H., and Johnson, F. M. (1967). Isozyme variability in species of the genus *Drosophila*. II. A multiple allelic isozyme system in *Drosophila busckii*: A stable polymorphic system. *Biochem. Genet.* **1**:73.
- Richardson, R. H., Richardson, M. E., and Smouse, P. E. (1975). Evolution of electrophoretic mobility in the *Drosophila mulleri* complex. In Markert, C. L. (ed.), *Isozymes*, Vol. 4: *Genetics and Evolution*, Academic Press, New York, pp. 533-545.
- Scandalios, J. G., Liu, E. H., and Campeau, M. A. (1972). The effects of intragenic and intergenic complementation on catalase structure and function in maize: A molecular approach to heterosis. *Arch. Biochem. Biophys.* **153**:695.
- Singh, R. S., Hubby, J. L., and Lewontin, R. C. (1974). Molecular heterosis for heat-sensitive enzyme alleles. *Proc. Natl. Acad. Sci.* **71**:1808.
- Singh, R. S., Hubby, J. L., and Throckmorton, L. H. (1975). The study of genic variation by electrophoretic and heat denaturation techniques at the octanol dehydrogenase locus in members of the *Drosophila virilis* group. *Genetics* **80**:637.
- Stone, W. S. (1955). Genetic and chromosomal variability in *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* **20**:256.
- Sturtevant, A. H. (1942). The classification of the genus *Drosophila*, with descriptions of nine new species. *Stud. Genet. Drosophila II: Univ. Tex. Publ.* **4213**:7.
- Thompson, E. A. (1973). The method of minimum evolution. *Ann. Hum. Genet.* **36**:333.
- Wasserman, M. (1962). Cytological studies of the repleta group of the genus *Drosophila*. V. The *mulleri* subgroup. *Stud. Genet. II: Univ. Tex. Publ.* **6205**:85.