STADLER SYMP. Vol. 7 (1975) University of Missouri, Columbia - 91

ENZYME POLYMORPHISM AND ADAPTATION

(enzyme polymorphism, electrophoresis, neutral hypothesis)

GEORGE B. JOHNSON

Department of Biology
Washington University
St. Louis, Missouri 63130

SUMMARY

After a brief resume of the controversy concerning the adaptive value of enzyme polymorphisms, a physiological hypothesis is advanced that heterosis for enzymes of intermediary metabolism results from the differential kinetic behavior of the alleles, which in heterozygotes serve to buffer rate-determining reactions from environmental perturbations. Polymorphism within a population of alpine butterflies is examined in some detail, and the results strongly implicate ongoing selective processes. A more detailed understanding would seem to require more pointed in vivo physiological analyses of the polymorphic variants. A characterization of the physical nature of the electrophoretic variants suggests that many variants do not involve a charge difference, while almost all involve a significant conformational difference. The value of explicit error estimates associated with each data characterization is stressed throughout.

INTRODUCTION

With the extensive use in the last ten years of zone electrophoresis as a genetic survey technique, it has become clear that levels of genetic variability are quite high in natural populations. The average levels of heterozygosity exceed 10%, 30% of examined loci are polymorphic. This degree of variability is much higher than had been expected, and its interpretation has become one of the central questions of population genetics.

Kimura and others have advanced the idea that this level of genetic variation reflects random events, the polymorphic enzyme alleles detected by electrophoresis actually having

little or no differential effect on fitness. Polymorphic enzyme alleles are seen as functionally equivalent, the differences between them thus being "neutral" to the action of selection. The idea is that the electrophoretic differences detected as altered band mobilities represent minor changes in the protein, sufficient to alter its net charge, but not to affect its activity. It is this concept of enzyme polymorphism representing randomly originating non-selective differences that forms the basis of the hypothesis of neutrality concerning selection.

Other workers have taken the opposite position that the high levels of genic variability are actively maintained by selection. This view implicitly assumes that there are functional differences between the alleles which have a differential effect upon fitness.

INITIAL APPROACHES

There has been a great deal said on this issue in the last few years, both pro and con. The initial empirical confrontation concerned biogeographic patterns which could be seen in enzyme allele frequencies. Populations sampled from very different locations in the distribution of a species show similar arrays of allele frequencies (PRAKASH, LEWONTIN and HUBBY 1969, BERGER 1971, BURNS and JOHNSON 1971, LAKOVAARA and SAURA 1971, ROCKWOOD-SLUSS et al. 1973, and others), which suggests parallel selection. A similar line of argument concerned patterns of environmental correlation. An example is allozyme variation in freshwater fish, where the frequencies of some polymorphic enzyme alleles appear to mirror water temperature (KOEHN 1969). Many other such correlations have been cited.

The initial responses of "neutralists" to these arguments have been very much to the point. Maruyama and others point out that very low levels of migration can produce similarity in gene frequency between populations, even for neutral alleles. Current investigations on Drosophila dispersal suggests that the point is a good one. As for environmental correlation, there is a simple and very powerful counter-argument proposed for this class of data. If a polymorphic locus is linked to another locus which is undergoing selection, then "neutral" variability can be maintained by selection on this other locus. Many loci may in this manner exhibit variability due to linkage to a single selected locus. In addition, if selection upon this other locus correlates with the environment, then the polymorphic ones will appear to correlate as well. Such "associative overdominance" is not unlike invoking divine intervention, as it is an argument that can explain practically any pattern of field data.

The "second round" of empirical arguments have gone something like this: migration may be eliminated as a cause of similarity between populations if the populations are closely-related but separate species. In a detailed study of the *Drosophila willistoni* group, similar patterns of allele frequencies are seen in different species (AYALA 1972), even though

there is no gene exchange between them. Environmental correlation arguments have also been defended: biochemical analyses of the variant enzymes (the cold-K_m allele in colder water, etc.) indicate correlation with function (MERRITT 1972). Neutralists have replied that the various *D. willistoni* species may not be in genetic equilibrium (e.g., genic similarity simply reflects recent divergence), and that the correlations of kinetics with function are based upon little data and may be spurious, or reflect linkage.

Thus, field investigation has not succeeded in resolving the issue, although attempts continue. Current work centers on the detailed analysis of multiple locus interactions. While differential migration or other demographic or historical phenomena may be invoked to account for clines or other patterns in allele frequency at any one locus, other loci in the same individuals would then be expected to show similar patterns. Preliminary studies indicate different patterns in allele frequency at different loci within single populations (CHRISTENSEN and FRYDENBERG 1974), rather than the locus-to-locus uniformity predicted under the neutral hypothesis. It is to this point that I have addressed the field studies described below.

PHYSIOLOGICAL AND BIOCHEMICAL APPROACHES

The essential disagreement between selectionists and neutralists is in their view of the functional significance of allele differences at enzyme loci. A selectionists makes two implicit assumptions: one, he assumes that polymorphic alleles at an enzyme locus function differently; secondly, he assumes further that this difference is physiologically important enough to affect fitness. The neutralist hypothesis argues that there is no difference in enzyme function, or that it makes no difference to fitness. Relatively few attempts have been made to distinguish between the two hypotheses at this level: the first point requires a rather sophisticated biochemical approach, and the second point seems on the face of it to require rather difficult physiological analysis.

Generalized comparisons of levels of polymorphism in functionally different classes of enzymes have tended to support the selectionists view. Loci whose enzymes utilize substrates originating from the external environment are far more polymorphic than loci whose enzymes utilize internal metabolites (GILLESPIE and KOJIMA 1968, KOJIMA, GILLESPIE and TOBARI 1970, JOHNSON 1973); (Table 1). These observations suggest that genetic polymorphism reflects a physiological response to environmental variation.

The finding of a consistent correlation between genetic variability and variability in the in vivo "task" of an enzyme suggests a more pointedly physiological approach. From a physiological viewpoint, selection will act upon each of the various steps of a biochemical process as a function of how each affects the output of that process, the metabolic "phenotype". Those reactions which critically affect flux through pathways

should be the most sensitive targets of selection. If enzyme polymorphisms reflect selection, then they ought to occur predominantly at rate-limiting steps (JOHNSON 1971b). Fortunately, a great deal of hard in vivo physiological data is available to test this hypothesis, as population geneticists have tended to score for loci whose enzymes are well studied physiologically.

TABLE 1. Allelic diversity in Drosophila as function of source of substrate. From JOHNSON (1973).

Drosophila Species	Enzymes with source of substrate internal		Enzymes with source of substrate external	
	S	$\bar{k}$	S	$\bar{k}$
D. subobscura	3	2.67	7	3.14
D. equinoxialis	2	2.00	4	2.50
D. pseudoobscura	3	2.67	5	3.60
D. obscura	3	2.00	6	3.60
D. willistoni	11	2.84	7	3.55
D. melanogaster	10	1.48	5	2.86
D. paulistorum	3	1.31	7	2.86
D. bipectinata	11	2.27	6	3.67
D. parabipectinata	11	1.54	6	2.50
D. malerkotliana pallens	11	1.82	6	3.67
D. athabasca	9	1.33	5	3.40
D. simulans	10	1.40	5	3.60
D. affinis	7	1.43	5	4.00
D. tropicalis	2	1.00	4	3.25

S: number of loci sampled with n>100 genomes. $\bar{k}$: mean number of alleles observed at a frequency $\geq$ 0.01 in samples with n> 100 genomes.

