A MOLECULAR APPROACH TO THE STUDY OF GENIC HETEROZYGOSITY IN NATURAL POPULATIONS. I. THE NUMBER OF ALLELES AT DIFFERENT LOCI IN *DROSOPHILA PSEUDOOBSCURA*[1]

J. L. HUBBY AND R. C. LEWONTIN

Department of Zoology, University of Chicago, Chicago, Illinois

Received March 30, 1966

A cornerstone of the theory of evolution by gradual change is that the rate of evolution is absolutely limited by the amount of genetic variation in the evolving population. FISHER's "Fundamental Theorem of Natural Selection" (1930) is a mathematical statement of this generalization, but even without mathematics it is clear that genetic change caused by natural selection presupposes genetic differences already existing, on which natural selection can operate. In a sense, a description of the genetic variation in a population is the fundamental datum of evolutionary studies; and it is necessary to explain the origin and maintenance of this variation and to predict its evolutionary consequences. It is not surprising, then, that a major effort of genetics in the last 50 years has been to characterize the amounts and kinds of genetic variation existing in natural or laboratory populations of various organisms.

The results so far have told us a great deal about cytological variation such as polymorphisms for inversions and translocations, about frequencies of rare visible mutations at many loci, and about frequencies of chromosomes that are deleterious when homozygous together with the degree of that deleterious effect. In addition, we know of some striking single-locus polymorphisms. These results are familiar to all students of population genetics and evolution, and have been well reviewed by DOBZHANSKY (1951) and more recently by MAYR (1963).

Yet, for all the wealth of observation and experiment, the techniques of population genetics have not allowed us to ask directly the most elementary question about the genetic structure of a population: *At what proportion of his loci can we expect a diploid individual to be heterozygous?* Put in another way, this is the question of how much genetic variation there is in any given population. That this question remains unanswered is best shown by a statement of MAYR (1963) at the end of more than 100 pages of review of our present knowledge.

> "It has been estimated (WALLACE 1958) that 50 percent or more of the loci of a given individual in an open, natural population may be occupied by dissimilar alleles. Others calculate that 12–20 variable loci already place a severe strain on a population, owing to the segregation of inferior genotypes. Far more research and a classification of the basic concepts is needed to narrow the gap between these widely diverging estimates."

[1] The work reported here was supported in part by grants from the National Science Foundation (GB 3213) and the Public Health Service (GM-11216).

They are indeed widely diverging, for 20 loci cannot represent more than 1% of the genome of Drosophila, on which WALLACE's estimate of 50% is based. Although statements giving such specific limits as "12 to 20 loci" are not in the published literature, essentially this idea is embodied in CROW's statement (1961) that it is "highly improbable that more than a minute fraction of the inbreeding effect is due to alleles maintained by balanced selective forces."

The reason for our present lack of knowledge about the amount of heterozygosity per locus in a population is that no technique has been available capable of giving a straightforward and unambiguous answer even under ideal experimental conditions. Any technique that is to give the kind of clear information we need must satisfy all of the following criteria: (1) Phenotypic differences caused by allelic substitution at *single loci* must be detectable in *single individuals*. (2) Allelic substitutions at one locus must be distinguishable from substitutions at other loci. (3) A substantial portion of (ideally, all) allelic substitutions must be distinguishable from each other. (4) Loci studied must be an unbiased sample of the genome with respect to physiological effects and degree of variation.

Requirements 1 and 2 really amount to the condition that the phenotypes used have a simple Mendelian inheritance without important environmental variation. Requirements 3 and 4 come from the need to make statements about variation in the genome as a whole from a necessarily restricted sample.

The methods of analysis previously available have satisfied some of these requirements, but no method has met them all. Studies of the occurrences of visible recessive mutations in populations (SPENCER 1947) meet requirements 1 and 2, but are not random over all loci and certainly come nowhere near meeting requirement 3. Some very small, but unknown, proportion of allelic substitutions show up as visible morphological changes. Particular biochemical and morphological polymorphisms do not meet requirement 4 since they are studied precisely because they *are* polymorphic. Thus, the fact that a number of blood group polymorphisms are known in man is not sufficient information, since we do not know how many loci are *not* polymorphic. Nor do we know whether blood group antigens are an unbiased sampling of the genome.

In an effort to sample gene effects broadly, considerable attention has been paid to genetic variation in fitness traits, like viability, fertility, longevity, and so on. But whether these are examined in Drosophila by special breeding experiments (for example, DOBZHANSKY and SPASSKY 1954), or in man by a study of mortality and morbidity rates in inbred families (for example, MORTON 1960), they fail to satisfy requirements 1 and 2. Because of the nature of the phenotypic differences, it is impossible, except in special cases, to observe the effect of single gene substitutions or to identify the genotype of individuals. In order to detect genetic differences for such characters with high environmental variation, it is necessary to compare the means of groups of individuals so that environmental variation is cancelled out. These groups must be made to differ genetically from each other and this is done either by deliberate inbreeding or by using closely related individuals. The difficulty of this technique is that after inbreeding, the groups or inbred lines differ from each other at many loci simultaneously. In fact,

the number of loci that are represented in different inbred lines by different homozygotes depends upon the amount of variation in the original population, the very thing we are trying to estimate.

Thus the geneticist finds himself in the curious position that allelic substitutions of large enough effect to be detected as single segregating entities are clearly a special class of mutations, while the most general class of variant alleles, so called isoalleles, are of such small effect that they cannot be detected by simple Mendelism. What we require is a technique that will detect a large portion, if not all, of the isoallelic variation at a locus, yet detect it through clear-cut, discrete, nonoverlapping phenotypic classes comparable to the usual visible mutant. It is the purpose of this paper to describe such a method and the genetic results obtained when it was used in *Drosophila pseudoobscura*. The accompanying paper in this journal (LEWONTIN and HUBBY 1966) then shows the result of its application to populations of *D. pseudoobscura* in an attempt to answer the question of how much genic variation exists in a wild population of a sexually reproducing species.

The method to be described is predicated, very roughly, on the following argument. Any mutation in a structural gene should result in the substitution, deletion, or addition of at least one amino acid in the polypeptide produced by the gene. Such a substitution, deletion, or addition will, in some fraction of cases, result in a change in the net electrostatic charge on the polypeptide. This change will in turn change the net charge on the enzyme or other protein of which such a polypeptide is a constituent. Since enzymes and proteins in general are, as far as we know, made up of polypeptides from one or sometimes two different structural genes, then we can expect that electrophoretic differences in enzyme proteins will segregate as single Mendelizing genes. Thus, if we survey a large number of enzymes and other proteins and if we determine the electrophoretic mobility of such proteins from single individuals, it should be possible to detect variability from individual to individual at single loci.

The study of electrophoretic mobility of enzymes and proteins does in fact satisfy the four requirements given above almost perfectly. The phenotypic differences are detectable in single individuals. Allelic substitutions at different loci are distinguishable from each other because the simple genetics of each difference can be investigated as for any phenotypic character. As it turns out, *all the electrophoretic differences to be described turn out to be single Mendelizing entities*. This fact is most important because it frees the method of any *a priori* assumptions about gene action. Moreover, it allows us, as a first order of approximation, to equate a protein without any detectable variation to a gene without detectable variation. That is, we can count up the number of loci in our sample that show no variation, as well as the number that do have alternative alleles. This is the cornerstone of the method for it then allows us to estimate the proportion of all loci that show variation in populations. The third requirement, that a substantial portion of all possible changes is detectable, is only met in part. As it turns out, however, the accompanying paper (LEWONTIN and HUBBY 1966) shows that we must be detecting a substantial fraction of the variation. Finally,

the enzymes and high concentration proteins used in the study have been chosen without reference to their known variability in the population, but only because assay techniques exist for them.

