GENETICS OF NATURAL POPULATIONS. VII. THE ALLELISM OF LETHALS IN THE THIRD CHROMOSOME OF DROSOPHILA PSEUDOOBSCURA

SEWALL WRIGHT, TH. DOBZHANSKY, AND W. HOVANITZ[1]

University of Chicago, Chicago; Columbia University, New York; and California Institute of Technology, Pasadena

Received November 25, 1941

INTRODUCTION

HETEROZYGOSIS for recessive mutant genes, some of which are lethal or semilethal in homozygotes, is very common in natural populations of Drosophila. In *D. pseudoobscura*, from fifteen percent (in the Death Valley region of California and Nevada) to thirty percent (in Mexico and Guatemala) of the third chromosomes contain lethals or semilethals. DOBZHANSKY and WRIGHT (1941) have reported that the frequency of alleles among lethals found in the population of a given mountain range in the Death Valley region is greater than that among lethals derived from populations of different mountain ranges. The importance of this observation lies in the fact that it permits certain inferences to be drawn regarding the breeding structure of natural populations of *D. pseudoobscura*. Further data bearing on the same problem are given in the present article.

The work to be described in the following pages has been made possible by a grant from the CARNEGIE INSTITUTION of WASHINGTON. The authors wish to express their appreciation to that Institution, and also to MRS. MARIAN GIBBS-HOVANITZ, MR. K. J. MAMPELL, and MR. ALEXANDER SOKOLOFF who have helped in conducting the experiments. Acknowledgement is also made to the DR. WALLACE C. and CLARA A. ABBOTT MEMORIAL FUND of the UNIVERSITY of CHICAGO for assistance in connection with the calculations.

MATERIAL

Most of the samples of natural populations of *Drosophila pseudoobscura* which served as material for the present study were collected in 1939 in three localities on the slopes of Mount San Jacinto, California. These localities are (fig. 1): Keen Camp (this locality is shown on the U. S. topographic maps as Hemet Valley; it is referred to below as "Keen"), Piñon Flats (later referred to as "Pinon"), and Andreas Canyon ("Andreas").

Keen is an almond shaped valley with a marshy meadow in the center draining into a semi-artificial Hemet Lake. The meadow lies at an elevation of slightly more than 4300 feet above sea level; it is surrounded by

[1] Observational and experimental data by TH. DOBZHANSKY and W. HOVANITZ, mathematical analysis by SEWALL WRIGHT.

forest and brushland consisting chiefly of *Pinus ponderosa Jeffreyi, Quercus dumosa, Q. Kelloggii, Arctostaphylos glandulosa, Ceanothus cuneatus, Rhamnus californica, Adenostema sparsifolium, A. fasciculatum*, and *Artemisia tridentata*. The winters are long (freezing temperatures may be expected from November through April) and the rainfall relatively plentiful (normally 15 to 20 inches). The elevation of Pinon is approximately 4000 feet. Owing to a difference in soil and to Pinon being exposed to the hot and dry

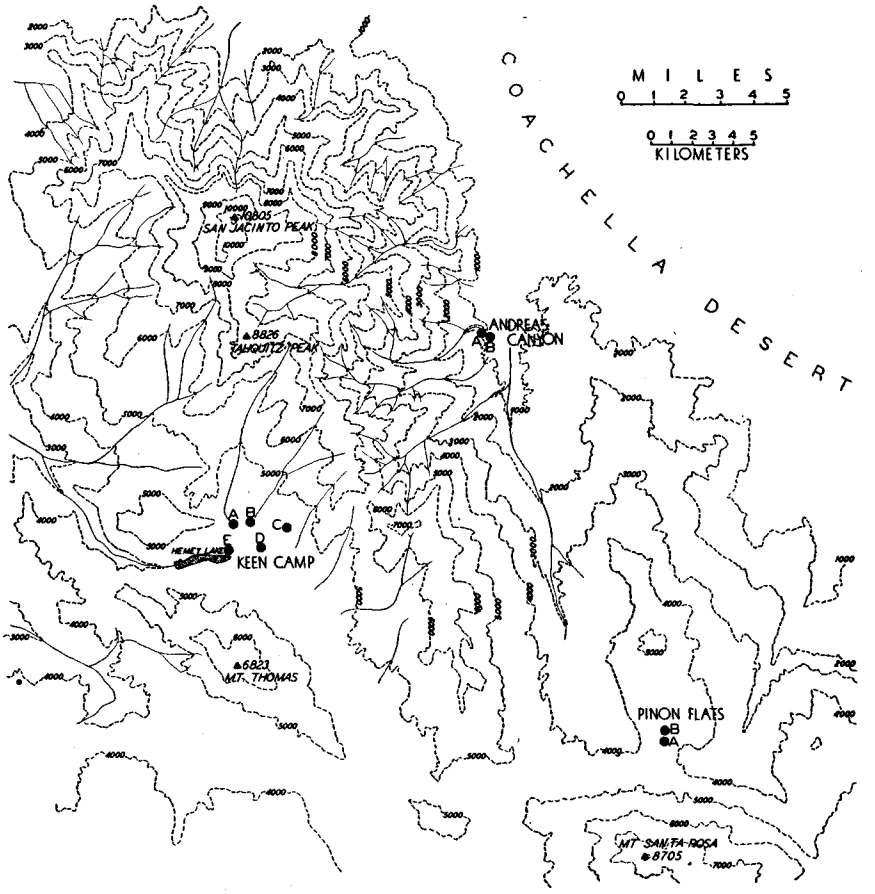

FIGURE 1.—Map showing the location of the collecting stations on Mount San Jacinto.