Extensive comparisons in Drosophila, small vertebrates, and man (JOHNSON 1974a) strongly confirm this prediction (Table 2). Rate-limiting "regulatory" reactions consistently exhibit far more polymorphism than those which are non-regulatory.

These results are important not only because they argue against a random basis for the bulk of the genetic variability, but also because they suggest a mechanism which might produce a generalized heterosis and thus account for the high levels of variability observed. Considered from a functional viewpoint, metabolism in a variable environment presents a particularly difficult evolutionary problem: How to coordinate and integrate separate metabolic processes, each of which might respond quite differently to an environmental change. Thus, the rate-limiting enzymes of glycolysis and of pentose metabolism may

respond quite differently to a 5° change in temperature, while the need for coordination of these two processes remains. In considering metabolism from an evolutionary viewpoint, the problem of maintaining an integrated metabolism in a changeable environment is of central importance.

TABLE 2. Metabolic patterns of polymorphism. Data are expressed as average heterozygosities. From JOHNSON (1974a).

Class of reaction	Drosophila	Small vertebrate	Man
Variable substrates	0.24	0.22	0.18
Specific substrates			
Regulatory	0.19	0.14	0.13
Nonregulatory	0.06	0.06	0.005
All loci	0.16	0.12	0.07

One cannot control a complex process without controlling the rate-determining elements. Whether it is building a house or running a metabolic pathway, the rate controlling steps in the process must be buffered against random stochastic changes, or the process output cannot be regulated. Enzymes at rate-limiting steps in metabolism usually operate at the very limit of their capability (this is why they are the slowest step and thus limit the overall rate); small changes in reaction conditions can affect their rates significantly. They thus present a critical problem in the physiological coordination of metabolism: The very properties which make them points of metabolic control also render them particularly sensitive to disorganizing influences from the environment.

However, a regulatory locus in a heterozygous individual presents quite a different situation. Unlike his homozygous counterpart, such an individual contains TWO allelic forms of the rate-limiting enzyme. Available data indicate that these two allozymes usually differ in their kinetic response to changes in reaction condition. An example is given in Fig. 1 for alcohol dehydrogenase in Drosophila. This displacement of reaction kinetics along the "environmental" axis has a very important consequence for loci of intermediary metabolism such as are usually scored in population genetics: In such reactions steady state substrate levels are usually quite low (when measured in vivo they may be of the order of the K_m or less). Thus under typical physiological conditions, with the two allozymes present in a heterozygote differing significantly in their affinity for substrate, only the more strongly binding

(low K_m) form will actually "see" any substrate. Although two allozymes are present, the system acts as if only one form were, and the realized kinetics are those of that form. Thus in Fig. 1 at 20°C, the αβ form determines the effective K_m. Now if αβ were the only form present, a 5° rise in temperature to 25°C (well within the actual thermal range of Drosophila habitat) would significantly raise the K_m, and the reaction rate under low substrate conditions would change drastically.

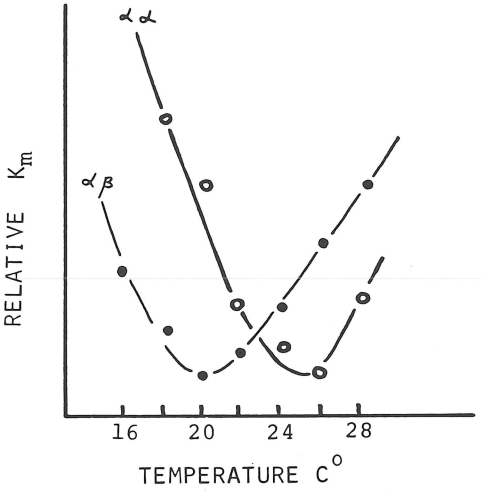


Figure 1. The effect of temperature upon substrate affinity of alcohol dehydrogenase for n-proponal for a homodimer and a heterodimer.

This is just what you don't want to happen at regulatory reactions. But in the heterozygote the αα form has a lower K_m at 25°C than the αβ form, and thus it now determines the realized reaction kinetics. The result is that the effective reaction kinetics in the heterozygote change far less due to temperature perturbations than in homozygotes. The kinetically different multiple forms serve to "buffer" the reaction against environmental variation.

The finding of high levels of heterozygosity at rate-limiting reactions (and very low levels at non-rate-limiting ones) now takes on added significance. In a typical physiological situation where substrate concentrations for metabolite-processing enzymes are low, a homozygous individual can maintain a constant flux at a rate-limiting reaction only over a narrow range of temperature; in a heterozygous individual, however, reaction flux may be maintained constant over a far broader range of environment. This both suggests a generalized

basis for heterosis at the highly polymorphic rate-limiting reactions and explains why non-rate-limiting reactions are generally not polymorphic.

The hypothesis that heterosis for enzymes of intermediary metabolism resides in kinetic differentiation of the alleles makes no assumptions about the nature of heterologous molecules (heterodimers are reported to form within the heterozygotes at most multi-subunit enzymes, giving a third band). "Hybrid multimers" need not be superior in any way to maintain the heterosis and, indeed, may be inferior. Results recently reported on the kinetic nature of heterodimers at one esterase locus in Drosophila (BERGER 1974) suggest that in Drosophila esterase heterodimers are indeed often inferior in stability, kinetics, etc. My own analysis, discussed below, indicates that heterodimers at a variety of loci in Colias butterflies exhibit marked conformational changes, very much in line with Berger's results for the esterase 5 locus. Thus, the hypothesis outlined above suggests that we must look to the properties of the alleles themselves and not to the heterologous multimers, if we wish to understand the physiological implications of polymorphism.

While relatively little attention has been given to direct kinetic analysis of allelic enzymes, work to date supports these arguments. Published investigations of isozyme and allozyme polymorphisms over the last ten years have uniformly reported significant kinetic differences (LICHT 1964, SOMERO and HOCHACHKA 1968, SOMERO 1969, KOEHN 1969, HOCHACHKA and LEWIS 1970, KOEHN et al. 1971, MOON and HOCHACHKA 1972, SCANDALIOS et al. 1972, MERRITT 1972, VIGUE and JOHNSON 1973, BALDWIN and ALEKSINK 1973, FELDER et al. 1973, BEWLEY 1973, CLARKE et al. 1973, MCNAUGHTON 1974, LOW and SOMERO 1974, BATTISTUZZI et al. 1974). Differences between variants are seen in binding affinity of enzyme for substrate or cofactor, in the thermal sensitivity of such binding, in the maximal reaction rate, in stability to high temperature, etc. There seems little question in the cases which have been investigated that electrophoretically detected enzyme variants differ kinetically. Nor is this surprising in view of the significant conformational differences which my own work has suggested exists between most alleles.

Thus, available data suggest that polymorphic enzyme alleles possess the potential for differential effects upon fitness. Whether or not this potential is realized is the key empirical point which must now be addressed.

