A MOLECULAR APPROACH TO THE STUDY OF GENIC HETEROZYGOSITY IN NATURAL POPULATIONS IV. PATTERNS OF GENIC VARIATION IN CENTRAL, MARGINAL AND ISOLATED POPULATIONS OF *DROSOPHILA PSEUDOOBSCURA**†

SATYA PRAKASH, R. C. LEWONTIN, AND J. L. HUBBY

Department of Biology, The University of Chicago, Chicago, Illinois 60637

Received November 12, 1969

ACCORDING to the synthetic theory of evolution, populations of a species diverge in their gene pool as a result of differential selection of genetic variation in these populations and as a result of random genetic drift. The kind and magnitude of genetic divergence between populations is governed mainly by interaction of population size, migration rate and selection intensity. In order then to understand evolutionary phenomena in natural populations, one has to start by determining the pattern of genetic variation in populations which come from different ecogeographic regions of the species range and represent different degrees of isolation from the other populations of the species. By using acrylamide gel electrophoresis we have shown that for a random sample of loci, a great deal of genic variation exists in *Drosophila pseudoobscura* (HUBBY and LEWONTIN 1966; LEWONTIN and HUBBY 1966). More recently we have demonstrated selection of different alleles of the *Pt-10* and *Amy* loci in different gene arrangements of *D. pseudoobscura* and *D. persimilis* (PRAKASH and LEWONTIN 1968). This variation is thus not merely isoallelic variation, of no significance in the process of natural selection.

It is the purpose of the present paper to examine the patterns of genetic variation in central, geographically peripheral and isolated populations of *D. pseudoobscura.* Such a study answers three questions: First, how widespread is the high degree of polymorphism we have previously reported? Is it characteristic of all populations of the species? Second, is there any relationship between the degree of heterozygosity in a population, and that population's location in the species range? For example, are populations at the margins of the species distribution more homozygous? Does isolation from the rest of the species lead to homozygosity? Finally, is there a significant variation in allele frequencies among populations, of clinal variation, of ecotypes? The answers to these questions will in turn tell us the relative roles of selection and random processes in determining the genetic makeup of the species.

We have studied 24 randomly chosen loci. We will show that while the central

* This work was performed under a Grant from the National Science Foundation GB 7135.

† Dedicated to Th. Dobzhansky on his seventieth birthday in recognition of his inspiring leadership of the study of evolutionary genetics.

populations are the most variable, it is the isolated (Bogota) population that has the least amount of genic variation and maximum genetic divergence when compared with the central (Strawberry Canyon in California) and marginal populations (Mesa Verde in Colorado and Austin in Texas). The divergence between the latter three populations is not very great. It is the similarity of these three populations that is striking and relevant to the questions we have just raised.

MATERIALS AND METHODS

Assay methods for larval proteins, leucine aminopeptidase, glucose-6-phosphate dehydrogenese, malic dehydrogenese and esterase have been previously described (HUBBY and LEWONTIN 1966). We have dropped α-glycerophosphate dehydrogenase because of poor resolution. Acid phosphatases of third instar larvae were studied rather than alkaline phosphatases as reported previously (HUBBY and LEWONTIN 1966). For detection of acid phosphatase activity the gels were incubated in 100 ml of 0.1 M acetate buffer pH 5.0 rather than Tris HCl pH 8.5 buffer. The rest of the method remains the same. For α-amylase, gels were prepared with 0.1 M Tris-borate buffer at pH 8.9 with 0.02 M $CaCl_2$. This buffer without $CaCl_2$ was used in the compartments of the gel box. The lower edges of the 12 pocket slot former were filed flush in order to have the enzyme sample run close to the surface of the gel. Third instar larvae or adults were used for the samples. Electrophoresis was carried out at 200 to 400v and 100mA for 3 hr. For detection of α-amylase the lower surface of the gel was wiped free of moisture and a staining plate was applied to this surface and weighted down. This sandwich was then held at 4°C for 2 days, after which the staining plate was removed and placed in KI-I solution for 2 min. The presence of α-amylase was shown by an unstained band on the plate (see Figure 2). The staining plates were prepared by a method given to us by DR. WINIFRED DOANE of Yale University to whom we are extremely grateful. A clean, dry glass plate, 12.5 × 17.5 cm, is edged with a 4 mm band of masking tape. 2.25 g hydrolyzed starch were boiled for exactly 5 min in 60 cc of water. This is mixed with 15 cc of 1 M Tris HCl at pH7.4 containing 3 ml of 1 M CaCl and 7.5 g cyanogum. The total mixture is placed on the edged glass and a clean dry plate is lowered over it, taking care to keep out all bubbles. This double plate is stored in buffer at 4°C. until needed, whereupon the top plate is separated and discarded.

For xanthine dehydrogenase, octanol dehydrogenase and acetaldehyde oxidase, gel, sample preparation, and electrophoresis were the same as for esterase (HUBBY and LEWONTIN 1966). For xanthine dehydrogenase detection, gels were incubated in a mixture of 100 ml 0.1 M Tris HCl buffer pH 7.5, 12 ml .05 M hypoxanthine, 30 mg NAD, 50 mg Nitro BTR, 0.2 ml of 1 M KCl. After 1–2 hr, 2 mg of phenazine methosulphate were added. To detect octanol dehydrogenase we incubated the gel in 100 ml 0.1 M Tris HCl buffer at pH 8.5 to which was added 30 mg NAD, 25 mg Nitro BTR and 0.5 ml octyl alcohol. After 2 hr, 2 mg of phenazine methosulfate were added. For aldehyde oxidase the gel was immersed in 100 ml 0.1 M Tris HCl buffer pH 8.5, with 25 mg Nitro BTR, 4 mg phenazine methosulfate and 8 ml acetaldehyde. The "oxidase" band is detected in dehydrogenase assays as a clear zone where no deposition of reduced Nitro BTR occurs.