A more complete discussion of the possible biases in our method is given in the accompanying paper where the results of the population analysis are described.

MATERIALS AND METHODS

Forty-three strains of *Drosophila pseudoobscura* have been investigated for the purposes outlined above. Each strain was established from a single, wild caught inseminated female and the 43 strains represent material sampled from five geographic locations. Thirty-three of the strains have been maintained in mass cultures for approximately five years. These include collections from Wildrose, California (10 strains); Mather, California (7 strains); Cimmaron, New Mexico (6 strains); Flagstaff, Arizona (9 strains); and Bogotá, Colombia (1 strain). Ten additional strains from Strawberry Canyon, California were assayed during their F_2 and F_3 generations. All strains were homozygous for the third-chromosome gene arrangement Arrowhead (AR), with the exception of the Strawberry Canyon material, which was segregating for several inversions on the third chromosome. Marker stocks employed in chromosomal location of the various enzyme and protein allelic differences were as follows: Chromosome 2—*up bx Ba gl*/Delta; Chromosome 3—*Bl sc pr*/ *or L*; Chromosome 4—*in hk j Cy* (*In*)/leth.

Electrophoresis in acrylamide gels was performed as described by HUBBY (1963). Polymerization was achieved with addition of tetramethylethylenediamine (TEMED) to 0.2% in place of the dimethylaminopropionitrile used previously. Buffer in the gel and in the cell compartments is the same as described previously except when noted in the assay method.

Enzyme assays: Esterases—(modified from SIMS 1965). Single one-day-old adults were ground in 15 μl of 0.1 M tris(hydroxymethyl)aminomethane-borate (Tris-borate) buffer pH 8.9 containing 1.5 mM ethylenediaminetetracetic acid (EDTA) and 5% sucrose in a 100 μl centrifuge tube. The slurry was centrifuged at 21,500 × *g* for 3 minutes (tubes and centrifuge obtained from Microchemical Specialty Corp., Berkeley, California). 10 μl of the supernatant was layered into the gel pocket and electrophoresis was conducted at 450–500 v and 110–160 ma for 90 minutes. The gels were preincubated at 5°C in 100 ml of a solution of 0.5 M boric acid for 1.5 to 2 hours to lower the pH of the gel to approximately 6.5. The gels were then incubated for 2 to 4 hours at 25°C in 100 ml of 0.1 M phosphate buffer pH 6.5 containing 20 mg α-naphthylacetate and 50 mg Fast Red TRN (both from Dajac Laboratories, Philadephia, Penna.).

Malic dehydrogenase—(modified from SIMS 1965). Gel and sample preparation, and electrophoresis were the same as for esterase determination. For enzyme detection the gels were incubated 1 to 2 hours at 25°C in 100 ml of a solution containing 1 mM Tris-HCl pH 8.5 1 mM Na malate, 0.1 mM KCN, 15 mg NAD, and 50 mg Nitro BT^R. At the end of the incubation period 2 mg phenazine methosulfate (PMS) were added and the gel was incubated an additional 30 minutes.

Glucose-6-phosphate dehydrogenase—(modified from YOUNG, PORTER, and CHILDS 1964). One-day- old adults were extracted as above as the source for this enzyme. 3 mg TPN was included in the gel mixture and electrophoresis was continued for 150 minutes. The gels were incubated 1 to 2 hr in 100 ml of a solution containing 5 mM Tris-HCl pH 8.5, 0.1 mM KCN, 50 μM $MgCl_2$, 50 mg sodium glucose-6-phosphate, 10 mg TPN, and 50 mg Nitro BT^R. At the end of the incubation period 2 mg PMS were added and the gel was incubated an additional 30 minutes.

Alkaline phosphatase—(modified from BECKMAN and JOHNSON 1964a). 0.1 M Tris-borate buffer at pH 8.9 not containing EDTA was used in preparation of the gel and in the cell compartments; in addition, the gel solution contained 5 mM $MgCl_2$. Third instar larvae were ground in 15 μl of buffer also containing $MgCl_2$. Electrophoresis was performed at 400 volts for 90 minutes. For enzyme detection the gels were incubated in 100 ml of a solution containing 1 mM Tris-HCl pH 8.5, 50 μM $MgCl_2$, 50 mg sodium α-naphthylphosphate, 2 g NaCl, and 500 mg

polyvinylpyrolidone. After 1 to 2 hours 50 mg of Fast Blue RR were added and an additional 30-minute incubation period followed.

α-glycerophosphate dehydrogenase—(modified from SIMS 1965). The gel mixture contained, in addition to the usual ingredients, 1.5 mM of α,β-glycerophosphate; the usual extraction buffer also contained an identical concentration of the substrate. One-day-old adults were the source of enzyme, and for enzyme detection the gels were incubated in 100 ml of a solution containing 1 mM Tris buffer at pH 8.5, 1 mM α,β-glycerophosphate, 0.1 mM KCN, 50 mg Nitro BT^R and 15 mg NAD. After 2 hours incubation, 2 mg of PMS were added for additional incubation period of 30 minutes.

Leucine aminopeptidase—(modified from BECKMAN and JOHNSON 1964b). The gel was prepared in the same manner as for alkaline phosphatase. Late pupae (one day before emergence) were ground singly in 10 μl of 0.1 M Tris-borate buffer pH 8.9 containing 5 mM $MgCl_2$ and 5% sucrose. For enzyme detection the gels were preincubated at 5°C in 0.5 M boric acid containing 5 mM $MgCl_2$. After rinsing, the gel was placed in 100 ml of a solution containing 2 mM Tris-malate buffer pH 5.2 and 20 mg L-Leucyl-β-naphthylamide HCl. After 2 hours the gel was transferred to 100 ml of the Tris-malate buffer containing 50 mg of Fast Black K.

Larval proteins: Late third-instar larvae were ground in 20 μl of Tris-borate buffer pH 8.9 containing 1.5 mM EDTA and 5% sucrose. The total supernatant from centrifugation was applied to the gels and electrophoresis was performed at 500 v and 110 ma for 2 hours. For protein detection the gels were placed in 100 ml of a solution of 5 parts methanol, 5 parts water, and 1 part acetic acid containing 50 mg Acid Black I. The gels were agitated 3 hr at room temperature. Excess stain was removed by additional washing with the methanol, water acetic acid mixture.