AN APPROACH TO THE STUDY OF NATURAL POPULATIONS

In an attempt to see whether allelic differences at polymorphic loci can be explained in terms of variation in specific environmental factors, I have for the last several years studied natural populations of butterflies of the genus Colias (known commonly as the Sulfur Butterfly), often working in close conjunction with Ward Watt of Stanford University. These butter-

flies are ideally suited to such studies: they are far larger than Drosophila, and thus offer more material per individual for biochemical analysis; they may be raised (albeit with some difficulty) in the laboratory, and variants subjected to formal Mendelian analysis; populations lend themselves readily to mark-release-recapture studies, providing necessary demographic information on population size, individual movements, etc.; most importantly, populations are known which reside in quite different habitats, making possible directed questions about differential adaptation.

My preliminary work indicated that at the α-glycerophosphate dehydrogenase locus there are two major electrophoretic alleles, apparently the same alleles in each of four species under study. One allele is common in the lowland species *Colias philodice*, and the other common in the alpine species *Colias meadii*, while the two montane species *C. alexandra* and *C. scudderi* exhibit both alleles in high frequency. When alpine or lowland species live in montane habitats, they also are heterozygous for both alleles. Whenever any of the four species live in a montane habitat, it is observed to be heterozygous for α-GPdH. This suggests a strong influence of habitat upon allele frequency. To investigate this apparent gene-environment relationship more pointedly, I examined populations of *C. meadii* at several locations where the populations seem to stradle timberline, occupying both alpine and montane habitats. In all of these populations, members thus live along a marked ecological cline. One population, at Mesa Seco, is particularly interesting as it is the subject of extensive demographic studies by Watt and coworkers. In this population, I examined individuals all of which had been mark-released-recaptured twice at the same site along a transect through the population. In each of several years, I observed strong clines in α-GPdH allele frequency that parallel the habitat cline (Fig. 2). Similar clines were also seen in each of two other alpine-montane populations. The pattern at all three field sites is that observed before, with alpine samples predominantly one allele, and montane samples heterozygous for two alleles.

In these early studies of clines in α-GPdH frequency in *C. meadii*, sample sizes were small, as each individual had been marked, released, and recaptured twice in the course of Dr. Watt's demographic studies (WATT, several manuscripts in preparation). Larger samples were collected in a subsequent study and each individual analyzed electrophoretically for 10 enzymes (Table 3). This study both confirmed the α-GPdH clines seen in the earlier sample of Fig. 2. and revealed a variety of alternative microgeographical patterns in allele frequency at other loci within the population. A clear and important pattern emerges from this within-population analysis: different loci are "viewing" the same environmental heterogeneity differently! The "neutral hypothesis" would certainly not predict this pattern of data. Any demographic or historical explanation advanced to explain the marked cline in α-GPdH would predict the same pattern at other loci. Clearly strong differential selection is operating within this population.

Because the selective factors are not know, it is impossible to state from the data that these loci are themselves the targets of selection, rather than others linked to them. This serves to emphasize the futility of single locus genetic studies of natural populations, a point cogently argued recently by R. C. LEWONTIN (1974). The diverse micro-geographic patterns seen at different loci within the same genetic population of butterflies suggest that adaptive complexes of particular alleles may be organized into metabolically coherent functional units, the

Altitude of Site	# Individuals Analyzed	Allele Frequency Fast	Allele Frequency Slow	Observed Frequency of α-GPdH Heterozygotes
12,400'	21	0.91	0.09	0.19
12,200'	9	0.94	0.06	0.11
11,600'	24	0.69	0.31	0.21
10,800'	21	0.40	0.60	0.43
10,700'	5	0.40	0.60	0.80

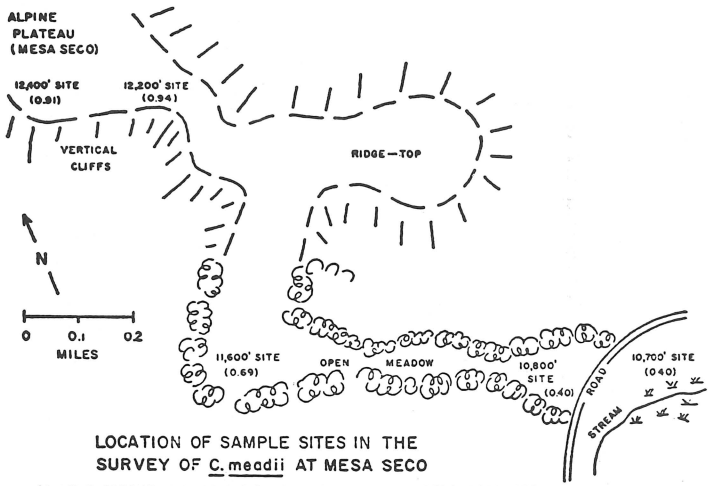


FIGURE 2. Clinal variation at the α-GPdH locus in the Mesa Seco population of *C. meadii*, after JOHNSON (1975b).

optimal assemblage differing at different points along the transect. This hypothesis may be directly studied by in vitro comparisons of the kinetic nature of the allele assemblages.

The complex adaptive patterns seen in the data of Table 3 suggest that future studies of the nature of adaptation at the molecular level will be most successful when the selective factor is known a priori rather than surmised a posteriori,

TABLE 3. Enzyme polymorphism along a transect within a single population of *C. meadii*

Site #	Character	Average Heterozygosity									
		α-GPdH	G6PdH	MdH-I	MdH-II	ME	Fum	AK	HK	PGM	TPI	$\overline{HET}$
1	Alpine	.11	.80	0	.15	.05	0	.30	.10	.60	.05	.17
2	Timberline/ Montane Meadow	.21	.85	0	.20	0	.35	.26	.35	.60	0	.24
3	Montane Forest	.38	.80	.15	.15	.05	0	.25	.15	.60	.05	.22
4	Forest Meadow	.43	.85	0	.45	.15	.55	.15	.10	.65	.05	.28
5	Open Meadow	.30	.60	0	.15	.05	.30	.20	.15	.65	0	.21

n = 40 at each site; samples collected in August 1974 on the Mesa Seco in Hinsdale County, Colorado; every individual butterfly was scored for all enzyme systems; mendelian inheritance has been verified for α-GPdH, MdH, and Fum in my laboratory, and for PGM in the laboratory of W. Watt (personal communication).

and that such future studies need to involve direct in vivo comparisons of alternative allele function under differing degrees of stress. Such a pointedly functional analysis is by no means impractical. Indeed, I am currently initiating such a study in a monocot grass.

THE PHYSICAL NATURE OF OBSERVED ELECTROPHORETIC MOBILITY

Much of the recent dialogue within population genetics concerning the adaptive or neutral nature of enzyme polymorphisms has involved implicit assumptions about the physical nature of the electrophoretically detected variants. Often key theoretical assumptions are naive (JOHNSON 1974b) or based upon no information. Thus it is important to examine in more detail the physical nature of the variants under discussion.

Electrophoretic variants have usually been reported as gel bands of altered mobility. Many experimental factors (temperature, buffer, voltage, etc.) affect electrophoretic mobility of proteins, although it is possible with suitable procedures to adequately standardize against these sources of experimental errors (JOHNSON 1971). Even in a fully standardized system free of systematic experimental errors, however, it is important to realize that a protein migrates in an electrophoresis gel at a rate which reflects not only its charge, but also its size and shape. Not only the electric field but also friction generated by the protein's movement through the gel matrix affects the observed rate of migration. Conformational differences between proteins thus may have major effects upon migration rate. The chemical forces which hold a protein in a particular shape and conformation are weak ones, primarily involved with hydrophobic interactions with water. Many amino acid substitutions not involving a charge change may have a major effect on these weak interactions and produce major conformational changes. It thus becomes important to ask whether such conformational variants would be detected by electrophoretic analysis.