winds from the Coachella and Imperial deserts, the ecological conditions are very different at Pinon and at Keen. The larger and most characteristic plants at Pinon are *Pinus cembroides monophylla, Juniperus californica, Arctostaphylos glauca, Condalia Parryi, Bernardia myricaefolia, Rhus trilobata, R. ovata, Purshia glandulosa* and *Quercus dumosa*. Andreas is a narrow canyon with a permanent stream rising in the upper reaches of San Jacinto and disappearing on the desert not far from the collecting locality, which

lies at an elevation of about 800 feet. The characteristic flora of the canyon bottom is *Populus Fremonti, Alnus rhombifolia, Washingtonia filifera, Platanus racemosa, Prunus eriogyna, Salix* sp., *Pluchea sericea, Prosopis chilensis, Acacia Greggii, Vitis Girdiana*, and others.[2]

Pinon is about 13 miles air line distant from either Keen and Andreas, while the distance between Andreas and Keen is about ten miles. Almost the entire territory shown in figure 1 which lies at 3500 or more feet above sea level is covered with trees or shrubs and therefore supports large and permanent populations of *D. pseudoobscura*. The ribbon of trees along the bottom of Andreas Canyon continues upward without major interruptions and eventually joins the mountain forest of San Jacinto. All the collecting localities described above consequently lie within a continuously populated region.

Nine collecting stations were chosen in the three localities; they are designated by letters following the locality name. Each "station" is a territory of at most 100 meters in diameter, on which 15 collecting bottles with fermenting banana were placed on previously marked trees or bushes. At Keen B, C, D, and E, and Pinon A and B the bottles were arranged more or less on the periphery of a circle. At Keen A and Andreas A and B owing to the local topography, the bottles were arranged in an irregular line from 50 (Andreas) to 100 (Keen A) meters long. The distances between the stations within a locality vary from about 200 meters (between Andreas A and B) to about 3.5 kilometers (between Keen C and E). Owing to the local peculiarities of the terrain, it is very difficult to estimate the areas of our "stations" and "localities"—that is, the areas for which our collections may be considered representative samples. The attractive radius of a banana trap bottle is also unknown, although certain unpublished experiments conducted in the summer of 1941 indicate that it cannot be much larger than 50 meters. As very rough approximations, we believe that the "stations" at Keen and Pinon are about 10,000 square meters, and at Andreas 5000 square meters or less. The area of the Keen "locality" is probably in the neighborhood of six square kilometers, that of the Pinon locality about 0.3 square kilometers, and that of the Andreas locality less than 50,000 square meters.

The trap bottles were exposed for a few hours in the late afternoon and at sunset, or else during the early morning hours. Occasionally, especially in winter at Andreas Canyon, flies may be procured during the day. In making repeated samplings of the same collecting station, the bottles were exposed, of course, always on the same trees.

A few of the flies included in the material of the present study came not

[2] The writers are indebted to DRS. H. BONNER, J. BONNER, and C. C. EPLING for the above lists of the characteristic plant species.

from Mount San Jacinto but from Wildrose Canyon, Panamint Mountains, California. Panamint is about 200 miles distant from San Jacinto and separated from the latter by a wide desert area. The population of Wildrose was sampled in 1937 (DOBZHANSKY and QUEAL 1938a) and in 1938 (KOLLER 1939). Two collecting stations, Wildrose A and B, were established there in 1939. Wildrose A corresponds almost exactly to the territory from which KOLLER'S sample was derived. DOBZHANSKY'S 1937 sample came from a larger territory, embracing Wildrose A, Wildrose B, as well as a distance of about 200 meters intervening between the two.

THE SEASONAL CYCLES IN POPULATIONS OF DROSOPHILA PSEUDOOBSCURA

Populations of *Drosophila pseudoobscura* behave differently in the several localities described above. At Keen, *D. pseudoobscura* is by far the commonest species of Drosophila, and its population at the height of the season must be very large. The flies appear early in April, become very abundant in June and early July, decline in numbers during July and August (especially at Keen C, which is the driest of the Keen collecting stations), remain at about the same level in September, and ostensibly disappear with the advent of cold weather in October or November. Although nothing is known about the refuges in which the flies spend the winters, the hibernating population at Keen is probably rather large. At Pinon, *D. pseudoobscura* is the commonest species of the genus, followed by *D. simulans* and *D. hydei*. It is abundant from March to May, rare during the hot and dry period of the summer, and again common from September to October or November (depending upon the occurrence of rain at that time). It is not in evidence during the winter months, although it seems possible that some individuals may be active on exceptionally warm days. At Andreas, *D. pseudoobscura*, *D. simulans*, *D. hydei*, and *D. victoria* become the most frequent species at different times of the year. *D. pseudoobscura* reaches its maximum abundance from January to March, declines thereafter, disappears almost completely from June to August, and gradually builds up its numbers from September to January. It is certain that its population is not destroyed entirely during the hot part of the summer, since single individuals may occasionally be trapped even in mid July. However, the population density must be very low in summer. The cycle at Wildrose is insufficiently known and is probably variable from year to year in this region of erratic rainfall. In two years the population was very dense in early June but scarce in September. Owing to the high elevation (close to 7200 feet), the winters are long and severe; how large is the hibernating population may only be conjectured.