RESULTS

Strain analysis: In the initial phase of each assay, 15 to 20 individuals from a given strain were examined. If no variation among them was observed for the particular protein under study, the strain was arbitrarily chosen as a standard reference. In the case of variable proteins, testing was continued until a uniform strain was found. Analysis of the rest of the strains was then performed on gels with 12 sample pockets. The outer pockets (1 and 12) always contained samples from the standard strain while the remaining ten pockets contained five individuals from each of two strains. The result of a typical screening gel for the Esterase-5 protein is shown in Figure 1. The figure shows, in addition to the standard, some flies giving single bands (positions 2, 3, 5, 6, 7, 8, and 9) and some giving double bands (positions 10 and 11). These latter proved to be heterozygotes. The case with three clear bands (pocket 4) was shown to be a heterozygote with both parental enzymes and hybrid enzyme (see below). The homozygotes of the two alleles in question are shown flanking it in positions 3 and 5. In most cases the sampling procedure was repeated one or more times so that 10, 15, or more individuals from each strain were examined. When variants were detected in a strain, pure breeding stocks were obtained for genetic analysis and for cross checks of interstrain variability. When variants were obtained from two or more strains that apparently had the same mobility, they were in all cases examined together on the same gel to verify the finding. Single individuals from each strain that displayed no variability were also examined on single gels using sufficient combinations to verify this conclusion. Figure 2 illustrates the constancy of electrophoretic mobility obtained in this study from individuals from different strains.

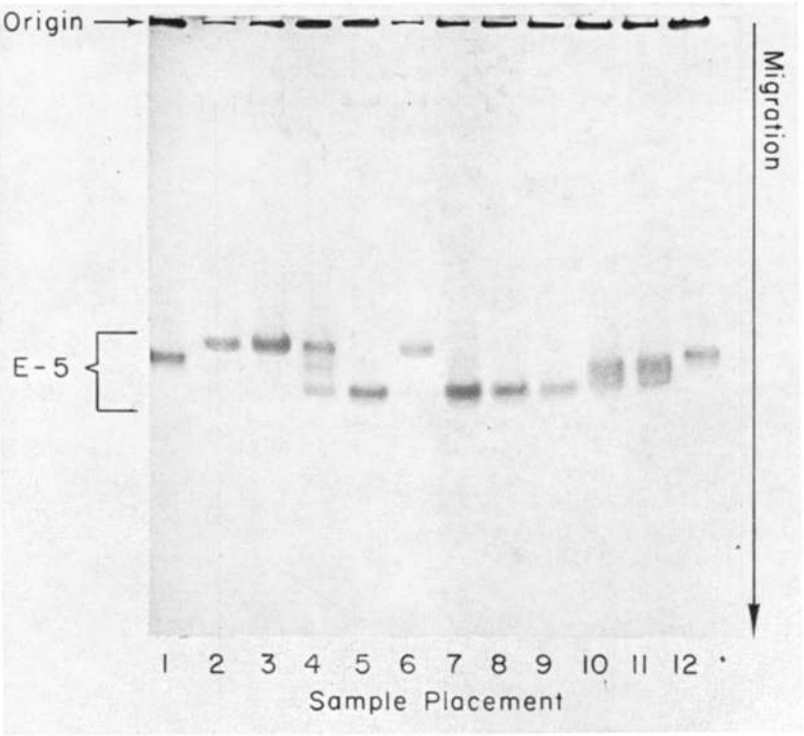

FIGURE 1.—Gel illustrating sample placement and typical results of strain analysis for Esterase-5. The first and the last samples were derived from the standard reference strain (E-$5^{1.00}$), while positions 2 through 6 were obtained from five individuals of one strain and positions 7 through 11 are from five individuals of a second strain. Positions 2, 3, and 6 contain Esterase-$5^{.95}$, position 5 contains Esterase-$5^{1.12}$, and position 4 contains Esterase-$5^{.95}$, Esterase-$5^{1.12}$, and a site of activity between them. Positions 7, 8, and 9 contain Esterase-$5^{1.12}$ and positions 10 and 11 contain Esterase-$5^{1.00}$ and Esterase-$5^{1.12}$. A site of activity midway between the latter two is barely discernible. In all the figures the direction of migration of the protein is down toward the anode.

Esterases: A number of esterases can be detected in *Drosophila pseudoobscura* with the present methods. Ten sites of activity have been seen when a number of individuals were pooled and used as the source of the enzyme. However, using single flies there is one conspicuous site of activity. Under the electrophoretic conditions employed here, this enzyme is the fifth esterase from the origin and will be designated Esterase-5 (E-5). From the populations sampled, individual female flies showed a variety of phenotypes with respect to Esterase-5. One strain, in which all of the individuals tested had a single site of activity with an average mobility of 6.3 cm, was chosen as a standard and this enzyme was assigned a relative mobility value of 1.00. Individual females from the various strains had a major site of esterase activity with relative mobilities of 0.85, 0.95, 1.00, 1.03, 1.07, and 1.12 (Figure 3A). Certain females exhibited two less intensely staining sites, each with one of the above mobilities and a third site of activity midway