Populations Studied: The limits of the range of the species not including Bogota are given in DOBZHANSKY and EPLING, 1944 (p. 12). Information about relative population sizes comes from personal observation and anecdotal evidence from other collectors.

Strawberry Canyon (California): This population is highly polymorphic for inversions in the third chromosome. The frequencies of different gene arrangements in this population are ST = .466, AR = .091, PP = 0.028, CH = 0.191, TL = 0.140, EP = 0.062, SC = 0.014, OL = 0.007 (STRICKBERGER and WILLS 1966). The population is dense and is in the center of abundance of the species. A total of one hundred and ten strains were collected throughout 1966 by ALAN WICK and CHRIS WILLS. By a strain we mean a line descended from a single wild-caught inseminated female. Each strain thus originally carried at least four wild autosomal complements and three *X* chromosomes. Larval proteins and esterase were studied from single indi-

viduals from each strain in F_1. The other assays were done on single individuals between the F_1 and F_5 generation of laboratory culture.

Mesa Verde (Colorado): A medium sized population near the eastern margin of the species. One hundred and twenty strains were collected in August, 1966. This sample had AR = 0.985, ST = 0.005, PP = 0.010 and so is nearly monomorphic for inversions. Proteins and esterase were studied in the F_1 by examining a single individual from each strain. The rest of the assays were carried out in succeeding generations up to the F_5.

Austin (Texas): A small population from the extreme southeastern margin of the species. Twenty-three strains were collected in April, 1967. Our sample had the following frequencies of third chromosome gene arrangements: ST = 0.06, PP = 0.78, AR = 0.08, TL = 0.06, CH = 0.01, OL = 0.01. For proteins and esterase, eight to ten individuals were examined in F_1 and the mating in nature thus deduced. Each strain then generally gave information on four wild autosomes and three wild *X* chromosomes. The rest of the enzyme assays were made by studying at least five individuals in F_2 to F_5. This was done in order to get the maximum information about the genetic variability in these relatively few strains.

Bogota (Colombia): This is a large population isolated from the rest of the species range to the north by 1500 miles. It carries SC and TL gene arrangements in frequencies of 0.627 and 0.374, respectively (MAYHEW, *et al.* 1966). For all of the assays, at least five individuals were examined from each strain in F_2 to F_6 since only 19 strains were available.

All the strains from different populations were maintained in large culture bottles at 18°C in optimum conditions.

RESULTS

Strain analyses were carried out as discussed in HUBBY and LEWONTIN, 1966. In brief, homozygous lines were derived for all the variants of a locus. We propose the term *allozyme* for the variant proteins produced by allelic forms of the same locus, to avoid the now common confusion with "isozymes" which are the various polymers produced from monomers specified by different loci. For classification, allozymes from different strains were always compared on the same gel with the variants from the derived homozygous lines. This procedure, though tedious and time consuming, avoids the risk of misclassification due to uncontrolled variations in the gel. Mendelian genetics has been done of the following new loci which are polymorphic: leucine aminopeptidase (*LAP*), xanthine dehydrogenase (*XDH*) and α-amylase (*Amy*). All three are autosomal loci. *Amy* is on the third chromosome (PRAKASH and LEWONTIN 1968). *XDH* is on chromosome II. *LAP* is not yet located on a specific autosome. Octanol dehydrogenase must be autosomal since both males and females give heterozygous band patterns. However, the genetics of octanol dehydrogenase (*ODH*), acetaldehyde oxidase-2, *Pt-12* and *Pt-13* remains to be done.

Both parental allozymes are present in the heterozygotes of different alleles of *LAP* (Figure 1), *Amy* (Figure 2), *Pt-12* and *Pt-13*. Three bands consisting of the two parental bands and an intermediate hybrid band were observed in the heterozygotes of different alleles of *XHD* and octanol dehydrogenase (Figure 3). For acetaldehyde oxidase-2 three bands are seen in *0.93/1.0* heterozygotes, but because of close mobilities, it is hard to say whether the *1.00/1.02* heterozygotes have two or three bands.

Calculation of gene frequencies: While the identification of single individuals from a strain can provide information about two wild genomes only, examina-

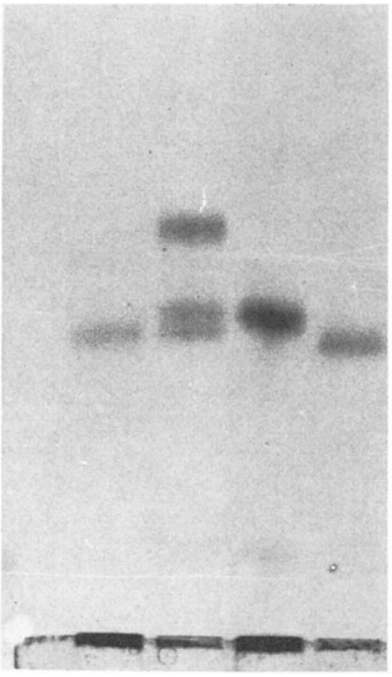

FIGURE 1.—A gel showing different genotypes of adult leucineaminopeptidase (from left to right) $LAP^{1.0/1.0}$, $LAP^{1.0/1.10}$, $LAP^{1.10/1.10}$, $LAP^{1.0/1.0}$.

tion of eight or more individuals from a strain provides information on all four wild genomes with a probability of .98.