DETECTION OF LETHALS AND SEMILETHALS IN THE THIRD CHROMOSOME

To detect the recessive mutant genes in the third chromosome, wild males or single sons of wild females are crossed individually to females homozygous for orange and purple (*or pr*). Single F_1 males are crossed to females carrying orange, Blade, Scute, and purple in one of their third chromosomes and an inversion and a lethal in the other (*or Bl Sc pr/C l*). In the next generation, flies showing *Bl* and *Sc* but not *or* and *pr* are selected and intercrossed; among the progeny, 33.3 percent of the zygotes are expected to be homozygous for the wild third chromosomes. The details of this technique have been described by DOBZHANSKY and QUEAL (1938b), and only certain features need be mentioned here. Provided a wild third chromosome carries no genetic factors affecting the viability, the progeny of the *Bl Sc* flies will consist of 33.3 percent wild type and 66.7 percent *Bl Sc* individuals. If the wild third chromosome carries a lethal, no wild types appear. Genes reducing the viability cause the frequency of wild types to fall between 0 and 33.3 percent. Such genes are very common in natural populations (DOBZHANSKY and QUEAL 1938b, DOBZHANSKY 1939). Those which give more than 0 but less than 16.5 percent of wild type flies in the cultures are arbitrarily designated semilethals. However, since cultures with from 10 to 20 percent of wild types are relatively rare, this arbitrariness introduces no serious error in our estimates of the frequencies of semilethals in nature.

Since the purpose of the present experiments has been to detect only the lethals and semilethals, but not the minor viability modifiers, no complete counts were made in a majority of the cultures. Cultures were examined when at least 50 to 100 flies were hatched in each of them; if a quarter or more of the flies were wild type, the culture was considered free of lethals or semilethals. Cultures in which few or no wild types were present on the first count were preserved, and further hatches were counted to determine whether the original deviation was due to the presence of lethals or semilethals, or to chance. We are confident that where the wild third chromosomes had gene arrangements other than the Standard, no lethals or semilethals escaped detection (these gene arrangements are mostly Arrowhead and Chiricahua I, occasionally Tree Line and Santa Cruz; for description of them see DOBZHANSKY and STURTEVANT 1938). The fact that some of the wild chromosomes in the populations studied had the Standard arrangement introduces a complication. Since the gene arrangement in the *or Bl Sc pr* chromosome is Standard, crossing over is free to take place in the *or Bl Sc pr/* Standard wild females. A semilethal which lies in the chromosome far from the loci of the dominant marking genes *Blade* and *Scute* may be missed. Furthermore, in the chromosomes with the Standard gene arrangement the lethals are distinguishable from the semilethals chiefly

TABLE 1

Numbers of the third chromosome lethals and semilethals in samples of natural populations of Drosophila pseudoobscura. The figures in parentheses indicate the numbers of the chromosomes having the Standard gene arrangement.

COLLECTING STATION	DATE OF COLLECTING	LETHALS	SEMI-LETHALS	CHROMOSOMES EXAMINED
Keen A	April 25	1 (0)	—	7 (1)
	May 15	3 (2)	—	21 (8)
	June 20–21	9 (1)	5 (1)	67 (7)
	July 25–26	5 (2)	4 (2)	54 (18)
	Aug. 31–Sept. 1	1 (0)	—	14 (0)
	October 22	1 (0)	—	6 (1)
	Total	20 (5)	9 (3)	169 (35)
Keen B	April 25	—	1 (1)	6 (2)
	May 15	—	—	4 (2)
	May 20	3 (0)	—	24 (6)
	June 20–21	5 (0)	1 (0)	28 (4)
	June 29–30	2 (1)	1 (0)	21 (5)
	July 26–27	4 (1)	—	38 (8)
	Aug. 31–Sept. 1	—	1 (0)	18 (2)
	Total	14 (2)	4 (1)	139 (29)
Keen C	May 20	4 (2)	—	18 (4)
	June 2	8 (2)	—	37 (4)
	June 29–30	3 (0)	2 (0)	43 (2)
	July 25–26	1 (0)	—	15 (3)
	Sept. 1–2	1 (0)	—	7 (1)
	Total	17 (4)	2 (0)	120 (14)
Keen D	May 21	—	—	4 (0)
	June 3	6 (2)	1 (0)	62 (10)
	July 14–15	10 (2)	3 (0)	53 (14)
	Sept. 1–2	6 (0)	1 (0)	30 (5)
	Total	22 (4)	5 (0)	149 (29)
Keen E	May 22	1 (0)	—	3 (1)
	June 3	4 (2)	—	61 (15)
	July 14–15	3 (0)	2 (0)	40 (8)
	Aug. 18–19	1 (0)	—	26 (6)
	Total	9 (2)	2 (0)	130 (30)
Keen	Grand Total	82 (17)	22 (4)	707 (137)