The theory of polyacrylamide gel electrophoresis has been extensively developed within the last few years (CHRAMBACH and RODBARD 1971) and now permits independent determination of the relative contributions of protein charge and of protein size and shape to electrophoretic mobility. Thus a direct empirical assessment of the importance of conformational variability is now possible.

The essence of electrophoretic theory may be stated by a simple equation describing the migration of a protein species in polyacrylamide gel electrophoresis (CHRAMBACH and RODBARD 1971, GONENNE and LEBOWITZ 1975):

$$R_f = \frac{Mo}{u_f} \cdot e^{[K_r \%T]}$$

where R_f = mobility of protein relative to the front
u_f = mobility of moving boundary (a constant known for most common buffer systems)
Mo = free electrophoretic mobility of protein
K_r = the retardation (frictional) coefficient
%T = % acrylamide (which determines pore size)

The important aspect of this equation is that it completely describes electrophoretic behavior in terms of two measurable variables: the one the protein's charge contribution and the other the protein's conformation (frictional) contribution. A plot of ln R_f vs %T is a linear function with a slope of K_r and an intercept whose antilog is Mo/u_f. These two parameters K_r and Mo provide a complete description of the contributions of the macromolecule to its migration rate: Mo is the free electrophoretic mobility that the protein would have in an ideal system with infinitely large gel pore size and thus no gel friction to restrain migration; it thus reflects net charge independent of any size or shape effects. K_r, the retardation coefficient, is a frictional coefficient which measures the degree to which a molecule is retarded by the gel matrix; it reflects molecular weight and shape independent of any charge effects.

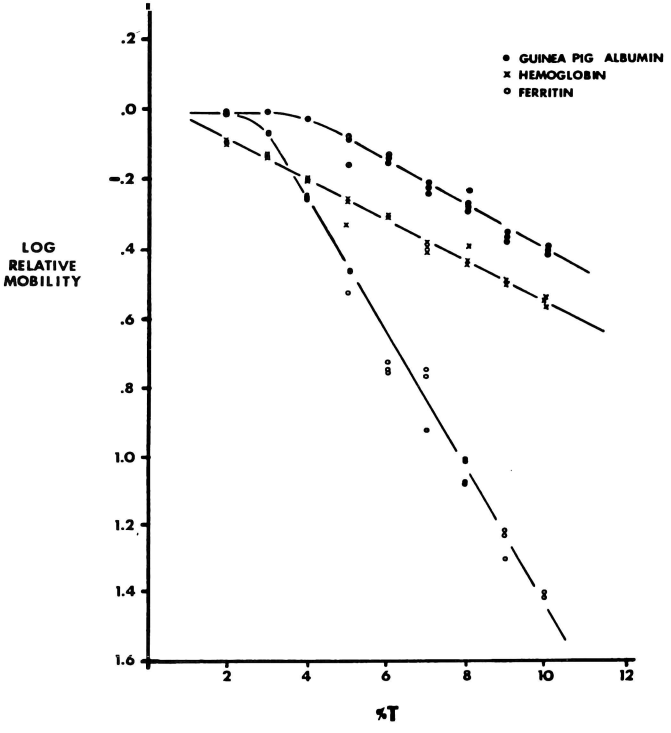


FIGURE 3. A Typical Plot of Mobility as a Function of Acrylamide Pore Size.

In this analytic approach one experimentally measures electrophoretic mobility at a variety of acrylamide concentrations. This effectively varies the gel pore size. Extrapolation to zero % acrylamide estimates migration rate when the pore size is infinitely large and permits direct estimation of Mo. The slope K_r measures the degree to which increasing %T (decreasing mean gel pore size) slows the migration rate, and is thus a measure of molecular size and conformation: a larger molecule would show a greater reduction in migration rate for a given increase in %T, and hence a greater K . K is a very sensitive function

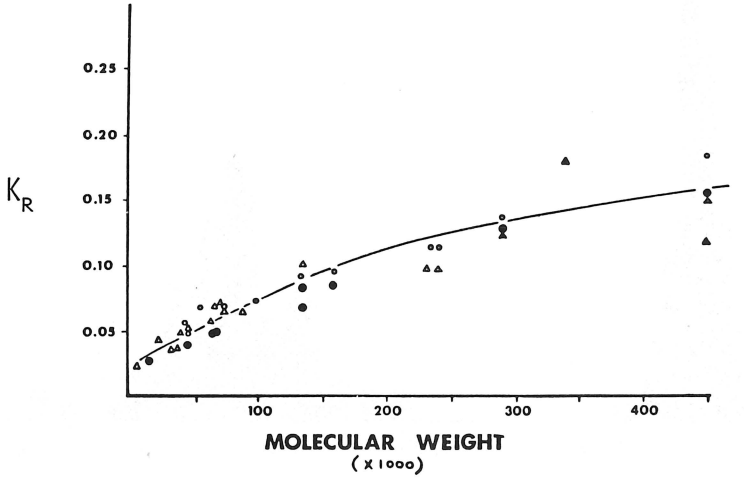


FIGURE 4. The effect of molecular size upon K ; after JOHNSON (1975).

of effective molecular radius in the Chrambach-Rodbard formulation and thus offers a promising probe of conformational difference.

Electrophoretic surveys of naturally-occurring proteins are usually carried out in 10% starch gel, a pore size equivalent to about 4-5% acrylamide; such work thus surveys variability in terms of R_f at about 4%T. It provides a *rate* measurement and is thus subject to many experimental error factors, even when internally standardized as described in my previous work (JOHNSON 1971).

The analytic approach described above provides far more rigorous allele characterization. Somewhat more work is required, as it involves analyzing each individual on a variety of gels of various %T, in order to estimate K and Mo. I feel this approach offers a major improvement because: 1. It separately estimates size and charge contributions to electrophoretic mobility; 2. It provides information on state properties

of proteins, equilibrium values which could be obtained by any workers; 3. Every determination is in effect a statistical regression line, so that 95% confidence intervals may be placed upon each value of Mo and K . This last point is a very important one, for it means that work from different laboratories may be compared objectively, with relative experimental errors known, and even offers the potential for computer-banking data; it means that different alleles can be objectively defined in terms of their physical properties ± known error, and that two

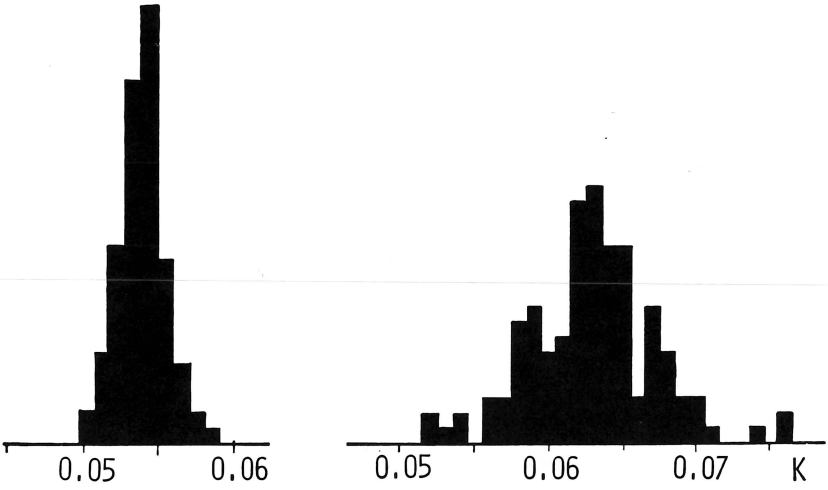


FIGURE 5. Retardation coefficients of α-GPdH in a natural population compared to hemoglobin standards run in the same gels.

variant proteins can be objectively classed as different on the basis of straightforward statistical tests.