Table 1 presents the loci that are polymorphic and the frequencies of alleles

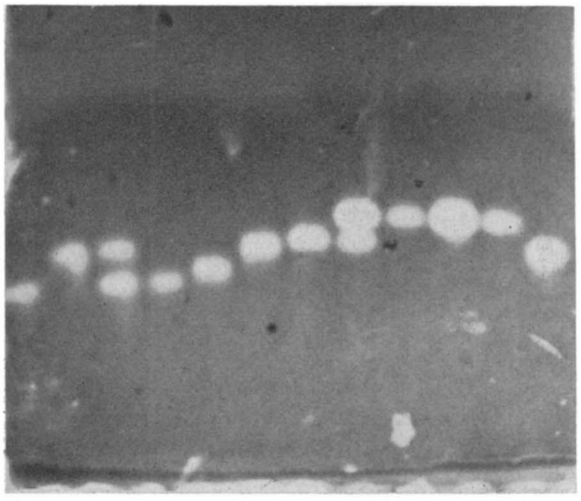

FIGURE 2.—A gel showing different genotypes of amylase in *D. pseudoobscura* and *D. persimilis*. (From left to right) $Amy^{.84/.84}$, $Amy^{1.0/1.0}$, $Amy^{.84/1.0}$, $Amy^{.84/.84}$, $Amy^{.92/.92}$, $Amy^{1.0/1.0}$, $Amy^{1.0/1.0}$, $Amy^{1.0/1.09}$, $Amy^{1.09/1.09}$, $Amy^{1.09/1.09}$, $Amy^{1.09/1.09}$, $Amy^{1.0/1.0}$.

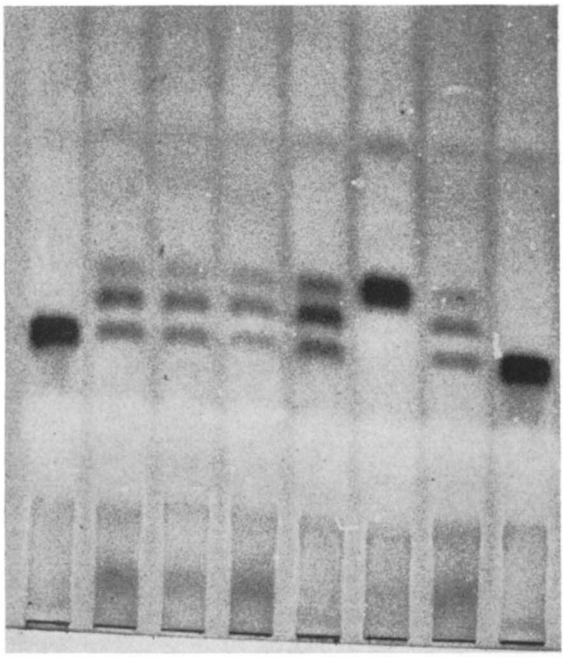

FIGURE 3.—A gel showing different genotypes of octanol dehydrogenase. Pockets 1 and 8, $ODH^{1.0/1.0}$; Pocket 6, $ODH^{1.22/1.22}$; Pockets 2, 3, 4, 5, and 7, $ODH^{1.0/1.22}$.

TABLE 1

Frequencies of alleles at various loci in the different populations of D. pseudoobscura

A

Pt-7

Chromosome II

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.68	.005	...	...	...
.73	.005	.010	.012	.050
.75	.954	.954	.966	.925
.77	.036	.036	.023	.025

B

Leucine aminopeptidase

Autosome

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.90	.008	.025	.043	...
.95	.050	.008	.022	...
1.00	.892	.940	.870	.95
1.10	.050	.025	.054	.05
1.12	...	...	.011	..

TABLE 1—Continued

C

Pt-8

Chromosome II

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.80	.014	.009	.011	.870
.81	.472	.410	.441	.100
.83	.514	.576	.512	.030
.85	...	.005	.035	...

D

Xanthine dehydrogenase

Chromosome II

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.90	.053	.016	.018	...
.92	.074	.073	.036	...
.99	.263	.300	.232	...
1.00	.600	.580	.661	1.00
1.02	.010	.032	.053	...

E

Esterase-5

X-Chromosome

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.85	...	.035	...	...
.90	...	...	.015	...
.95	.123	.114	.031	.03
.97	...	...	.031	...
1.00	.424	.364	.292	.97
1.02	.014	.048	.108	...
1.03	.080	.039	...	...
1.04	.004	.101	.154	...
1.07	.193	.197	.262	...
1.09	.009	...	...	...
1.12	.132	.101	.046	...
1.16	.019	...	.062	...

F

Pt-12 and *Pt-13* unmapped

Locus	Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
Pt-12	*1.18*	.94	1.00	1.00	1.00
	1.20	.06	...	...	...
Pt-13	*1.30*	1.00	1.00	1.00	.87
	1.37	...	...	...	.13

G

Larval acid phosphatase-4
X Chromosome

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.93	...	...	.028	...
1.00	1.0	1.0	.860	1.0
1.05	...	...	.112	...

H

Glucose-6-phosphate dehydrogenase
X Chromosome

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
1.00	1.00	.994	1.0	1.0
1.10	...	.006	...	...

I

Acetaldehyde oxidase-2
Unmapped

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.90	.01	...	...	...
.93	.03	...	...	.05
1.00	.94	1.0	1.0	.83
1.02	.02	...	...	.12

J

Malic dehydrogenase
Chromosome IV

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
1.00	.97	.95	.97	1.00
1.20	.03	.05	.03	...

K

Octanol dehydrogenase-1
Autosome

Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
.86	...	.013	...	...
1.00	.977	.960	1.00	1.00
1.22	.023	.026	...	...