COLLECTING STATION	DATE OF COLLECTING	LETHALS	SEMI-LETHALS	CHROMOSOMES EXAMINED
Andreas A	April 24	1 (0)	—	54 (28)
	May 13	1 (0)	1 (0)	17 (9)
	June 4	2 (0)	1 (0)	5 (0)
	Sept. 21	10 (4)	1 (1)	73 (37)
	Oct. 28–29	7 (1)	2 (0)	59 (21)
	Dec. 9	5 (4)	—	42 (22)
	March 2	8 (4)	2 (1)	93 (43)
	Total	34 (13)	7 (2)	343 (160)
Andreas B	April 24	—	1 (1)	19 (9)
	May 13	1 (1)	—	20 (10)
	June 4	3 (0)	—	19 (4)
	Sept. 21	9 (4)	3 (0)	103 (45)
	Oct. 28–29	9 (3)	—	47 (13)
	Dec. 9	3 (1)	—	23 (10)
	March 2	7 (3)	4 (0)	86 (40)
	Total	32 (12)	8 (1)	317 (131)
Andreas	Grand Total	66 (25)	15 (3)	660 (291)
Pinon A	April 23	—	1 (0)	13 (6)
	May 14	12 (2)	1 (0)	81 (19)
	June 21–22	3 (1)	—	23 (9)
	August 19–20	3 (0)	2 (0)	35 (7)
	Sept. 19–21	6 (2)	2 (1)	72 (25)
	Oct. 21–22	10 (6)	1 (1)	72 (29)
	Total	34 (11)	7 (2)	296 (95)
Pinon B	April 23	—	—	6 (6)
	May 14	6 (1)	3 (1)	64 (16)
	June 21–22	1 (0)	—	7 (2)
	August 19–20	1 (0)	—	28 (9)
	Sept. 19–21	5 (1)	2 (1)	57 (21)
	Oct. 21–22	1 (1)	1 (0)	69 (28)
	Total	14 (3)	6 (2)	231 (82)*
Pinon	Grand Total	48 (14)	13 (4)	527 (174)
San Jacinto	Grand Total	196 (56)	50 (11)	1894 (602)
Wildrose A	June 10	9 (2)	6 (0)	102 (23)
Wildrose B	June 10	14 (3)	5 (0)	94 (23)
Wildrose	Total	23 (5)	11 (0)	196 (46)

* The calculations are based on 79 as the number of Standard chromosomes at Pinon B. The original data were not accessible to the authors when the discrepancy was noted in correction of proof. Whatever its location, the effect on the conclusions is negligible.

if the latter make the surviving homozygotes phenotypically abnormal (which is frequently, but not invariably, the case). If the homozygotes are normal, a semilethal in a Standard chromosome is either recorded as a complete lethal or not detected at all.

THE FREQUENCIES OF LETHALS IN WILD CHROMOSOMES

The data are summarized in table 1. In view of the technical complications in the case of the Standard gene arrangement, it is desirable to compare the outcomes of the observations on the Standard and the Non-Standard chromosomes. The last two columns of table 2 show the frequencies and their standard errors of the Standard gene arrangement in various populations. There are highly significant differences in this respect among the regions considered here: San Jacinto 31.8 ± 1.1 percent, Death Valley 18.0 ± 1.2 percent, and it may be added, Mexico and Guatemala 0 percent (DOBZHANSKY 1939). There are, indeed, highly significant differences among the localities on San Jacinto: Keen 19.4 ± 1.5 percent, Pinon 33.0 ± 2.0 percent, and Andreas 44.1 ± 1.9 percent ($n = 2$, $\chi^2 = 99.6$, probability of such differences arising from accidents of sampling negligible). The significance of differences of this type between populations of neighboring localities will be discussed in another publication.

TABLE 2

The frequencies of lethals and semi-lethals.

	NON-STANDARD							STANDARD							% STANDARD	
	NO. CHROM.	LETHALS NO.	%	SE	SEMILETHALS NO.	%	SE	NO. CHROM.	LETHALS NO.	%	SE	SEMILETHALS NO.	%	SE	SE	
San Jacinto																
Keen	570	65	11.4	1.3	18	3.2	0.7	137	17	12.4	2.8	4	2.9	1.5	19.4	1.5
Pinon	353	34	9.6	1.6	9	2.5	0.8	174	14	8.0	2.1	4	2.3	1.1	33.0	2.0
Andreas	369	41	11.1	1.6	12	3.3	0.9	291	25	8.6	1.6	3	1.0	0.6	44.1	1.9
Total	1292	140	10.84	0.86	39	3.02	0.48	602	56	9.3	1.18	11	1.83	0.55	31.8	1.1
Death Valley																
1937	707	91	12.9	1.3	26	3.7	0.7	142	10	7.0	2.1	0	0		16.7	1.3
Wildrose	150	18	12.0	2.7	11	7.3	2.1	46	5	10.9	4.5	0	0		23.5	3.0
Total	857	109	12.72	1.14	37	4.32	0.69	188	15	7.98	1.98	0	0		18.0	1.2
Grand Total	2149	249	11.59	0.69	76	3.54	0.40	790	71	8.99	1.02	11	1.39	0.42		