PHYSICAL CHARACTERIZATION OF α-GPdH ALLELES

I have estimated K and Mo values, ± 95% confidence intervals, for the enzyme α-glycerophosphate dehydrogenase (α-GPdH) in each of four Colorado Colias species (*C. philodice, C. alexandra, C. scudderi,* and *C. meadii*). The result was startling: unlike preliminary surveys of serum albumins (JOHNSON 1975), I obtained a surprisingly wide variety of values for the K of α-GPdH, even within single species (from 0.05 to 0.08) (Fig. 5)! Examination of the molecular weight calibration curve (Fig. 4) reveals that this change in K is equivalent to that which would be produced by a large molecular weight change and suggests very major conformational differences or post-translational modifications.

Tentative allele assignments were made on the basis of multiple banded individuals: The distribution of Mo values relative to the hemoglobin standard was determined, and the two distal bands of any presumptive heterozygote were always assigned to different allelic classes (all such assignments were only preliminary and are being subject to genetic confirmation). Evaluation of all heterozygotes obtained (Fig. 6) suggested four Mo classes (although two of them are relatively less common). When the K values for each of the four different Mo classes (Fig. 7) are plotted, a clear pattern emerges.

Two classes (I and IV) are relatively less frequent and have distinctive K values. One of the two commoner Mo classes (III) appears to have two K variants associated with it! These

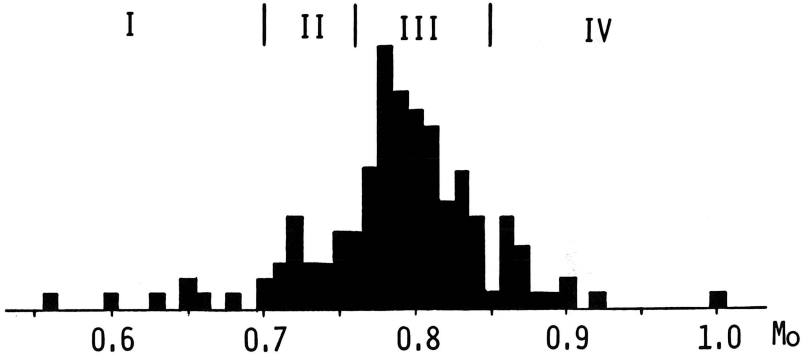


FIGURE 6. Free electrophoretic mobilities of colias, expressed relative to hemoglobin standard run in the same gel.

are statistically different by t-test ($P<0.01$), and, more importantly, several heterozygous individuals were detected which had identical Mo values for all bands, while having bands in each of these two K classes. Thus the evidence indicates an additional variant, identical in free electrophoretic mobility to one of the others, but differing in K ! This presumably represents a variant with a different conformation. Other possibilities, such as the post-translational modification system described by SCHLESINGER et al. (1974) are unlikely, as these two K variants mendelize in laboratory crosses. I would emphasize the dependability of the data used in this analysis: in almost every K regression, experimental standard error was less than 5%, and 95% confidence intervals for K and Mo values were typically less than 10% of experimental values!

One may thus list for α-GPdH five alleles (±σ)

Phenotype Class	Mo (Free Electrophoretic Mobility)	K (Conformational Retardation Coefficient)
1.	0.63±.04	0.70±.08
2.	0.73±.02	0.80±.03
3.	0.80±.02	0.82±.02
4.	0.80±.02	0.89±.03
5.	0.87±.02	0.95±.05

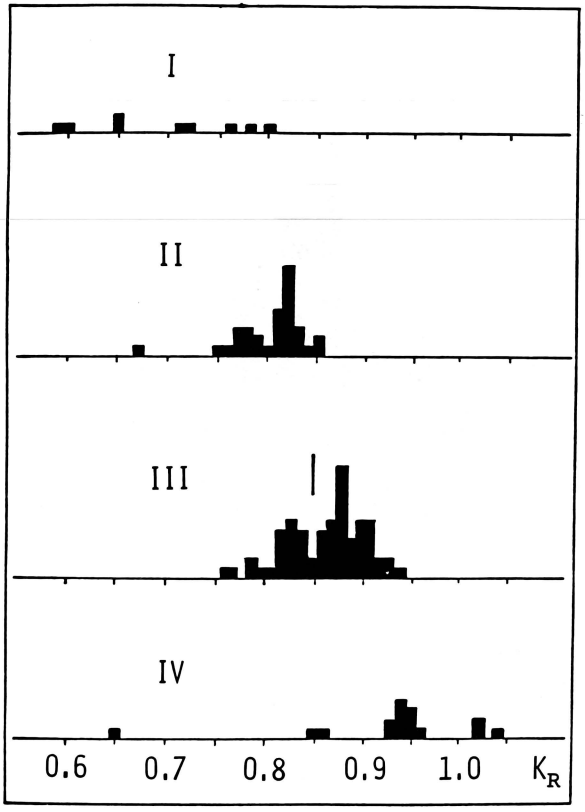


FIGURE 7. Retardation coefficients of each of the Mo classes. Values are expressed relative to hemoglobin standards run in the same gel; after JOHNSON (1975c).

HIDDEN ALLELES AND THE NATURE OF "MOBILITY VARIATION"

Variation in protein conformation, and its effects upon protein migration in gels, suggests an explanation for previously reported variability in allele gel mobility (JOHNSON 1971). Working with 7% acrylamide gels, I had consistently detected variability in electrophoretic band mobility which I could not explain on the basis of experimental error (Fig. 8).

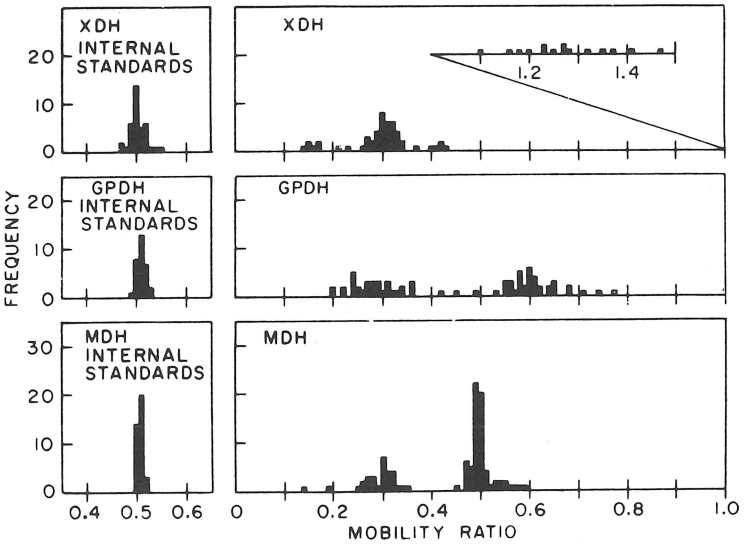


FIGURE 8. Electrophoretiv variability of alleles in excess of that predicted by the ratio of two internal standards run in the same gels; after JOHNSON (1971a).