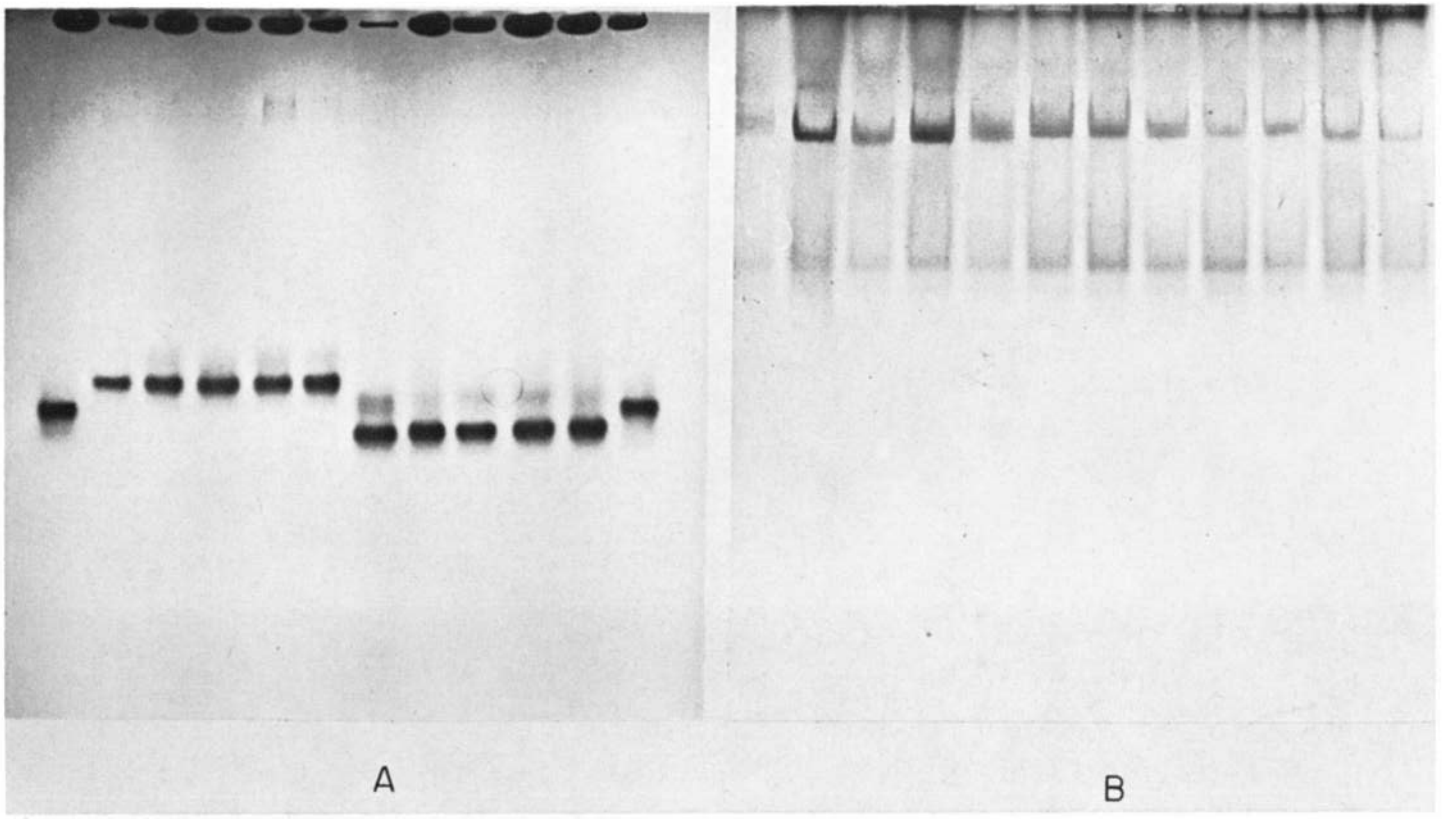

FIGURE 2.—Gels illustrating the constancy of electrophoretic mobility of the assays employed in this study. **A**. Assay of Esterase-5 activity. The sample placement is described in Figure 1. The three forms of the enzyme depicted here are Esterase-$5^{1.00}$ (sample 1 and 12), Esterase-$5^{.95}$ (samples 2 through 6), and Esterase-$5^{1.07}$ (samples 7 through 11). **B**. Assay of malic dehydrogenase. All individuals display Malic Dehydrogenase$^{1.00}$. (See text for explanation of nomenclature.)

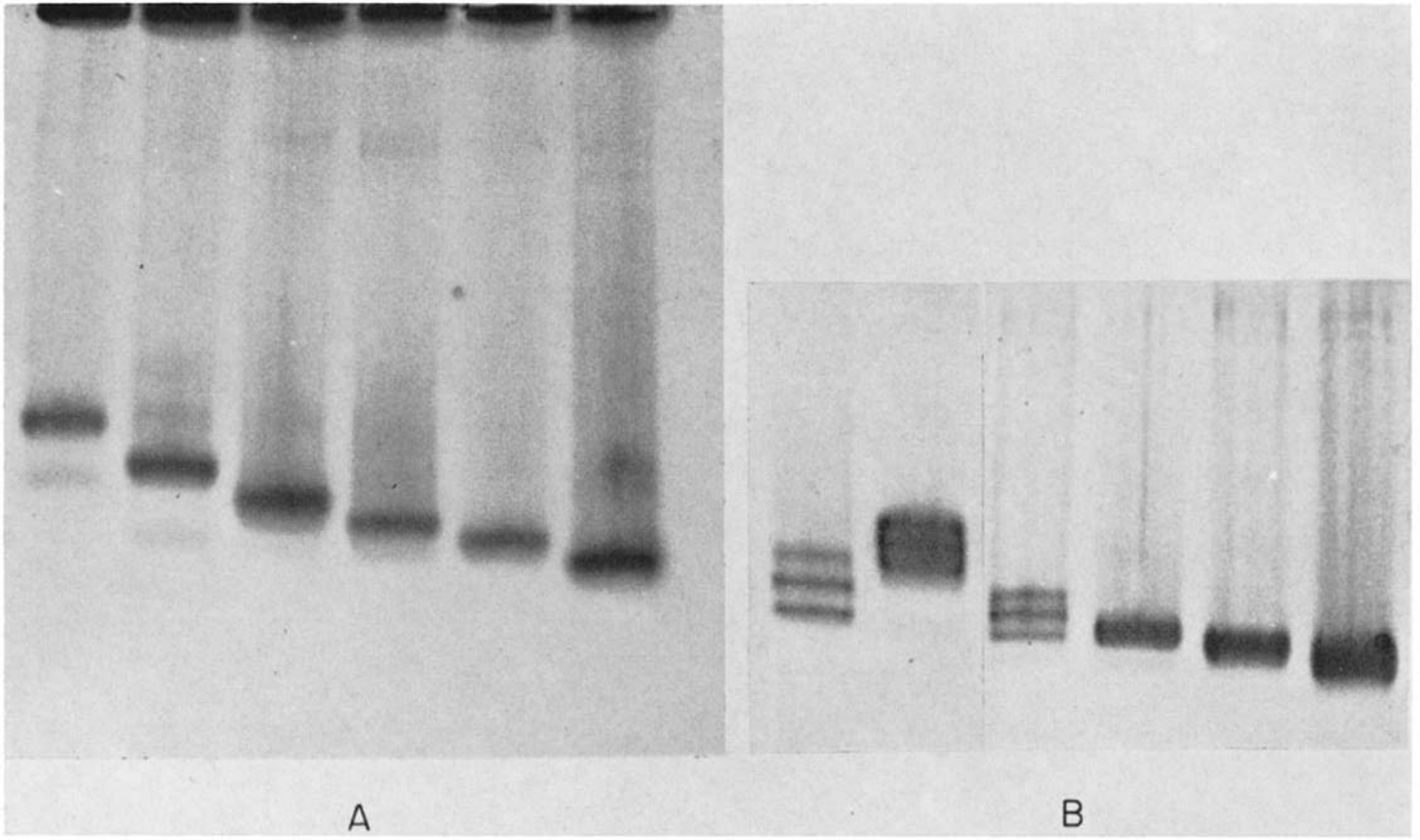

FIGURE 3.—**A**. Gel illustrating the six allelic forms of esterase-5. The results were obtained by assaying females homozygous for the following alleles (from left to right): $e\text{-}5^{.85}$, $e\text{-}5^{.95}$, $e\text{-}5^{1.00}$, $e\text{-}5^{1.03}$, $e\text{-}5^{1.07}$, and $e\text{-}5^{1.12}$. The other sites of esterase activity are not associated genetically with esterase-5. **B**. Gel illustrating hybrid enzyme formation in females heterozygous for various esterase-5 alleles. The genotypes of the individuals assayed are as follows (from left to right): $e\text{-}5^{.95}/e\text{-}5^{1.07}$, $e\text{-}5^{.85}/e\text{-}5^{.95}$, $e\text{-}5^{.95}/e\text{-}5^{1.03}$, $e\text{-}5^{1.00}/e\text{-}5^{1.03}$, $e\text{-}5^{1.03}/e\text{-}5^{1.07}$, and $e\text{-}5^{1.03}/e\text{-}5^{1.12}$.

between the two major sites. Thus one female exhibited major sites of activity at 1.00 and 1.12 and had an additional site at 1.06 (Figure 1A, 4th pocket). Males, on the contrary, never contained more than one Esterase-5 site of activity. It was concluded that the locus specifying the structure of this enzyme was probably sex-linked and that the three sites of activity encountered in females were a result of hybrid enzyme formation in a heterozygote. Genetic crosses fully supported these conclusions. A typical example will illustrate these findings. After sufficient eggs were laid from a pair mating, the parents were sacrificed and their Esterase-5 phenotype was determined. The female was found to have Esterase-5 activity with a relative mobility of 1.12 and the male was found to have a site of activity at relative mobility 0.85. All of the F_1 female progeny had sites of activity at 1.12, 1.02, and 0.85. All F_1 male progeny had one site of activity at a relative mobility of 1.12. Comparable results were observed for each of the variant forms. Moreover, in backcrosses of heterozygous females to assumed hemizygous males, sons were produced with either one or the other but never both of the grandparental forms, while both heterozygous and homozygous daughters were found. Thus the Esterase-5 differences behave as simple sex-linked characters.