As shown in table 2, the frequencies of lethals are higher in Non-Standard than in Standard chromosomes, both at San Jacinto and in Death Valley, although not significantly so. The difference may be considered as probably significant in the grand total ($P = 0.05$). The deficiency of semilethals in Standard chromosomes, on the other hand, is clearly significant in the data from Death Valley ($P = 0.004$) and in the grand total ($P = 0.002$). This deficiency is due probably to the above mentioned difficulty

in the detection of semilethals in Standard chromosomes. Table 3 shows the relative frequencies of semilethals and of complete lethals among the mutants detected in natural populations and among the mutants obtained in the laboratory. In Non-Standard chromosomes, the percentages of semilethals among all lethals do not differ significantly among the three regions considered, the average being 24.1 ± 2.3 percent ($n=2$, $\chi^2=1.4$, $P=0.5$). This average, moreover, does not differ significantly from the corresponding figure in laboratory lethal mutations, 21.5 ± 5.1 percent ($P=0.7$). Since semilethals as a class may be less strongly selected against in heterozygotes than complete lethals, this similarity of the percentage in nature and in the laboratory contributes somewhat to the evidence that there has been no important differential elimination of lethal mutations in nature.

TABLE 3

Percentage of semilethals among the total lethals found in nature or occurring in the laboratory. The Standard and Non-Standard chromosomes are distinguished in the wild populations. The laboratory mutants include only Non-Standard chromosomes.

	NON-STANDARD				STANDARD			
	COMPLETE AND SEMI-LETHALS	SEMILETHALS			COMPLETE AND SEMI-LETHALS	SEMILETHALS		
		NO.	%	SE		NO.	%	SE
Death Valley	146	37	25.4	3.6	15	0	0	—
San Jacinto	179	39	21.8	3.1	67	11	16.4	4.5
South of U. S.	36	11	30.6	7.7	0	—	—	—
Total (in nature)	361	87	24.1	2.3	82	11	13.4	3.8
Laboratory mutants	65	14	21.5	5.1				

In the Standard chromosomes, relatively fewer semilethals are recorded (13.4 ± 3.8 percent) than in Non-Standard chromosomes (24.1 ± 2.3 percent as noted above). The difference appears to be significant ($P=0.03$).

It is obvious that accurate comparisons cannot be made in data that include semilethals in Standard chromosomes. It would be fairly satisfactory to use the data for all chromosomes but to consider only complete lethals. However, the best estimates of lethal frequencies are probably those based on Non-Standard chromosomes alone but including semilethals as well as lethals. Henceforward in the present paper the expression "lethal" will be used to include the semilethals, "complete lethal" being used where none of the recessives survives.

Table 4 shows the frequency of lethals by region and locality in Non-Standard chromosomes and in all chromosomes. Confining attention to Non-Standard, there are clearly significant differences in lethal frequency

between either San Jacinto (13.9 ± 1.0 percent) or Death Valley (17.0 ± 1.3 percent) and the region south of the United States (30.0 ± 4.2 percent). The slightly higher frequency in Death Valley compared with San Jacinto is probably significant (P = 0.02). There are no significant differences, however, among the localities of the same region (for example, in San

TABLE 4

The frequency of lethals, including semilethals, by region and locality. The frequencies in Non-Standard chromosomes are given in addition to those in all chromosomes because of the greater reliability of Non-Standard than Standard.

	NON-STANDARD				TOTAL			
	CHROMOSOMES NO.	LETHAL	% LETHAL	SE	CHROMOSOMES NO.	LETHAL	% LETHAL	SE
San Jacinto								
Keen	570	83	14.6	1.5	707	104	14.7	1.3
Pinon	353	43	12.2	1.7	527	61	11.6	1.4
Andreas	369	53	14.4	1.8	660	81	12.3	1.3
Total	1292	179	13.85	0.97	1894	246	12.99	0.77
Death Valley								
1937	707	117	16.5	1.4	849	127	15.0	1.2
Wildrose	150	29	19.3	3.2	196	34	17.3	2.7
Total	857	146	17.04	1.28	1045	161	15.41	1.12
South of U. S.								
Mexico	82	23	28.0	5.0				
Guatemala	38	13	34.2	7.7				
Total	120	36	30.0	4.2				

TABLE 5

The frequencies of lethals, including semilethals and of Standard, in all tested chromosomes at the various stations and localities of San Jacinto and of Wildrose (1939).

	NO. TESTED	% LETHAL	% STAND.		NO. TESTED	% LETHAL	% STAND.
Keen A	169	17.2	20.7	Andreas A	343	12.0	46.6
B	139	12.9	20.9	B	317	12.6	41.3
C	120	15.8	11.7				
D	149	18.1	19.5	Total	660	12.3	44.1
E	130	8.5	23.1				
Total	707	14.7	19.4				
Pinon A	296	13.9	32.1	Wildrose A	102	14.7	22.5
B	231	8.7	34.2	B	94	20.2	24.5
Total	527	11.6	33.0	Total	196	17.3	23.5

Jacinto, $n = 2$, $\chi^2 = 1.1$, $P = 0.6$) in marked contrast with striking differences in the frequencies of various gene arrangements in the third chromosomes in those same localities.