TABLE 1—Continued
L
Pt-10 and *α-Amylase-1* Loci
Chromosome III

Locus	Allele	Strawberry Canyon	Mesa Verde	Austin	Bogota
Pt-10	*1.02*	.005	.021	.010	...
	1.04	.615	.970	.935	...
	1.06	.380	.008	.054	1.00
α-Amy-1	.74	.030	...	...	...
	.84	.290	.210	.125	1.00
	.92	...	...	...	...
	1.00	.680	.790	.875	...

segregating at different loci in different populations. We did not observe any segregation at 10 loci (*Pt-4, -5, -6, -9A, -9B, -11,* "*oxidase*" and *larval acid phosphatase-5, -6, -7*), and all populations had the same allele at each locus. It is clear from scanning Table 1 that a very great difference exists in the kind and amount of genetic variation between the North American populations of Strawberry Canyon, Mesa Verde, and Austin on the one hand and the isolated population of Bogota on the other. We will, therefore, discuss them separately.

North American populations:—We would first like to know the degree of genetic divergence between Strawberry Canyon (SC), Mesa Verde (MV), and Austin (AU) populations. We will consider the patterns of variation in these populations under the following categories:

1. *Widespread polymorphisms with no evidence of differentiation between populations:* There is no significant difference in the gene frequencies of different alleles at the *Pt-7*, *Pt-8*, *LAP*, *XDH* and *MDH loci.* Although there is no differentiation among the populations, the loci involved are by no means alike in their variation. They run the gamut from the nearly monomorphic *Pt-7* with one allele in frequency above .95, to the highly polymorphic *XDH* locus in which four different alleles are present in greater than .05 frequency in Strawberry Canyon. For this locus and for *Pt-8* with its 50–50 polymorphism of two alleles, no wild-type allele can be designated. Nevertheless, all three populations have the same allele frequencies. We will return to this point.

2. *Widespread polymorphisms with evidence of differentiation between populations:* The *Est-5* locus shows 12 alleles, with each population segregating for 8 or 9 forms with different frequencies of different alleles in different populations. Some alleles like *.85*, *.90*, *.97* and *1.09* were encountered only in one of the three populations. For other alleles there is evidence of differentiation as one moves from the western Strawberry Canyon to the eastern Mesa Verde and the southeastern Austin populations. While the frequency of the alleles *.95, 1.0, 1.03, 1.12* is highest in Strawberry Canyon, intermediate in Mesa Verde and lowest in Austin, the reverse appears to be true for the alleles *1.04 and 1.07*. A χ^2 test for homogeneity of allele frequencies in the three North American populations gives

a value of $\chi^2 = 95.45$ ($p << .001$). The presence of differentiation at this locus, together with the general similarity of allelic frequencies between populations, suggests that selection is holding these alleles in different frequencies in different populations.

3. *Local Polymorphisms: Pt-12, AP-4, G-6-PD* and *acetaldehyde oxidase-2* loci are each polymorphic in only one of the three North American populations, the *ODH-1* locus is polymorphic in Strawberry Canyon and Mesa Verde but monomorphic in Austin. In all cases, that allele which is present in the monomorphic populations is also the one in high frequency where the locus is variable. At one extreme is *G-6PD* which is essentially monomorphic with a single heterozygous individual for an aberrant allele (*1.10*) found in Mesa Verde. Presumably selection strongly favors the wild-type allele, *1.00*. At the other extreme is *ODH* with three alleles and with a suspicious similarity between Strawberry Canyon and Mesa Verde. *Pt-12* is probably a good deal more polymorphic than we report here, on the basis of some preliminary trials with a new very sensitive stain for larval proteins. The allele *1.22* certainly is not a newly arisen mutant in the process of elimination but a stable although minor part of the species gene pool. As for the other two classes of polymorphisms, selection seems to be the agent controlling the variation.

4. *Polymorphisms associated with the third chromosome:* The *Pt-10* and *Amy* loci are on the third chromosome of *D. pseudoobscura*. The third chromosome shows a great deal of inversion polymorphism and different populations of this species differ greatly in regard to the kind and amount of inversion polymorphism. As has been pointed out in the MATERIALS and METHODS section of this paper, Strawberry Canyon is the most polymorphic and Mesa Verde the least for third chromosome gene arrangements. The pattern of variation at these two loci provides the most striking differences observed between populations. At both *Pt-10* and *Amy* loci the greatest amount of polymorphism exists in Strawberry Canyon, while Mesa Verde is less heterogeneous than Austin at the *Pt-10* locus but more variable at the *Amy* locus. This pattern of variation of these loci is due to association of different alleles with different third chromosome gene arrangements. The differences in frequencies of alleles at these two loci in different populations are a reflection of strong differential selection of various gene arrangements in these populations. This relation between inversions and alleles at these loci has been thoroughly discussed by PRAKASH and LEWONTIN (1968).

Bogota: This population differs strikingly from the North American populations. In North America the greatest differentiation in various populations was observed at the *Pt-10* and *Amy* loci. At the *Pt-10* locus allele *1.06* is 38% in Strawberry Canyon, 0.8% in Mesa Verde, and 5.4% in Austin. At the *Amy* locus the allele *.84* is 29% in Strawberry Canyon and only 12.5% in Austin. At other loci, the differentiation between the three North American populations was not nearly as striking. Bogota, on the other hand, differs from the North American populations not only at the *Pt-10* and *Amy* loci but at most other segregating loci. While *Pt-7* shows a similar polymorphism to that of the rest of populations, at the *Pt-8* locus the allele. *80* is 87%. This allele is about 1% in other populations.