The sex-linked genetic locus responsible for part or all of the structure of this protein will be designated esterase-5 (*e*-5) and the allelic forms of this locus will be distinguished by a superscript referring to the relative mobility of the protein produced by the allele. Thus the six alleles discovered thus far are $e\text{-}5^{.85}$, $e\text{-}5^{.95}$, $e\text{-}5^{1.00}$, $e\text{-}5^{1.03}$, $e\text{-}5^{1.07}$, and $e\text{-}5^{1.12}$, and the corresponding protein will be called Esterase-$5^{.85}$ (E-$5^{.85}$) etc. Figure 3A and B illustrates the six allelic forms of the enzyme detected and the results of assaying individuals heterozygous for certain pairs of these alleles.

Hybrid enzyme formation was detected in heterozygotes between all six of the allelic forms which had sufficient electrophoretic mobility differences to allow discrimination (Figure 3B). It was not possible to detect hybrid formation between E-$5^{1.00}$ and $E^{1.03}$, nor between $E^{1.03}$ and E-$5^{1.07}$. In both cases, however, the stained areas merged in such a manner to indicate that a hybrid molecule was probably present.

It was possible to show that the intermediate band was indeed hybrid enzyme by creation of hybrid molecules *in vitro*. Flies from two strains differing in mobility were macerated and the mixed supernatants frozen in the presence of 1 M NaCl overnight. These were then thawed and applied to a gel for electrophoresis and examination of Esterase-5. An intermediate band was produced by this treatment together with bands corresponding to the original strain proteins (HUBBY and NARISE, in preparation).

An effort was made with this enzyme to determine if alleles producing an enzyme with the same electrophoretic mobility were identical in other respects when they came from different populations. Accordingly a series of heat treatments of Esterase-5 enzymes were performed. Esterase-$5^{1.03}$ differed sharply from the other five forms of the enzyme. Heating the enzyme extract for 5 minutes at 50°C completely destroyed the activity of E-$5^{1.03}$ and had no apparent effect on the other five forms of Esterase-5. E-$5^{1.03}$ was found in two localities,

Wildrose and Mather, California. Both isolates from these localities showed the same degree of heat sensitivity. No differences in heat sensitivity at other temperatures were observed with the other five alleles.

Malic dehydrogenase: Malic dehydrogenase exists in multiple forms in *D. pseudoobscura*. The majority of the strains studied had a major band of enzymatic activity approximately 4.7 cm from the origin. This enzyme was chosen as the standard and was given a relative mobility value of 1.00 (Figure 2B). A second band of activity was usually observed at approximately 5.2 cm from the origin in forms which also contained the standard form of the enzyme. This second staining area was considerably lighter than the major band. Also under the assay condition another staining area was uniformly present. However, this enzyme was not substrate specific, and only required the presence of NAD, PMS and the dye for detection (see SIMS 1965).

Three other forms of malic dehydrogenase were detected from these strains. In all cases, the dominant site of activity had a different relative mobility from the standard. Thus, forms with relative mobility 0.90, 1.11, and 1.20 were demonstrated. In pure breeding strains for these enzymes the multiple form of the enzyme had a corresponding change in mobility. Those forms with the major site of activity at 1.00 had a minor site at 1.10, while those individuals with 0.90 and 1.20 had minor sites at 1.0 and 1.3, respectively. Strains pure breeding for the four forms of the enzyme were isolated and crosses were made reciprocally between all possible pairs. F_1 females and males did not differ in their phenotypes. In each cross, however, the F_1 individuals displayed a new site of activity midway between the mobility of the parental forms. In a cross of individuals with the 1.2 form of the enzyme by individuals with the 0.9 form, the F_1 flies had sites of activity at 1.2, 0.9, and 1.05. The staining activity was roughly in a 1:1:2 ratio (Figure 4). Faint bands of the multiple forms of the enzyme were still discernable.

To determine which chromosome controls the differences in malic dehydrogenase, a cross was made between a strain pure breeding for the 1.20 form of the enzyme and each of the dominant marker stocks (all pure breeding for the 1.00 form of the enzyme). For each cross, F_1 males heterozygous for the dominant marker were then backcrossed to females from the marker stock. The backcross progeny not carrying the dominant marker were then assayed for their enzyme constitution. In the case of crosses involving the 2nd and 3rd chromosome markers, the wild-type backcross progeny fell into two classes: flies with only the 1.00 form (i.e., homozygotes) and flies with both forms (heterozygotes 1.00/1.20). In the case of the 4th-chromosome marker crosses, however, every wild-type backcross individual possessed both forms and was presumably a heterozygote, 1.00/1.20. Thus, it was concluded that the locus specifying malic dehydrogenase was on the fourth chromosome. The alleles at this locus will be designated $mdh^{.90}$, $mdh^{1.00}$, $mdh^{1.11}$, and $mdh^{1.20}$.

Glucose-6-phosphate dehydrogenase: Glucose-6-phosphate dehydrogenase was detected as a substrate specific site of activity which migrated an average of 3.8 cm under the conditions employed. No electrophoretic variants of this enzyme were detected within the strains analyzed.

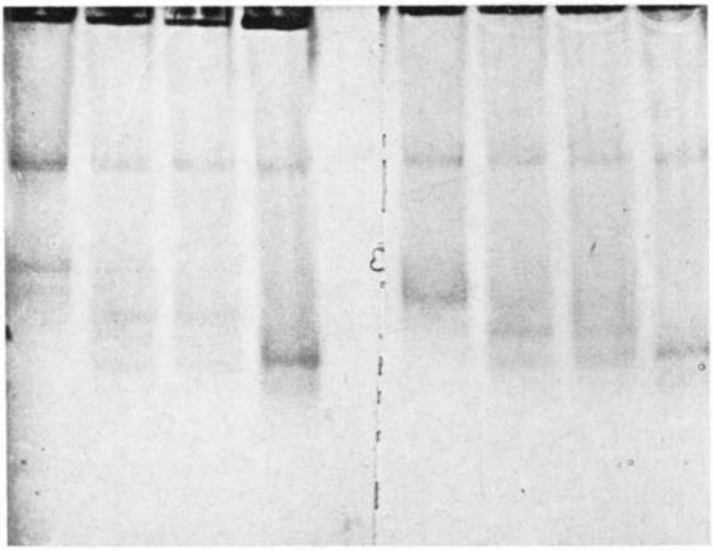

FIGURE 4.—Gel illustrating the behavior of malic dehydrogenase in individuals heterozygous for different alleles. The first site of activity (from top to bottom) is not specific for malate (see text). The genetic constitution of the individuals assayed is as follows (from left to right): $mdh^{.90}/mdh^{.90}$, $mdh^{.90}/mdh^{1.20}$, $mdh^{1.20}/mdh^{.90}$, $mdh^{1.20}/mdh^{1.20}$, $mdh^{1.00}/mdh^{1.00}$, $mdh^{1.00}/mdh^{1.20}$, $mdh^{1.20}/mdh^{1.00}$, $mdh^{1.20}/mdh^{1.20}$. Note the hybrid enzyme present in the heterozygote and the very faint multiple sites of activity in both homozygotes and heterozygotes.