Tables 5 and 6 break down the data from localities according to station and date of collection, respectively. For the purpose here, the slight inaccuracy brought by inclusion of the Standard chromosomes is probably more than compensated for by the increase in numbers by such inclusion. The data accordingly refer to all chromosomes.

The frequency of the Standard chromosome shows no significant differences among the stations of the same locality (total, $n = 7$, $\chi^2 = 8.3$, $P = 0.3$). There are the marked differences, however, between localities already

TABLE 6

The frequencies of lethals, including semilethals, and of the Standard arrangement, according to month of collection at three localities of San Jacinto.

	KEEN			PINON			ANDREAS		
	NO. TESTED	% LETHAL	% STAND.	NO. TESTED	% LETHAL	% STAND.	NO. TESTED	% LETHAL	% STAND.
April 1939	13	14.9	27.6	19	14.0	28.6	73	2.7	50.7
May 1939	74			145			37	14.8	37.7
June 1939	319	14.7	14.7	30	13.3	36.7	24		
July 1939	200	16.0	25.5						
Aug. 1939	95	11.9	14.9	63	9.5	25.4			
Sept. 1939	–			129	11.6	35.7	176	13.1	46.6
Oct. 1939	6			141	9.2	40.4	106	17.0	32.1
Dec. 1939							65	12.3	49.2
Mar. 1940							179	11.7	46.4
Total	707	14.7	19.4	527	11.6	33.0	660	12.3	44.1

referred to above. The possibility that this may be related to differences in conditions is suggested by significant differences in the frequency of Standard according to date of collection in the same locality (total, $n = 12$, $\chi^2 = 30.8$, $P = 0.003$). This is brought out more adequately by other more extensive data which will be presented elsewhere.

With respect to the frequency of lethals, there are not only no significant differences among stations of the same locality (total, $n = 7$, $\chi^2 = 11.2$, $P = 0.1$) or among localities as previously discussed but also none according to date of collection (total, $n = 12$, $\chi^2 = 11.9$, $P = 0.5$). Thus the percentage of lethals as a group appears to be rather stable within all subdivisions of a region, irrespective of very great diversity of conditions, and at all times of the year. That there is some factor, however, that can make a difference in the percentage of lethals is shown by the probably significant differ-

ence between Death Valley and San Jacinto and the highly significant difference between both of these and the Mexico-Guatemala region discussed above.

We shall use 13.85 ± 0.97 percent based on the Non-Standard chromosomes as the most reliable estimate of the frequency of lethal bearing third chromosomes at San Jacinto. The corresponding figure for the 1937 collections at Death Valley, 16.5 ± 1.4 percent, is probably preferable for the region in 1937 to the figure based on all chromosomes (15.0 ± 1.2 percent) which was used in the previous paper (DOBZHANSKY and WRIGHT 1941), but the difference is not great. The former figure will be used in this paper as a basis for comparative analysis.

ALLELISM OF LETHALS

In an indefinitely large random breeding population, there would be no appreciable chance that two lethals taken at random have had a common origin. However, they may prove to be alleles in a certain proportion of cases merely because of recurrent mutation in a finite number of loci. In populations of limited size, this basic chance of allelism is supplemented by that due to the chance of common origin. As brought out in the preceding paper, comparison of the frequencies of allelism of lethals from the same or different populations is among the most important sources of information on the breeding structure of the species.

The technique of testing the allelism of third-chromosome lethals and semilethals is, briefly, as follows. Lethal-bearing third chromosomes are perpetuated in the form of balanced stocks of the constitution: lethal/*or Bl Sc pr*. If stocks containing non-allelic lethals are intercrossed, the progeny consists of *Bl Sc* and wild-type flies in a ratio approaching 2:1; if the lethals are alleles, the progeny consists of *Bl Sc* flies only. For details consult DOBZHANSKY and WRIGHT 1941.

To test the allelism of a group of lethals, every lethal-bearing strain must be outcrossed to all other strains of the group; with k lethals this requires $k(k-1)/2$ crosses. Since as many as 246 lethals and semilethals have been recovered from the San Jacinto populations alone (table 1), the labor of testing them all is prohibitive. In view of this, we have performed all the possible intercrosses of the lethals and semilethals found in the samples from the same stations, while only a part of the intercrosses of lethals recovered from different stations and localities have been made.

The basic chance of allelism, due merely to recurrent mutation, may be determined by testing lethals from such remote populations that there is no appreciable chance of common origin.

The results of intercrosses of lethals from different localities on Mount San Jacinto and of lethals from Mount San Jacinto and Wildrose (Death

Valley region) are shown in table 7. The numbers of the lethals examined are shown in parentheses after the name of locality; thus, "Keen (15) ×Pinon (15)" signifies that fifteen lethals from Keen were intercrossed with fifteen lethals from Pinon; the number of the intercrosses necessary to test this group of lethals is evidently $15 \times 15 = 225$.

TABLE 7

Numbers of intercrosses made and numbers of alleles found among lethals from populations of different localities and regions.

LOCALITIES	NUMBER OF INTERCROSSES	NUMBER OF ALLELES
Keen (15)×Pinon (15)	225	1
Keen (15)×Andreas (15)	225	1
Pinon (16)×Andreas (16)	256	2
Keen (15)×Wildrose (15)	225	0
Pinon (15)×Wildrose (15)	225	1
Andreas (15)×Wildrose (15)	225	1
Total	1381	6

These data may be supplemented by the data published previously on the frequency of allelism of lethals from different mountain forests of Death Valley. Table 8 below gives a condensed summary of all the data bearing on this question.