However, the alleles. *.81* and *.83* are still present in Bogota, but allele *.83* is down to 3% from about 50% in the northern populations. At the *leucine aminopeptidase* locus we find only two alleles in Bogota while there are four to five alleles in other populations. The frequency of *LAP* allele *1.10* is about the same in Bogota as elsewhere, about 5%. At the *Est-5* locus only two alleles were found in Bogota as opposed to 8 or 9 alleles in each of the other populations. The allele $Est\text{-}5^{1.00}$, which is the major allele of the northern populations, has reached a frequency of 0.97 in Bogota. Bogota also has about 3% of allele $Est\text{-}5^{.95}$. This allele is present in frequencies of 3 to 12% in the other samples. While rare alleles are maintained at *Pt-7*, *Pt-8*, *LAP* and *Est-5* loci, allele *1.0* appears to be fixed at *XDH*, *ODH*, *malic dehydrogenase*, *Pt-12* and *AP-4* loci. At the *XDH* locus allele *.99* is 23% to 30% in the North American populations but absent in Bogota. At the *Pt-10* and *Amy* loci it is the rare alleles $Pt\text{-}10^{1.06}$ and $Amy^{.84}$ that appear to be fixed These two loci show this pattern because of their association with the Santa Cruz and Treeline gene arrangements found in Bogota. The only rather surprising polymorphism is that of the *acetaldehyde oxidase-2* locus. This locus is segregating only in Strawberry Canyon and Bogota populations. At the *Pt-13* locus there is evidence of evolution of a different polymorphism as can be seen by the presence of allele *1.37* not found elsewhere. This locus is less reliable than others, however, and some recent tests of a new sensitive stain show evidence of greater polymorphism, reliably demonstrated.

DISCUSSION

At the beginning of this paper we posed three questions about the genetic structure of populations of *D. pseudoobscura*:

How widespread is the genetic heterozygosity revealed in our original work? From the original 18 loci (HUBBY and LEWONTIN 1966) we have expanded our study to 24 and in place of our original small samples of a dozen or fewer strains, most having been cultured in the laboratory for many years (LEWONTIN and HUBBY 1966), we have studied between 76 and 220 genomes newly sampled from the wild, for each population. These more complete samples have confirmed the original estimates of LEWONTIN and HUBBY and amplified them slightly, as shown in Table 2. The mean proportion of polymorphic loci in the three North American populations is .42 and the average heterozygosity per individual is .123. These compare to .30 and .115, respectively, in our original paper (LEWONTIN and HUBBY 1966, Table 3). Thus, the larger samples of loci and genomes have increased by one third the estimate of the proportion of loci polymorphic but have increased only slightly the estimate of average heterozygosity per individual. We may take the figure of 12% heterozygosity, then, as a firm estimate within the limits of detection of our method. As we explained in our 1966 paper, electrophoretic differences may be apparent in as few as 25% of all amino acid substitutions.

A somewhat more conservative estimate of heterozygosity results from ignoring the loci on chromosome III which are intimately bound up with the inversion

TABLE 2

Proportion of loci polymorphic (total loci = 24) and the proportion of the genome estimated to be heterozygous in an average individual for each population studied

Population	Number of loci polymorphic	Proportion of loci polymorphic	Proportion of genome heterozygous per individual	
			Including *Pt-10* and *Amy-1*	*Pt-10* and *Amy-1* not included
Strawberry Canyon	11	.46	.140	.110
Mesa Verde	10	.42	.110	.102
Austin	9	.38	.117	.112
Bogota	6	.25	.044	.048
Mean excluding Bogota	10	.42	.123	.108
Grand Mean	9	.378	.103	.093

polymorphism on that chromosome (PRAKASH and LEWONTIN 1968). A significant reduction in the estimate of heterozygosity results in Strawberry Canyon (last column of Table 2) but the other populations are unaffected. For comparison with our earlier work, we have not included Bogota in the average heterozygosity and average polymorphism, since that population was specifically included in the present work as *a priori* atypical owing to its isolation from the rest of the species range. Even if it is included, however, average heterozygosity per individual is 10% and the average proportion of loci polymorphic is 38%.

An expected result of the larger sample of genomes in the present study is an increase in the number of alleles detected in each population. The most dramatic case is that of the *Est-5* locus which now has 12 alleles known from natural population, with 8–9 present in each population. Uncommon new alleles have also been detected at the loci for *Pt-7* (4 alleles) *Pt-8* (4 alleles) and *LAP* (5 alleles). Whereas $Pt\text{-}8^{.80}$ was originally thought to be an allele unique to Bogota it is now found to be a low frequency component of the North American populations.

In summary, the answer to the first question is that at least 38 to 42% of the loci are polymorphic and 10%–12% of the loci are in a heterozygous state for each individual, in populations from the main range of the species.

Are these differences in genetic variation between populations according to their location in the species range?

Table 2 shows remarkably little difference in the genetic heterogeneity within the three populations from the main species range. There is some suspicion of a decline in the proportion of loci polymorphic from the center of abundance at Strawberry Canyon to the extreme southeastern margin at Austin. Austin, at the edge of the Edwards Plateau, lies on a sharp boundary of the distribution of *D. pseudoobscura.* The loci monomorphic in Austin that are polymorphic in Strawberry Canyon, however, are *octanol dehydrogenase, acetaldehyde oxidase,* and *Pt-12,* all three of which have one allele in very high frequency even in populations in which they are polymorphic. Thus, the monomorphism of these loci in Austin does not make a significant difference in the genetic heterogeneity.

When we turn to heterozygosity per individual, Strawberry Canyon again seems to be more heterozygous than either Austin or Mesa Verde, but the last column in Table 2 shows that this excess heterozygosity is entirely contained in the third chromosome inversion polymorphism. If the two chromosome III loci are omitted, there are no differences at all among the three North American populations in average heterozygosity.