α-glycerophosphate dehydrogenase: Assays of α-glycerophosphate dehydrogenase revealed two substrate sites of activity migrating to an average distance of 2.9 and 3.3 cm under the electrophoretic conditions. The enzyme with the fastest migration rate displayed the most activity. No electrophoretic variants of these proteins were observed. As in the malic dehydrogenase assay system, a site of activity was found which required only the presence of NAD, PMS, and dye for detection (Figure 5).

Alkaline phosphatase: Five sites of alkaline phosphatase activity were usually demonstrated from late third-instar larva (Figure 6). Occasionally two other faintly stained areas appeared on the gel. These seven sites will be designated Alkaline phosphatase (Ap) 1 through 7 in order of increasing mobility. These seven sites had an average absolute mobility of 1.0, 2.5, 2.6, 3.3, 3.9, 4.4, and 5.0 cm. Ap-4 displayed the most activity under our assay conditions. Ap-1 and Ap-2 did not vary electrophoretically in the samples studied, but were positively identified on only about 60% of the gels. Ap-3 and Ap-5 were also found too infrequently for analysis, thus they will not be considered further in this study. Ap-4 had two electrophoretic variants in the populations sampled. In one of the 43 strains, all individuals sampled had a site of activity slightly slower than the usual site of Ap-4. In reciprocal crosses between individuals pure breeding for the fast and the slow migrating forms of the enzyme, all F_1 females yielded a site of activity intermediate and overlapping the parental forms of the enzyme; males always displayed the form of the enzyme characteristic of their female

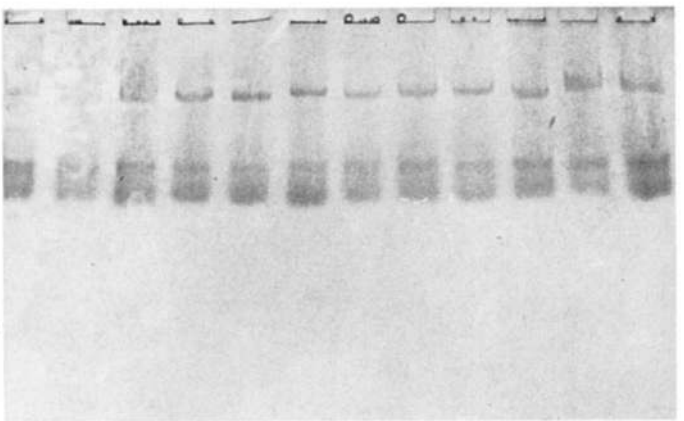

FIGURE 5.—Gel illustrating α-glycerophosphate dehydrogenase activity. From the top of the figure to the bottom is shown (a) site of activity not specific for substrate (see text), and (b) the two substrate specificities of activity.

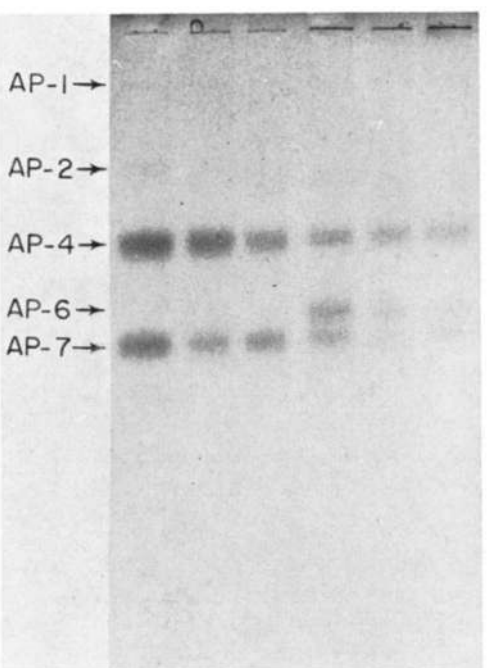

FIGURE 6.—Gel illustrating alkaline phosphate activity. Alkaline phosphatase 1 and 2 are not discernible in this picture. Alkaline phosphatase-4, -6 and -7 are shown as darkly stained areas. The three individual samples on the left display no activity for Alkaline phosphatase-6 (Ap-6$^-$) while the samples on the right display activity for this enzyme.

parent. Thus it was concluded that the locus controlling the enzyme was sex-linked. This conclusion was supported by production of one or the other parental forms of the enzyme in male progeny from F_1 females. The sex-linked locus

specifying Ap-4 will be designated *ap-4* and the two alleles detected in this study will be known as *ap-4*$^{.93}$ and *ap-4*$^{1.00}$, respectively, for the slow and the fast forms of the enzyme they produce. Ap-4 activity was extremely sensitive to the presence of Mg^{++} ions in the gel. Gels made without including this ion revealed the other sites of alkaline phosphate activity but rarely showed Ap-4 activity.

Ap-6 showed no electrophoretic variant but varied considerably in amount of activity among strains and among individuals of the same strain. Ap-6 was absent in all individuals of one strain.

In reciprocal crosses between individuals with and without a site of activity at Ap-6, all F_1 individuals displayed activity at that site. Because of the variable activity of this enzyme, no clear statement as to whether the F_1 individuals differ from the parental type that produces the enzyme may be made at present. In backcross of F_1 females to males without Ap-6 activity, progeny displaying and not displaying enzymatic activity at Ap-6 were found. The linkage of the gene responsible for Ap-6 has yet to be determined. The alleles at the locus specifying this enzyme will be designated *ap-6*$^+$ and *ap-6*$^-$ for the active and inactive forms, respectively (Figure 6).

Ap-7 was found to behave in a strictly analogous fashion to Ap-6. Thus no electrophoretic differences were found (however, see below), the activity varied considerably, and one strain was found in which no activity was detected in the great majority of individuals. From inbred lines derived from this strain, a stock uniformly deficient in the enzymatic activity was derived and reciprocal crosses were made to flies uniformly displaying this particular enzymatic activity. F_1 progeny did not differ in activity of the enzyme in relation to sex and uniformly possessed activity. Progeny from a backcross of F_1 females to males lacking the enzyme segregated for presence and absence of this activity. The alleles producing active and inactive forms are designated *ap-7*$^+$ and *ap-7*$^-$, respectively. During the course of analysis of strains, a single individual was found to have two sites of activity in the area of Ap-7 activity. Both sites were of equal intensity, one occupied the usual migratory area of Ap-7 and the other was slightly faster. This could be explained by the presence of heterozygous codominant alleles at *ap-7*, but since this has not been found again, this interpretation must remain conjectural.

Leucine aminopeptidase: Leucine aminopeptidase proved to be the most difficult assay system to analyze. As many as eight sites of activity with differing mobility were detected in inbred lines using single individuals. The enzyme migrating second from the origin had the greatest activity and was most consistently observed. Strains were found that had four different electrophoretic forms of this enzyme. The difference in migration rates were extremely slight and could only be detected by side-by-side comparison in adjacent pockets of the gel. These differences, though slight, were consistent and characteristic of the strain. F_1 individuals obtained by crossing strains with the greatest difference in migration rates had a site of activity overlapping the two parental forms. This site could not be further characterized as to number or intensity of more discrete areas. There were no detectable differences between reciprocal crosses. Results of an examina-

tion of individuals from a backcross was consistent with an interpretation of a single parental type and a hybrid type. But we feel at present that more elaborate techniques will be necessary for full substantiation of these finds. The linkage relation of the gene controlling this enzyme, which we shall christen *lap-2*, is at present not known.