TABLE 8

Frequency of allelism in tests of lethals, including semilethals, from different localities either of the same or different regions.

REGION	NO. TESTS	SAME LOCUS	%	SE
San Jacinto	706	4	0.57	0.28
Death Valley	4913	20	0.41	0.09
S.J. and D.V.	675	2	0.30	0.21
Total	6294	26	0.413	0.081

Taking the figures at face value, the percentage of allelism decreases with distance, but the differences are not significant ($n = 2$, $\chi^2 = 0.65$, $P = 0.7$). The grand average, 0.413 ± 0.081 percent will be used as an estimate of the chance of allelism at localities so remote that independent origin may be assumed.

At the opposite extreme is the chance of allelism within a station. The data for the intra-station crosses are presented in table 9. As explained

above, the number of crosses necessary to test the allelism, for example, of the 27 lethals of Keen A is 351. Table 1 however, shows that the 29 lethals, recovered from the population of this station (two of which were lost before the allelism tests were made) came from six samples taken at approximately monthly intervals from April to October 1939. Some of the 351 crosses consequently were between lethals collected simultaneously and

TABLE 9

Numbers of intercrosses made and numbers of alleles found among lethals from the populations of the same station.

STATION	LETHALS EXAMINED	COLLECTED SIMULTANEOUSLY		COLLECTED AT DIFFERENT TIMES		TOTAL	
		INTERCROSSES	ALLELES	INTERCROSSES	ALLELES	INTERCROSSES	ALLELES
Keen A	27	109	—	242	4	351	4
Keen B	17	25	1	111	2	136	3
Keen C	17	37	1	99	6	136	7
Keen D	25	108	2	192	7	300	9
Keen E	11	16	—	39	1	55	1
Pinon A	34	119	6	442	4	561	10
Pinon B	16	38	1	82	2	120	3
Andreas A	19	52	1	119	1	171	2
Andreas B (1939)	21	62	1	148	2	210	3
Andreas B (March 2, 1940)	8	28	2	—	—	28	2
Wildrose A (June 10, 1939)	15	105	—	—	—	105	—
Wildrose B (June 10, 1939)	13	78	1	—	—	78	1
Total	223	777	16	1474	29	2251	45

others between lethals from samples collected at different times. Since it is desirable to compare the frequency of alleles among lethals from the same sample with that among lethals collected at different times, the data in table 9 are differentiated accordingly.

Table 10 summarizes the results of the intercrosses between lethals found in the populations of different stations within a single locality. Since the population of every station (except Wildrose) has been sampled repeatedly during 1939, we may distinguish the intercrosses involving lethals detected in samples taken on the same day or within a short time interval (less than a month), and the intercrosses of lethals derived from samples taken at times from a month to several months apart. The former are shown in tables 9 and 10 in the columns labelled "Collected simultaneously," and the latter in the columns labelled "Collected at different times."

Table 11 gives a condensed summary of the data from San Jacinto according to whether the flies came from the same station or different sta-

TABLE 10

Numbers of intercrosses made and numbers of alleles found among lethals from populations of different stations within a locality.

STATIONS	COLLECTED SIMULTANEOUSLY		COLLECTED AT DIFFERENT TIMES	
	NUMBER OF INTERCROSSES	NUMBER OF ALLELES	NUMBER OF INTERCROSSES	NUMBER OF ALLELES
Keen A×Keen B	112	2	180	0
Keen A×Keen C	21	0	156	3
Keen A×Keen D	96	2	178	2
Keen A×Keen E	3	0	68	0
Keen B×Keen C	36	1	78	0
Keen B×Keen D	3	0	89	0
Keen B×Keen E	3	0	9	0
Keen C×Keen D	56	1	86	0
Keen C×Keen E	36	0	22	0
Keen D×Keen E	100	0	37	0
Pinon A×Pinon B	126	2	398	5
Andreas A×Andreas B	99	1	292	1
Wildrose A×Wildrose B	230	2	—	—
Total	921	11	1593	11

tions of the same locality and according to whether they were collected simultaneously or not.

In the earlier study of data from Death Valley, the tests for allelism of lethals from the same locality were based wholly on flies collected simul-

TABLE 11

Frequency of allelism of lethals (including semilethals) according to whether the flies were collected at the same or different stations of the same locality and according to whether collected simultaneously or at different times (San Jacinto only).

	SAME STATION				DIFFERENT STATIONS			
	NO. TESTS	SAME LOCUS	%	SE	NO. TESTS	SAME LOCUS	%	SE
Simultaneous	594	15	2.53	0.64	691	9	1.30	0.43
Different times	1474	29	1.97	0.36	1593	11	0.69	0.21
Total	2068	44	2.13	0.32	2284	20	0.88	0.20

taneously from a few traps. It was pointed out that there was a possibility that the proportion of alleles might be higher than among random flies from

the locality (DOBZHANSKY and WRIGHT 1941). In the present data, the percentage of alleles in tests of flies caught simultaneously (2.53 ± 0.64 percent is indeed slightly higher than in the tests of flies caught at the same station several months apart (1.97 ± 0.36 percent), but the difference is not significant ($P = 0.4$). The difference is in the same direction in the case of flies caught in different stations (simultaneously 1.30 ± 0.43 percent alleles, non simultaneously 0.69 ± 0.21 percent alleles), but again the difference is not significant ($P = 0.2$).