Bogota stands in contrast. As far as is known, no population of *D. pseudoobscura* is resident south and east of the Guatemalan highlands except for the isolated, but *large* population around Bogota, Colombia. This population is clearly depauperate in its variation with only .25 of its loci polymorphic and average heterozygosity per individual of less than 5%. Nor is the loss of heterogeneity restricted only to those loci that are nearly fixed in the northern populations. Bogota is completely homozygous for the 1.00 allele of *XDH*, while the other populations are all segregating for 5 alleles with an average heterozygosity at this locus of .55 (Table 1D). Even more striking is the virtual monomorphism of the Bogota population for the *Est-5* locus (Table 1E).

With the current data it cannot be proven that isolation from the rest of the species range is responsible for the relative homozygosity of Bogota. To make the point it will be necessary to examine other populations at the southern end of the distribution, specifically southern Mexico and Guatemala. Recent attempts to collect in these localities have been unsuccessful partly because of the great increase in human occupation in the last 30 years. Another approach is the examination of other isolated populations such as those from the Channel Islands off the California coast, and from the isolated mountain ranges in Death Valley, Nevada, and Arizona. These latter collections have been made and are in the process of analysis. The degree of isolation of these populations is, however, uncertain so the effect of isolation can be given only a one-sided test.

Whatever the effect of isolation, there is clearly no significant effect of marginality on genic heterogeneity. This is unlike the situation usually observed. MAYR (1963) points out that "A study of (polymorphic) species reveals almost invariably that the degree of polymorphism decreases toward the border of the species and that the peripheral populations are not infrequently monomorphic . . ." Moreover, CARSON (1959) has postulated that peripheral populations of a species are subject to "homoselection" producing specialized, more homozygous populations. *D. pseudoobscura* seems to be an exception, although MAYR is talking about morphological polymorphisms and inversion polymorphisms rather than the general variation of the genome. It may be, of course, that *D. pseudoobscura* simply does not have the appropriate breeding structure ("frequent isolation and low population density," according to MAYR) that both MAYR and CARSON consider as preconditions for marginal homozygosity. A study of genic variation in central and peripheral populations of *D. robusta* or *D. willistoni*, which show the classical picture for inversion polymorphisms, would be most illuminating.

What does the pattern of allelic variation within and between populations reveal about the forces governing the variation?

Variation at a locus *within* a population will be maintained by some form of balancing selection like heterosis, by selection in favor of one allele, the "wild type" opposed by mutation to aberrant alleles, or else by selection of different alleles in different populations with enough migration between populations to prevent local fixation. On the other hand isolation, small population size or strong selection in favor of one allele with no flow of alternate alleles from other populations will destroy intra-population variation. Variation *between* populations will result from local selection in different directions with insufficient migration between populations to swamp the effect of divergent selection, or else from small population size and virtually complete isolation between populations, producing random divergence. The effects of divergent selective forces and random divergence are counteracted by gene flow between populations, so that the observed variations in allelic frequencies among populations are the result of the balance. In assessing the strength of these forces in the case of *D. pseudoobscura* populations we must account for the following facts:

a) The high degree of genic polymorphism within the three North American populations, b) The remarkable overall similarity of the polymorphism from population to population in that the same alleles are in high and intermediate frequencies in all three North American populations, e.g., $Pt\text{-}8^{.81}$ and $Pt\text{-}8^{.83}$, $XDH^{.99}$ and $XDH^{1.00}$, $Est\text{-}5^{1.00}$, and $Est\text{-}5^{1.07}$. c) The repeated appearance at the same frequency of the same low frequency alleles in all populations, e.g., $Pt\text{-}7^{.73}$ and $Pt\text{-}7^{.77}$, $Pt\text{-}8^{.80}$, $LAP^{1.10}$, and $LAP^{.95}$, etc. d) The evidence of a cline in gene frequencies at the *Est-5* locus, of high frequency of allele *.80* in Bogota at the *Pt-8* locus and of a polymorphism unique to Bogota, *Pt-13*. All are evidences of some local differentiation.

Now it is not necessary that the same explanation apply to all loci since selection is certainly different for the different functions, but some explanations are ruled out by these observations and others are made unlikely. We will list the possible explanations of the genic variation, together with an analysis of each as it pertains to our observations.

1) Selection is uniformly for a "wild-type" allele at each locus with the observed variation being maintained by mutation. This explanation is clearly out of the question for all the loci we have designated as polymorphic, 46% of the loci in Strawberry Canyon, for example. It is obvious that a balance between selection for a wild type and mutation to disadvantageous alleles cannot explain the $Pt\text{-}8^{.81}$, $Pt\text{-}8^{.83}$ and the $XDH^{.99}$, $XDH^{1.00}$ polymorphism, since in all these cases we are dealing with alleles at intermediate frequencies. But even the lopsided polymorphisms at the *malic dehydrogenase* locus and the *octanol dehydrogenase* locus where the less frequent allele is present in 3% to 5% will not admit this explanation. In case of no dominance where selection operates against the rare allele the equilibrium gene frequency (q) is equal to $2u/s$ where u is the mutation rate to the rare gene and s the selection coefficient against the rare homozygote. To maintain an alternate allele at a frequency of .03, even if s were equal to .01 would require mutation rates of the order of 10^{-4}. There is no evi-

dence of such high mutation rates for these loci in Drosophila, and it appears that rates for these amino acid substitutions are comparable to rates for visible mutants (10^{-6}–10^{-5}). (unpublished experiments)