In order to render my electrophoretic analysis more reproducible, I had been running two internal standard compounds in each polyacrylamide disc gel; the standards are chosen such that any change in experimental conditions will be expected to affect the migration of one of them more than the other (JOHNSON 1975). The ratio of the two internal standards thus provides a means of standardizing diverse data. Under controlled conditions I was able to maintain a constant standard ratio for data collected over many months. Experimental material, however, showed far more variability than that of the standards (JOHNSON 1971). Clearly some factors other than experimental error were contributing to the polymorphic heterogeneity.

The nature of the mobility variation I had previously observed at 7%T now becomes apparent. It is also clear why workers employing starch gels had not detected that mobility varia-

tion (Fig. 9), as they were working at a pore size where such conformation differences would have little differential effect on R_f.

THE GENETIC BASIS OF THE VARIATION

I have performed genetic analyses of the α-GPdH variants in the lowland species *Colias philodice*. Employing the methods developed by Watt and made available to me by him, I have ex-

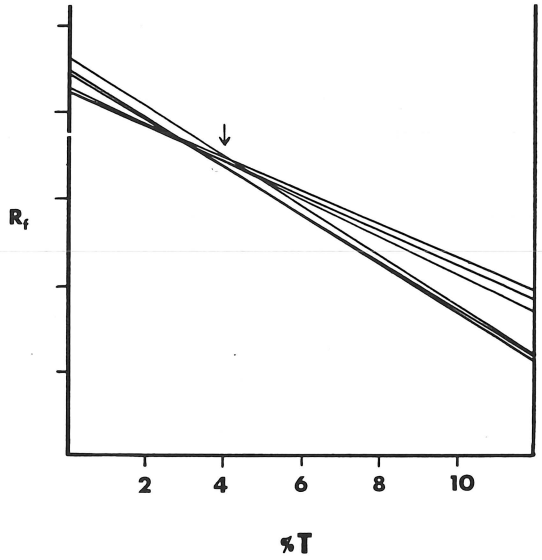


FIGURE 9. Six C. meadii individuals, three of allele class 3 and three of class 4. Starch gels would fail to distinguish these classes, as common starch pore sizes correspond to about 4% acrylamide.

perienced no polyhedral viral disease problems and can typically raise 50 to 100 adults from a single mating. From hatched egg to adult, larval mortality has been consistently about 60%. In all crosses involving a homozygote and a heterozygote, 1:1 segregation was observed. In only one case was I successful in obtaining a heterozygote x heterozygote cross (2/3 x 2/3). The results were 18 (2): 46 (2/3): 12 (3), consistent with a 1:2:1 segregation pattern.

The genetic basis of the K_r difference between phenotypic classes 3 and 4 is another matter. Preliminary crosses of "mobility variation" extremes several years ago gave no heterozygote progeny, suggesting a non-genetic basis. And yet there were

class 3/class 4 heterozygotes in my more recent survey. To settle the matter, I obtained eggs from wild-caught *C. eurytheme* females determined by direct analysis of K to be class 3 and class 4, and with several replicates, individually crossed virgin ♀ 3 x ♂ 4 homozygotes; several reciprocal crosses were also made. All the crosses were fertile, and F_1 progeny were multiple banded at high %T. The results of these crosses thus clearly establish the genetic nature of these two variant classes.

CONFORMATION VS. CHARGE: THE TRUE NATURE OF ELECTROPHORETIC VARIANTS

An analytic approach to allele characterization results in error estimates and it permits a direct assessment of the relative contributions of charge and of frictional retardation to observed electrophoretic mobility. In Table 4 are presented such characterizations for several typical "heterozygote" individuals of *C. meadii*. It is clear that in heterozygotes the additional bands resulting from heterologous molecules can reflect multimers with conformations very different from their non-hybrid parent forms. In this light, BERGER's results (1974) are not at all surprising.

TABLE 4. Typical K_r values for some presumptive heterozygote individuals of *C. meadii*.

Fe/Hb (Ratio of K of internal standards)		$R_f^{5\%}$	K (± 95% confidence intervals)
G6PdH	.167	.41	.072 ± .002
			.085 ± .014
MdH-II	.164	.28	.066 ± .006
		.23	.064 ± .010
		.19	.129 ± .021
Fum	.169	.11	.290 ± .023
		.04	.097 ± .306
HK-II	.166	.85	.071 ± .012
		.80	.048 ± .003
		.75	.078 ± .009
PGM	.162	.63	.059 ± .003
		.52	.058 ± .011
TPi	.164	.63	.068 ± .007
		.61	.068 ± .003
		.58	.066 ± .006

The standard error associated with K estimates of the ratio of the two internal standards (Fe/Hb) is in every case less than 10%.

From the K characterization for several heterozygous loci of *C. meadii* presented in Table 4, a number of interesting conclusions can be drawn:

1. For several enzyme loci, the multiple bands of presumptive heterozygous individuals exhibit identical K values (PGM and TPi), indicating that the different alleles and the heterodimer of TPi have similar conformations.

2. For other loci this is not the case. For HK-II, the presumptive heterodimer exhibits a radically different K value.

3. This same effect can also be seen clearly for MdH-II, where not only is the presumptive heterodimer conformationally dissimilar to its two parent alleles, but its K value is such that at 5% acrylamide (similar pore size to starch gel) the band appearing in the "middle" position is <u>not</u> the presumptive heterodimer.

4. For enzymes such as G6PdH, heterozygotes would not have been detected by examination of mobility at 5% acrylamide.

5. In almost all cases, 95% confidence intervals on the estimates of K are quite narrow, indicating that these estimates accurately reflect the behaviors of the proteins being examined.

6. The high reproducibility of the internal standard ratio and the narrow 95% confidence intervals suggest that these K characterizations are suitable for computer data banking. It is important to note that because statistical evaluation is explicitly associated with each K value, direct comparison may be made with data of other workers using similar approaches.

Table 5 presents data for three typical loci of *C. meadii* in which a population survey was carried out using an analytic approach. All alleles were detectable at R_f = 5%T (approximately equal in pore size to a 10-11% starch gel). In the case of malic enzyme (ME), the two alleles have identical free electrophoretic mobility, differing only in conformation. The same proliferation of conformational variants is seen for HK-I, in which four of the five detected alleles have identical charge.

Table 6 summarizes the results for 15 loci. Clearly a large fraction (almost 40%) of the variants detected by electrophoresis do not differ in charge! Calculations of the total level of genomic variability based upon the fraction of single base substitutions in the genetic code which would produce a change in amino acid charge are in the face of these results clearly useless.

It is very important to note that fully 90% of the detected variants exhibited significant differences in conformation. That enzyme polymorphism is so tightly coupled to conformational change would not have been expected under any "neutral" hypothesis and strongly suggests associated functional differences.

Table 5. Analytic survey in *C. meadii*.