Larval proteins: More than 13 proteins may be observed from extracts of 3rd instar larva (Figure 7). These presumably derive largely from the hemolymph. Of these, nine can be routinely demonstrated. In this instance, rather than using arbitrary standard strains, their mobilities were calculated from a reference standard of bovine serum albumen. These proteins are designated Proteins (Pt) 1 through 13, and by inference or genetic test the loci responsible for their production are called *pt-1* through *pt-13*. Pt-7, 8 and 10 were the most prominently stained proteins and displayed electrophoretic variability. In two cases, three

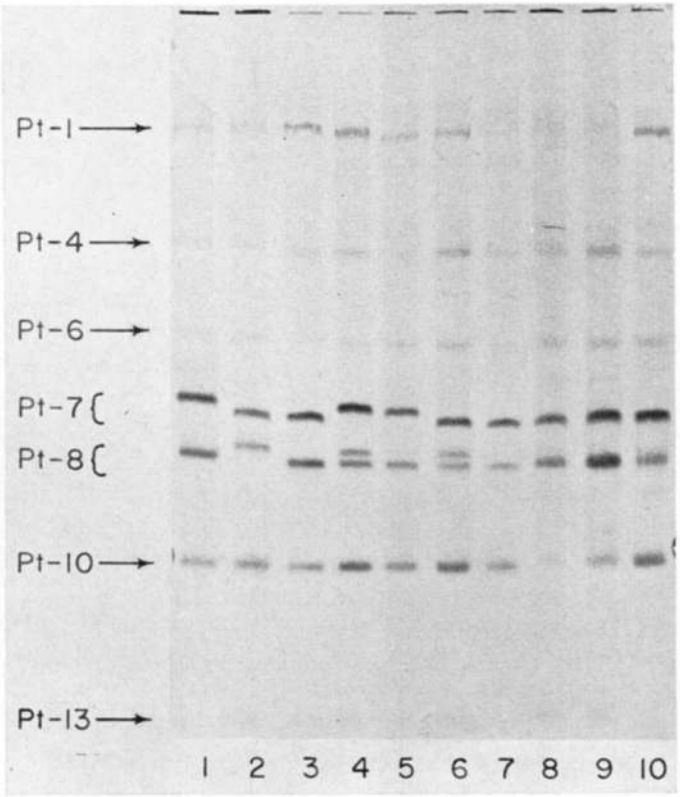

FIGURE 7.—Gel illustrating larval proteins. Only Pt-1, 4, 6, 7, 8, 10, and 13 are clearly reproduced photographically. This is an example of a screening gel in which both Pt-7 and Pt-8 were shown to vary. The genotypes, inferred from this photograph and substantiated by genetic analysis, of the individuals illustrated here are as follows (from left to right): (1) pt-$7^{.73}$/pt-$7^{.73}$, pt-$8^{.85}$/pt-$8^{.85}$ (2) pt-7/-$^{.75}$/pt-$7^{.75}$, pt-$8^{.83}$/pt-$8^{.83}$ (3) pt-$7^{.73}$/pt-$7^{.73}$, pt-$8^{.85}$/pt-$8^{.85}$ (4) pt-$7^{.73}$/pt-$7^{.73}$, pt-$8^{.83}$/pt-$8^{.85}$ (5) pt-$7^{.73}$/pt-$7^{.73}$, pt-$8^{.85}$/pt-$8^{.85}$ (6) pt-$7^{.75}$/pt-$7^{.75}$, pt-$8^{.83}$/pt-$8^{.85}$ (7, 8, 9, 10) pt-$7^{.75}$/pt-$7^{.75}$, pt-$8^{.85}$/pt-$8^{.85}$.

different forms of the protein were detected, while in the third, four forms were found. The forms of the proteins were named according to their relative mobilities and are as follows: Pt-$7^{.73}$, Pt-$7^{.75}$, Pt-$7^{.77}$; Pt-$8^{.80}$, Pt-$8^{.81}$, Pt-$8^{.83}$; and Pt-$10^{1.02}$, Pt-$10^{1.04}$, Pt-$10^{1.06}$, Pt-$10^{1.08}$. Single individuals contained at a maximum only two of the possible forms and strains pure breeding for nine of the possible ten forms were obtained (the exception was pt-$10^{1.08}$). In each case all F_1 individuals from reciprocal crosses of strains with different forms of the proteins displayed both parental forms. In backcrosses of F_1 females to one or the other pure breeding strain, two types of progeny were observed: those that displayed only a protein of the form of the backcross parent and those that displayed both forms as in their F_1 parent. Thus the genetic loci responsible for the production of these proteins are called *pt-7*, with three allelic forms *pt-7*$^{.73}$, *pt-7*$^{.75}$, and *pt-7*$^{.77}$; *pt-8*, with three allelic forms *pt-8*$^{.80}$, *pt-8*$^{.81}$, and *pt-8*$^{.83}$; and *pt-10*, with four allelic forms *pt-10*$^{1.02}$, *pt-10*$^{1.04}$, *pt-10*$^{1.06}$, and *pt-10*$^{1.08}$.

The chromosomal location of the three loci were determined as follows. First the genetic constitution of the marker stocks were assayed for these three loci. It was found that the second chromosome marker stock was homozygous for *pt-7*$^{.75}$, *pt-8*$^{.83}$, and *pt-10*$^{1.04}$. The marker stock for the third chromosome was homozygous for *pt-7*$^{.75}$ and *pt-8*$^{.81}$, but was consistently heterozygous for *pt-10*$^{1.04}$ and *pt-10*$^{1.06}$. The balanced lethal marker stock for the fourth chromosome contained both *pt-7*$^{.75}$ and *pt-7*$^{.77}$ in heterozygous and homozygous condition, and was homozygous for *pt-8*$^{.83}$ and *pt-10*$^{1.04}$. Thus it was concluded that *pt-10* was probably on the third chromosome and that *pt-7* was probably not on the fourth chromosome. Secondly, the marker stocks were then crossed with strains carrying different alleles; F_1 males and females carrying the dominant marker were then inbred and their progeny assayed as larva. If progeny were found that were homozygous for the allele characteristic of the marker stock, it was concluded that that allele was *not* linked to the chromosome carrying the dominant marker since the dominant markers were homozygous lethal and were associated with crossover suppressors. From such analysis it was concluded that *pt-7* and *pt-8* were located on chromosome 2 and *pt-10* was located on chromosome 3. *pt-10*$^{1.08}$ was found in a single individual (as a probable heterozygote with *pt-10*$^{1.06}$) and never recaptured on subsequent tests. Pt-1, 4, 6, 11, 12, and 13 were, in addition to Pt-7, 8, and 10, observed in the analysis of the strains though not positively identified in each sample. Of these, Pt-1, Pt-4, Pt-6, and Pt-13 were found in all 43 strains. Pt-12 was identified in 36 strains, Pt-11 in 34, Pt-5 in 18, and Pt-9 in 15. Pt-4, 6, 11, and 12 displayed no variability in mobility. The migration of Pt-1 ranged from a relative mobility of .21 to .23 (see Figure 7). This variation was not consistent or reproducible. It was not associated in any manner with the strain derivation of the individual, nor were any pure breeding strains obtained that differed consistently from one another. Since the migration of this protein was the slowest of all the proteins tested, it was consequently the least reliable. We, therefore, tentatively conclude that this variation was not genetically determined. Pt-13, the most rapidly migrating protein, was found in two conditions. In 39 of the 43 strains investigated, this protein migrated with a relative mobility of 1.30. In

two strains, a protein was seen at a relative mobility of 1.28, in addition to the usual form at 1.30. On subsequent tests with these two strains, this additional band was not observed.