2) Selection is for a different "wild type" in different populations and gene flow between populations maintains the variation within populations. This explanation is ruled out by the lack of variation in gene frequency *between* populations. If we assume, as we must, that the differences in gene frequency between such widely separated localities as Austin, Mesa Verde and Strawberry Canyon are at least as great as the gene frequency differences between much closer localities, we can take the observed frequency variation as an upper bound for populations that are close enough to exchange genes directly. If m, the migration rate per generation, and s, the selection for or against an allele, are of the same order of magnitude, the absolute difference in allelic frequency between two populations selected in opposite directions will be very close to:

$$| q_1 - q_2 | = | 1 - \sqrt{1-\bar{q}} - \sqrt{\bar{q}} | \text{ (Li 1955)}$$

where $\bar{q}$ is the average frequency of the allele over both types of populations. For a .45:.55 polymorphism like *Pt-8* this difference is then about .41; for $XDH^{.99}$ it is about .37 and for a rare allele like $LAP^{1.10}$ it is .185. Reference to Table 1C, 1D, and 1B, respectively, shows that the largest observed differences between populations are much smaller than these calculations. Thus the migration rate must be much larger than the intensity of the selection pulling the gene frequencies apart. In the case where $m > s$ the difference in allelic frequencies between two populations is approximately

$$| q_1 - q_2 | = 2s/m\, \bar{q}(1-\bar{q})$$

Applying the data from Table 1 we get

$Pt\text{-}8^{.81}$ or $Pt\text{-}8^{.83}$	$\lvert q_1 - q_2 \rvert \cong .495\ s/m$
$XDH^{.99}$	$\lvert q_1 - q_2 \rvert \cong .391\ s/m$
$LAP^{1.10}$	$\lvert q_1 - q_2 \rvert \cong .072\ s/m$

Using the largest observed differences between North American populations from Table 1C, 1D and 1B for these alleles we get estimates of the ratio of s to m as follows:

$Pt\text{-}8^{.81}$: $s = .125\ m$
$XDH^{.99}$: $s = .173\ m$
$LAP^{1.10}$ $s = .347\ m$

So, for example, if we assume that different populations living in environments different enough to have selection favoring opposite wild types exchange 10% of these individuals each generation, the selection coefficients for the genes can be only .0125 for *Pt-8*, .017 for *XDH* and 0.035 for *LAP*. But 10% gene exchange per generation is a great deal for *D. pseudoobscura* unless the populations are within a few kilometers of each other. (Dobzhansky and Wright 1943)

Our conclusion must be that unless selection coefficients favoring opposite

"wild types" are quite small we cannot explain the observed lack of difference in gene frequency between populations.

3) The observed variation is truly "isoallelic" and selection coefficients, whatever their direction, are very small or even zero on the average. This is the hypothesis that the observed genic variation is physiologically irrelevant.

At first sight this is an attractive hypothesis. A large amount of genetic variation can be maintained in a population if it is assumed that many alternative states of each gene are possible and arise by spontaneous mutation. As shown by KIMURA and CROW (1964) the effective number of alleles that can be maintained in such a case is approximately

$$4Nu + 1$$

where N = effective population size and u is the mutation rate. Assuming effective population sizes of the order of the reciprocal of the mutation rate an effective number of alleles of 5 per population at a locus is quite reasonable. At this stage of our knowledge of the ecology of *D. pseudoobscura* it is impossible to say whether effective population sizes might in some cases be 10^5 or 10^6. What evidence we have (WRIGHT, DOBZHANSKY and HOVANITZ 1942; DOBZHANSKY and WRIGHT 1943) suggests effective sizes of the order of 10^3–10^4. However, a little in-migration goes a long way in increasing the effective size of the population.

The observation of the lower heterozygosity in the isolated populations of Bogota and of the occurrence of unusual alleles in high frequency in that population ($Pt\text{-}8^{.80}$, $Pt\text{-}13^{1.37}$) are in accord with the isoallelic hypothesis. It is precisely in a population cut off from the flow of genes from the rest of the species that homozygosity should increase and alleles should undergo random changes in frequency, unrelated to the frequencies in the other populations. The North American populations, on the other hand, are connected with each other by migration so their effective population sizes are very great. Moreover, they will tend to have the same alleles in the same frequency because of gene flow. In MAYR's terms, migration maintains the "integrity of the species" while the Bogota population, cut off by 1500 miles, has drifted away in its genetic composition.

The difficulty with the isoallelic hypothesis is that it cannot explain the *identity* of the allelic configurations in populations so widely separated as Austin, Mesa Verde, and Strawberry Canyon. Such loci as *MDH*, *Pt-7*, *Pt-8*, *LAP* and *XDH* have allelic frequencies that are identical within the limits of sampling error over all three populations. Most telling are the repeatable frequencies of the rarer alleles which, under the isoallelic model, would be subject to frequent loss from populations. The *Est-5* locus which, by number of alleles seems a perfect example of the isoallelic hypothesis, does not fit because of the apparent clinal nature of the allelic frequency differences and the fact that the clinal differences are superimposed on an underlying similarity of frequencies. $Est\text{-}5^{1.00}$ is the most frequent allele and $Est\text{-}5^{1.07}$ the next most frequent in all populations.

The isoallelic hypothesis also does not apply to the two loci on Chromosome III, *Pt-10* and *Amy* loci. We have already shown (PRAKASH and LEWONTIN 1968) that these loci are subject to strong selection pressures.

The only loci for which very weak selection might provide a satisfactory explanation are those like *MDH*, *ODH* and *acetaldehyde oxidase*, but these loci with one allele in very high frequency in some populations and fixed in others, are of no use in distinguishing any hypothesis since they are on the borderline of compatibility with all.

4) The variation is maintained by some form of balancing selection with moderate to strong selection intensities and with the direction of selection roughly the same in all localities.