Data From Three Typical Loci

α-GPdH

Allele	n	$\bar{x}$	$\bar{K}_r$ ± σ	$\bar{Mo}$ ± σ
1	12	.30	.877 ± .010	.081 ± .001
2	16	.40	.835 ± .008	.080 ± .001
3	10	.25	.800 ± .016	.076 ± .001
4	2	.05	.782 ± .037	.065 ± .001

ME

Allele	n	$\bar{x}$	$\bar{K}_r$ ± σ	$\bar{Mo}$ ± σ
1	22	.52	.428 ± .010	.054 ± .001
2	20	.48	.386 ± .012	.054 ± .003

Hk-I

Allele	n	$\bar{x}$	$\bar{K}_r$ ± σ	$\bar{Mo}$ ± σ
1	15	.39	1.096 ± .006	.073 ± .002
2	9	.24	1.063 ± .009	.071 ± .002
3	7	.18	1.024 ± .010	.072 ± .003
4	2	.05	.991 ± .013	.073 ± .008
5	4	.11	.794 ± .026	.061 ± .002

Table 6. Summary of analytic survey of 15 loci in *C. meadii*. (n = 40; Alleles listed only if freq. ≥ .005).

	Total # Alleles	# Types Conformationally Distinct	# Types Different In Charge
α-GPdH	4	3	3
G6PdH	3	3	2
ME	2	2	1
FUM	5	4	5
PGM	7	7	2
TPi	4	4	2
Ak-I	2	2	1
Ak-II	3	3	2
Hk-I	5	5	2
Hk-II	3	3	1
MdH-I	5	4	4
MdH-II	4	3	3
EST-I	4	3	2
EST-1I	3	3	2
EST-III	2	2	2
	56	51 (91%)	34 (61%)

My initial studies of α-GPdH had assumed (on the basis of similar R_f values when run in 7% acrylamide) that the same two alleles were segregating in each of the four Colorado species. This same sort of homology criterion is widely employed in current estimations of "genetic distance": such genetic distance measurements estimate the relative probability of sampling identical alleles in the same vs. different populations, and thus critically depend upon assessment of allele homology. It seems clear from the work discussed above, however, that R_f is not a sufficient criterion of homology. Alleles with identical charge may yield different R_f values due to post-transcriptional events which affect K_r. "Electrophoretically silent" alleles may be revealed by K_r differences although having similar R_f values on starch gels, while alleles with different charges may be adjudged similar because of the interactive effects of net charge (reflected in Mo, the free electrophoretic mobility) and conformation (reflected in K_r, the retardation coefficient); this was true for α-GPdH in 10% starch gels, or in polyacrylamide gels at %T = 5 (see Fig. 9). Many of these effects can be seen in Table 7. Thus analyses of genetic distance, particularly betewen species or other taxonomic groups where genetic crosses are not possible, should be evaluated with caution when homology assessments are made solely in terms of R_f.

In Table 7, the K (retardation coefficient) and Mo (free electrophoretic mobility) estimates are presented relative to the Hb standard employed in that determination, ± standard error. Each distinction between alleles indicated by a difference in homology assignment was verified by a "t" test for the significance of difference between the two regression lines [as outlined in SOKAL and ROHLF (1969, p. 450)]; in every case $P>0.90$ that the classes are distinct. All claims of departure from homology can be verified by such a statistical analysis. In many cases apparently modest differences in K_r or Mo are highly significant because the 95% confidence intervals associated with the K_r and Mo estimates are small. It is worth noting that unstandardized R_f-5% values within an allelic class vary by as much as 5%, indicating the importance of standardizing disc gels.

Overall homology assessment based upon analysis of electrophoretic mobility at 5% acrylamide would have indicated that within this genus there are two rare alleles (#1 and #6) and one common allele, this same allele occurring in all species. An analytic characterization, with independent estimation of the charge and conformational factors contributing to observed mobility, reveals a more complex pattern of six distinctly different proteins within the genus. In many cases proteins similar in R_f are not homologous. Many studies of genetic distance are carried out by surveying allozyme variation at 5% alrylamide or 10-11% starch (with a pore size similar to 5% acrylamide). Any estimate of "genetic distance" based upon such mobility data is thus likely to grossly overestimate homology, at least among the species of this genus.

TABLE 7. Characterization of homology among some α-glycerophosphate dehydrogenase alleles of four species of the genus *Colias*.

	(Mobility) $R_f^{5\%}$	(Conformation) K_r^{Hb} ± S.E.	(Charge) M_o^{Hb} ± S.E.	Homology Assignment
C. philodice	0.53	0.65 ± .01	0.73 ± .02	1
	0.61	0.81 ± .02	0.77 ± .01	2
	0.60	0.84 ± .01	0.81 ± .02	3
	0.63	0.90 ± .01	0.81 ± .01	4
C. alexandra	0.62	0.80 ± .02	0.76 ± .01	2
	0.60	0.83 ± .01	0.80 ± .01	3
	0.63	0.90 ± .01	0.82 ± .01	4
	0.62	0.96 ± .02	0.87 ± .01	5
C. scudderi	0.63	0.81 ± .01	0.75 ± .01	2
	0.60	0.83 ± .01	0.81 ± .01	3
	0.63	0.89 ± .01	0.81 ± .01	4
	0.61	0.92 ± .02	0.85 ± .02	5
C. meadii	0.71	0.78 ± .02	0.65 ± .01	6
	0.62	0.80 ± .01	0.76 ± .03	2
	0.61	0.84 ± .01	0.80 ± .01	3
	0.62	0.88 ± .01	0.81 ± .02	4

I have essentially completed an analysis of genetic distance between the four Colorado Colias species *C. meadii*, *C. alexandra*, *C. philodice*, and *C. scudderi* using Mo and K_r values to assess homology. When alleles are characterized in terms of physical properties of proteins such as Mo and K_r, which assess the charge and size contributions to mobility independently, confusions such as indicated in Table 7 do not arise. The ability to assign 95% confidence intervals to all Mo and K_r values, and to assess homology on the basis of statistical criteria, with stated probability values, lends added confidence to the comparisons.

CONCLUSIONS

In this paper, I have attempted to illustrate the power of a physiological perspective in assessing the hypothesis that enzyme polymorphism is of general adaptive value. Such physiological approaches have to date generally supported an adaptive interpretation, and strongly suggest *in vivo* experimentation as a fruitful future line of inquiry. Informed multi-locus approaches involving metabolically related reactions and known physiological stresses seem to me to offer great promise of providing a two-tailed test of the adaptive hypothesis.

I further suggest that empirical approaches to the analysis of electrophoretic gene polymorphisms have failed to take full advantage of the power of the electrophoretic technique. Using internal standardization and analytic procedures, one may obtain independent estimates of the charge and conformational characteristics of an allele, with explicit 95% confidence intervals associated with each data value. These are real physical properties of the proteins and are not rate functions (as R_f values are). Thus a worker coming years later to the same population should arrive at the same characterization of a given allele, within the stated error limits. This analytic approach allows one to ask several old questions with new rigor:

1. Allelic variants may be described in objective terms, with stated error limits;
2. A range of alleles "hidden" in starch gel electrophoresis can be clearly characterized;
3. Heterodimers are seen to be conformationally quite different from non-hybrid enzymes and in no sense intermediate;
4. Much electrophoretic variability is seen not to involve a charge change;
5. Almost all detected electrophoretic variability involves significant conformational change.