GENERAL CONCLUSIONS

Direct evidence has been obtained for eight genetic loci and strong evidence was presented for a ninth locus controlling various proteins in *Drosophila pseudoobscura*. From analyses of these loci it was possible to distinguish 6 alleles at one locus (*esterase-5*), 4 alleles at three separate loci (*mdh*, *lap-2*, and *pt-10*), 3 alleles at two separate loci (*pt-7* and *pt-8*) and 2 alleles at three loci (*ap-4*, *ap-6*, and *ap-7*). Of these nine loci, six have been found to be linked to one of the five chromosomes of *D. pseudoobscura*. These include esterase-5 and alkaline phosphatase-4 on the X chromosome, *pt-7* and *pt-8* on the second chromosome, *pt-10* on the third chromosome, and malic dehydrogenase on the fourth chromosome; *ap-6*, *ap-7*, and *lap* are yet to be placed on particular chromosomes, but they are known to behave as single Mendelian genes in backcrosses. In addition 12 other proteins were found that did not vary electrophoretically. Thus no *direct* evidence could be obtained as to their genetic control. *However, we have no reason to believe that the dictums of modern genetic theory do not pertain to these proteins.* Consequently, we would include these as examples of gene products of loci showing no allelic variation. This is, of course, an approximation since not all genetic changes give rise to electrophoretic variability. Thus this technique only provides us with a measure of the *lower limit* of all possible genic variance at a particular locus. Nevertheless, of the 21 loci investigated, it was possible to reliably demonstrate more than one allele in nine of them.

The distribution of variable loci over the various functions needs to be considered when we use these differences to ask about genetic variation in the genome as a whole. First, we are obviously sampling only structural gene loci rather than controlling genes. Second, we have deliberately put weight on surveying a variety of enzymes and proteins rather than exhaustively studying many enzymes with the same general specificity. There are ten sites of esterase activity detectable on gels, but we have studied only the one with high activity under our conditions. In the case where a number of related enzymes have been studied, of the five alkaline phosphatases, three were variable and two constant. Among the remaining five enzyme proteins, three are also variable and two constant. Thus there does not seem to be any difference between a wide sample of function and a narrow one.

The larval proteins are single proteins in high concentration. None of the specific enzyme protein shows up with the acid Black I stain, presumably because they are in too low concentrations. Of the 11 proteins, three showed a reliable genetic variation and one, Pt-13, showed variation that has not yet been confirmed. This is less variation than the six out of ten among the enzymes analyzed, but not grossly different. It may in part be due to the smaller sample size analyzed for Pt-5, 9, 11, and 12. This overall similarity in variation within the two groups

of proteins allows us, at least as a first approximation, to use them as a representative sample of structural genes in the genome of *D. pseudoobscura.*

The chromosomal location of some of the genes involved deserves some brief comment as to their possible homology to other Drosophila species. In *D. melanogaster* both esterase-6 and esterase C have been located on the left arm of chromosome 3 (element D) (WRIGHT 1963; BECKMAN and JOHNSON 1964c). *Esterase-5* in *D. pseudooobscura* is on the X chromosome, which is composed of elements A and D. Thus the esterase in *D. pseudoobscura* may have had a common origin with one or the other of the enzymes in *D. melanogaster.* However, they are clearly not identical since they differ in electrophoretic mobility and in ability to form hybrid molecules in heterozygotes. An alkaline phosphate gene was also found on chromosome 3 of *D. melanogaster* (BECKMAN and JOHNSON 1964a). Since one of the enzymes with this activity is sex linked in *D. pseudoobscura,* there is again the possibility of genetic homology for this enzyme between these species. It should be pointed out that techniques like these are applicable to most other organisms and thus provide the means for similar analysis of other kinds of species.

We are greatly indebted to DR. SUMIKO NARISE who made very important contributions both to the techniques described and to the survey of the strains for enzyme differences. We also wish to thank MR. ALLAN NOVETSKY for his contribution to this work in the analysis of larval proteins.

SUMMARY

In order to estimate the amount of genic variation in a natural population, techniques are needed that will detect isoallelic variation at many individual loci. Using electrophoretic mobility in acrylamide gels, we have characterized ten enzyme proteins and 11 other proteins from strains of *D. pseudoobscura* representing five geographic localities. Electrophoretic variants, shown by genetic tests to be controlled by single loci, were found at 9 out of 21 loci as follows: an esterase, 6 alleles; malic dehydrogenase, 4 alleles; leucine aminopeptidase, 4 alleles; three different alkaline phosphatases, 2 alleles each; two different larval proteins, 3 alleles and one larval protein with 4 alleles. The 12 proteins showing no reliable genetic variation were glucose-6-phosphate dehydrogenase, α-glycerophosphate dehydrogenase, two different alkaline phosphatases, and eight larval proteins. Two of the variable loci were sex linked, two are on chromosome 2, one on 3 and one on chromosome 4.

LITERATURE CITED

BECKMAN, L., and F. M. JOHNSON, 1964a Variations in larval alkaline phosphatase controlled by *Aph* alleles in *Drosophila melanogaster.* Genetics **49:** 829–835. —— 1964b Genetic control of aminopeptidases in *Drosophila melanogaster.* Hereditas **51:** 221–230. —— 1964c Esterase variations in *Drosophila melanogaster.* Hereditas **51:** 212–220.

CROW, J. F., 1961 Population genetics. Am. J. Human Genet. **13:** 137–150.

DOBZHANSKY, TH., 1951 *Genetics and the Origin of Species.* Columbia University Press, New York.

DOBZHANSKY, TH., and B. SPASSKY, 1954 Genetics of natural populations. XXII. A comparison of the concealed variability in *Drosophila prosaltans* with that in other species. Genetics **39**: 472–487.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford.

HUBBY, J. L., 1963 Protein differences in Drosophila. I. *Drosophila melanogaster.* Genetics **48**: 871–879.

LEWONTIN, R. C., and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura.* Genetics **54**: 595–609.

MAYR, F., 1963 *Animal Species and Evolution.* Harvard University Press, Cambridge.

MORTON, N. F., 1960 The mutational load due to detrimental genes in man, Am. J. Human Genet. **12**: 348–364.

SIMS, M., 1965 Methods for detection of enzymatic activity after electrophoresis on polyacrylamide gel in Drosophila species. Nature **207**: 757–758.

SPENCER, W. P., 1957 Genetic studies on *Drosophila mulleri.* I. The genetic analysis of a population. Univ. Texas Publ. **5721**: 186–205.

WALLACE, B., 1958 The role of heterozygosity in Drosophila populations. Proc. 10th Intern. Congr. Genet. **1**: 408–419.

WRIGHT, T. R. F., 1963 The genetics of an esterase in *Drosophila melanogaster.* Genetics **48**: 787–801.

YOUNG, W. J., J. E. PORTOR, and B. CHILDS, 1964 Glucose-6-phosphate dehydrogenase in Drosophila: X-linked electrophoretic variants. Science **143**: 140–141.