This hypothesis is in accord with most of the observations and is compatible with all of them if it is assumed the Bogota population is of relatively recent origin. Both the high degree of polymorphism and the close similarity of allelic frequencies between populations is most easily explained by the assumption of heterosis or some other form of balancing selection dependent chiefly upon the inherent physiological properties of the polypeptides and only weakly on variations in the environment. The superimposition of a cline such as the one for the *Est-5* locus is not at variance with this general picture. The repeatability of the frequency of the less common alleles and their maintenance at frequencies between 3% and 10% is especially easy to explain under this balance hypothesis. Moreover, if the equilibrium frequency of an allele under balancing selection is on the average 3%, year to year variations in environment and accidents of sampling during the winter when population size is low can result in the temporary loss of an allele (*Est-5*$^{1.03}$ in Austin, *acetaldehyde oxidase*$^{.93}$ and *acetaldehyde oxidase* $^{1.02}$ in Mesa Verde and Austin, for example). Even in Bogota where variation is distinctly less than in North America, infrequent alleles at several loci are still maintained (*LAP*, *acetaldehyde oxidase-2*, *Est-5*, *Pt-7*). This is unlikely in an isolated population, especially if colonized by a very small number of individuals, without some form of balancing selection.

The depauperate variation of the Bogota population is of a special sort. With the single exception of *Pt-8*, the commonest or only allele in Bogota is the one in highest frequency in the North American populations. In some cases (*Pt-8*, *LAP*, *Est-5*) Bogota, while still polymorphic, has lost some of the less frequent alleles. These are exactly the results to be expected from a colonization by a small number of individuals. A small sample of genomes will consist either entirely of the most common allele or of that allele and one or at most two others. Infrequent alleles will be omitted and if the colonization has been recent, mutation will not have yet replaced the missing alleles. The *Pt-8* polymorphism in Bogota with a very high frequency of *Pt-8*$^{.80}$ can be explained by a shift in the selective value of different alleles because of drastic change in the genetic background which in turn was due to a "founder effect" and lack of gene flow from the main population.

In summary the hypothesis of widespread balancing selection at most of the polymorphic loci fits most easily all of the observations on the central, marginal and isolated populations, although we cannot exclude, for some loci, a model of selectively neutral isoalleles.

We are most grateful to Mr. C. V. Raman Murthy for technical assistance and to Chris Wills, Alan Wick, Louise Prakash and Michael Kambysellis for providing us with the

population samples. We are especially grateful to DR. ALICE HUNTER for the very great trouble she went to in getting the Bogota sample for us.

SUMMARY

Genetic variation at 24 randomly chosen loci of central (Strawberry Canyon, California), eastern marginal (Mesa Verde, Colorado), southeastern marginal (Austin, Texas) and isolated (Bogota, Columbia) populations of *D. pseudoobscura* has been studied. The central population has the largest amount of polymorphism and the isolate the least. Marginal populations are intermediate. *However, if we disregard the* Pt-10 *and* Amy *loci which are associated with the third chromosome gene arrangements, there is no difference in the amount of polymorphism in the central and the marginal populations.* On the average in the main populations 40% of the loci are polymorphic and the average proportion of heterozygous loci per individual is 12%.—In the North American populations we have found four kinds of polymorphic loci: (1) *Pt-7*, *Pt-8*, *LAP*, *XDH* and *MDH* loci reveal no differentiation between populations. (2) Various alleles of the *Est-5* locus show the presence of clines between these populations, (3) *acetaldehyde oxidase-2*, *Pt-12* and *AP-4* polymorphisms are restricted to a single population, (4) *Pt-10* and *Amy* polymorphisms, because of a pattern of association with the gene arrangements, have very different allelic frequencies in different populations.—The Bogota population which is isolated from the main body of the species by more than 1500 miles displays the greatest amount of genetic divergence and only about one-third as much genic variation as the main body of the species. The lowered variability and the large extent of genic divergence of this population are most likely due to recent colonization by a small number of founders.—From observations on the pattern of polymorphism in these four populations, we are forced to conclude that the most likely explanation for the maintenance of this genic variation is some form of balancing selection.

LITERATURE CITED

CARSON, H. L., 1959 Genetic conditions which promote or retard the formation of species. Cold Spring Harbor Symp. Quant. Biol. **24**: 87–105.

DOBZHANSKY, T., and C. EPLING, 1944 Contributions to the genetics, taxonomy and ecology of *Drosophila pseudoobscura* and its relatives. Carnegie Inst. Wash. Publ. 554, Washington, D.C.

DOBZHANSKY, T., and S. WRIGHT, 1943 Dispersion rate in *Drosophila pseudoobscura*. Genetics **28**: 304–340.

HUBBY, J. L., and R. C. LEWONTIN, 1966 A molecular approach to the study of genetic heterozygosity in natural populations. 1. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics **54**: 577–594.

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

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.

LI, C. C., 1955 *Population Genetics*. University of Chicago Press. Chicago and London.

MAYHEW, S. H., S. K. KATO, F. M. BALL, and C. EPLING, 1966 Comparative studies of arrangements within and between populations of *Drosophila pseudoobscura*. Evolution **28**: 646–662.

MAYR, E., 1963 *Animal species and Evolution*. The Belknap Press, Harvard University Press, Cambridge, Mass.

PRAKASH, S., and R. C. LEWONTIN, 1967 A molecular approach to the study of genic heterozygosity in natural populations III. Direct evidence of coadaptation in gene arrangements of Drosophila. Proc. Natl. Acad. Sci. U.S. **59**: 398–405.

STRICKBERGER, M. W., and C. J. WILLS, 1966 Monthly frequency changes of *Drosophila pseudoobscura* third chromosome gene arrangements in a California locality. Evolution **20**: 592–602.

WRIGHT, S., DOBZHANSKY, T., and HOVANITZ, 1942 The allelism of lethals in the third chromosome of *Drosophila pseudoobscura*. Genetics **27**: 363–394.