LITERATURE CITED

AYALA, F. 1972 Darwinian vs. non-Darwinian evolution in natural populations of Drosophila. Proc. Berkeley Symp. Math. Stat. Probl. (LeCan, L., J. Neyman and E. Scot, Eds.) 6th ed., Vol. 5. University of California Press, Berkeley.

BALDWIN, J. and M. ALEKSINK 1973 Adaptation of enzymes to temperature: Lactate and malate dehydrogenase from *Platypus* and *Echida*. Comp. Biochem. Physiol. 44B:363.

BATTISTUZZI, G., R. SCOZZARI, P. SANTOLAMAZZA, L. TERRENATO and G. MODIANO 1974 Comparative activity of red cell adnosine deaminase allelic forms. Nature 251:711.

BERGER, E. 1971 A temporal survey of allelic variation in natural and laboratory populations of *D. melanogaster*. Genetics 66:121.

BERGER, E. 1974 Esterases of *Drosophila*. II. Biochemical studies of esterase-5 in *D. pseudoobscura*. Genetics 78:1157.

BURNS, J. and F. JOHNSON 1971 Esterase polymorphism in the butterfly *Hemiargus isola:* Stability in a variable environment. Proc. Nat. Acad. Sci. U.S. 68:34.

CHRISTENSEN, B. and FRYDENBERG 1974 Geographical patterns of four polymorphisms in *Zoarces* as evidence of selection. Genetics 77:765.

CHRAMBACH, A. and D. RODBARD 1971 Polyacrylamide gel electrophoresis. Science 172:440.

DAY, T., P. HILLER and B. CLARKE 1974 The relative quantities and catalytic activities of enzymes produced by alleles at the alcohol dehydrogenase locus in *D. melanogaster*. Biochem. Genet. 11:155.

FELDER, M., J. SCANDALIOS and E. LIU 1973 Purification and partial characterization of two genetically defined alcohol dehydrogenase isozymes in Maize. Biochem. Biophys. 317:149.

GILLESPIE, J. and K. KOJIMA 1968 The degree of polymorphisms in enzymes involved in energy production compared to that in nonspecific enzymes in two *D. ananassae* populations. Proc. Nat. Acad. Sci. 61:582.

GONENNE, A. and J. LEBOWITZ 1975 Estimation of molecular weights of small proteins on polyacrylamide gel electrophoresis. Analytical Biochemistry 64:414-424.

HOCHACHKA, P. and J. LEWIS 1970 Enzyme variants in thermal acclimation: Trout liver citrate synthases. J. Biol. Chem. 245:6567-6573.

JOHNSON, G. 1971a Analysis of enzyme variation in natural populations of the butterfly *Colias eurytheme*. Proc. Nat. Acad. Sci. 68:997-1001.

JOHNSON, G. 1971b Metabolic implications of polymorphism as an adaptive strategy. Nature 232:347-349.

JOHNSON, G. 1973 Importance of substrate variability to enzyme polymorphism. Nature New Biology 243:151-153.

JOHNSON, G. 1974a Enzyme polymorphism and metabolism. Science 184:28-37.

JOHNSON, G. 1974b On the estimation of effective number of alleles from electrophoretic data. Genetics 78:771-776.

JOHNSON, G. 1975a The use of internal standards in electrophoretic surveys of enzyme polymorphism. Biochemical Genetics 13, December issue, in press.

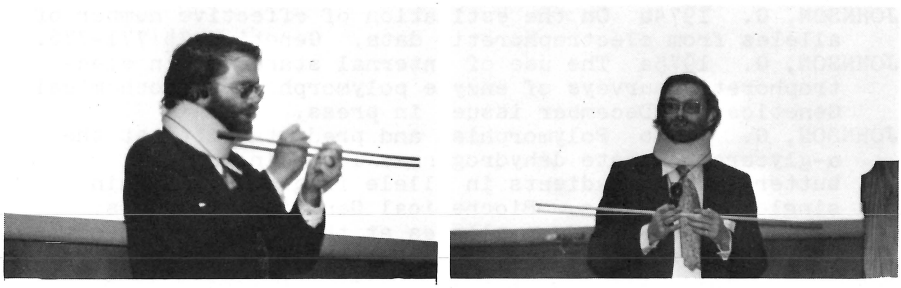
JOHNSON, G. 1975b Polymorphism and predictability at the α-glycerophosphate dehydrogenase locus in *Colias* butterflies: Gradients in allele frequencies within single populations. Biochemical Genetics, in press.

JOHNSON, G. 1975c Hidden alleles at the α-glycerophosphate dehydrogenase locus in *Colias* butterflies. In preparation.

KOEHN, R. 1969 Esterase heterogeneity: Dynamics of a polymorphism. Science 163:943-944.

KOEHN, R., J. PEREZ and R. MERRITT 1971 Esterase enzyme function and genetical structure of populations of the

freshwater fish *N. stramineus*. Am. Natur. 105:51-69.
KOJIMA, K., J. GILLESPIE and Y. TOBARI 1970 A profile of *Drosophila* species' enzyme assayed by electrophoresis. Biochem. Genet. 4:627-637.
LAKOVAARA, S. and A. SAURA 1971 Genic variation in marginal populations of *D. subobscura*. Hereditas 69:77-82.
LEWONTIN, R. C. 1974 The gentic basis of evolutionary change. Columbia University Press, New York.
LICHT, P. 1964 The temperature dependence of myosin-adenosine-triphosphatase and alkaline phosphatase in lizards. Comp. Biochem. Physiol. 12:331-340.
LOW, P. and G. SOMERO 1974 Temperature adaptation of enzymes: A proposed molecular basis for the different catalytic efficiencies of enzymes from ectotherms and endotherms. Comp. Biochem. Physiol. 49B:307-312.
McNAUGHTON, S. 1974 Natural selection at the enzyme level. Amer. Nat. 108:616-624.
MERRITT, R. 1972 Geographic distribution and enzymatic properties of lactate dehydrogenase allozymes in the fathead minnow. Amer. Nat. 106:173-184.
MOON, T. and P. HOCHACHKA 1972 Temperature and the kinetic analysis of trout isocitrate dehydrogenases. Comp. Biochem. Physiol. 42b:725-730.
ROCKWOOD-SLUSS, E., J. JOHNSTON and W. HEED 1973 Allozyme-genotype-environment relationships. I. Variation in natural populations of *D. pachea*. Genetics 73:135-146.
SCANDALIOS, J., E. LIU and M. CAMPEAU 1972 The effects of intragenic and intergenic complementation on catalase structure and function in maize: A molecular approach to heterosis. Arch. Biochem. Biophys. 153:695-705.
SOMERO, G. and P. HOCHACHKA 1968 The effect of temperature on catalytic and regulatory functions on pyruvate kinases of rainbow trout and the antarctic fish *T. bernacchii*. Biochem. J. 110:395-400.
SOMERO, G. 1969 Pyruvate kinase variants of the Alaskan King Crab. Biochem. J. 114:237-241.
VIGUE, C. and F. JOHNSON 1973 Isozyme variability in *Drosophila*. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in *D. melanogaster*. Biochem. Genet. 9:213-237.



George B. Johnson