What is really wanted is the chance of allelism in a random sample of a single generation. This might be expected to be a little less than that from flies caught in a few traps but a little larger than that from flies caught over a period in which the frequencies of lethals have had an opportunity to change. In view of recent studies indicating very rapid diffusion of flies

TABLE 12

Frequency of allelism of lethals, including semilethals, according to locality and according to whether collected at the same or different stations of the same locality (same or different locality of the region is the case of Death Valley 1937).

LOCALITY	SAME STATION				DIFFERENT STATIONS			
	NO. TESTS	SAME LOCUS	%	SE	NO. TESTS	SAME LOCUS	%	SE
San Jacinto								
Keen	978	24	2.45	0.49	1369	11	0.80	0.24
Pinon	681	13	1.91	0.52	524	7	1.34	0.50
Andreas	409	7	1.71	0.64	391	2	0.51	0.36
Total	2068	44	2.13	0.32	2284	20	0.88	0.20
Death Valley 1937	772	24	3.11	0.63				
Wildrose 1939	183	1	0.55		230	2	0.87	0.67

from a point of release (unpublished), it has become doubtful whether the former consideration has much weight. Thus it is probable that the grand averages for stations and localities tend to be slightly too low rather than too high. This is especially the case with the figure for localities, which should, ideally, be based on allelism tests of a large number of lethals collected at random throughout the territory designated as "locality." In our computations the allelic lethals found within stations are entirely excluded from consideration as far as the frequency of allelism within the localities is concerned, although they should have some weight in completely random tests of a locality.

Table 12 gives a condensed summary of the data according to locality. The data from the 1937 collections at Death Valley are included under the

caption "same station," but the flies were taken from areas comparable to the Andreas and Pinon localities of San Jacinto.

There are no significant differences between the localities at San Jacinto in spite of the great differences in conditions (for intra-station tests $n=2$, $\chi^2=1.0$, $P=0.6$; for inter-station tests $n=2$, $\chi^2=1.9$, $P=0.4$). There is a great difference between the high percentage of allelism in Death Valley localities in 1937 (3.11 ± 0.63) and either the intra- or inter-station results at one locality, Wildrose, in 1939. The numbers at Wildrose were small, but the combined percentage (0.73 ± 0.32 percent) is significantly less than the earlier figure ($P=0.01$). It is possible that conditions differed markedly in 1937 and 1939.

Returning to the data from San Jacinto, the percentage of alleles among flies from the same station (2.13 ± 0.32 percent) is significantly greater ($P=0.0005$) than among flies from different stations of the same locality (0.88 ± 0.20 percent). This in turn is significantly greater ($P=0.009$) than among flies from different localities or regions (0.413 ± 0.081 percent). These are the figures that will be used in further analysis, but, as noted, the two former are probably somewhat too low for estimates of allelism of random flies from the same station or same locality in a single generation.

LETHALS OF COMMON ORIGIN

In testing the allelism of third-chromosome lethals from the Death Valley region, DOBZHANSKY and WRIGHT (1941) have observed that the lethals found repeatedly within the population of a station on an isolated mountain range are not significantly commoner in the populations of other ranges than are the lethals encountered within a station only once. Conversely, the lethals encountered on two or more mountain ranges show no tendency to accumulate in populations of any one range. This observation indicates that the allelic lethals found within the population of a small territory are frequently descendants of a single original mutant, while the allelism of lethals on different mountain ranges is due mainly to independent origin of similar mutants in two or more populations. To put it another way, the lethals found repeatedly within a population are not necessarily those which arise most frequently in the species at large.

Although not all the possible intercrosses between the lethals detected in Mount San Jacinto populations have been made, the material of the present study may be used to check the validity of the above observations. Table 13 shows the number of lethals found once ("singles"), twice ("doublets"), and three times ("triplets") in the populations of the stations examined.

Among the 152 lethals, each of which was found singly within a station (table 13), only 18, or 11.9 percent, were detected in populations of other

stations within the same locality, and 11, or 7.2 percent, were encountered in other localities or regions. Among the 25 doublets, seven, or 28.0 percent, were found at other stations within the locality, and none was found in other localities or regions. Finally, three out of the seven triplets, or 42.9 percent, were met with at other stations within the same locality, and one out of the seven, or 14.3 percent, in a different region. The same lethal was found thrice at Keen C, thrice at Keen D, and once at Keen A, but it was not encountered at all outside the Keen locality. Another lethal was found

TABLE 13

Number of lethals found once, twice, and thrice within the populations of separate stations.

STATION	SINGLES	DOUBLETS	TRIPLETS	STATION	SINGLES	DOUBLETS	TRIPLETS
Keen A	19	4	—	Pinon B	10	3	—
Keen B	11	3	—	Andreas A	15	2	—
Keen C	9	1	2	Andreas B (1939)	15	3	—
Keen D	13	3	2	Andreas B (1940)	4	2	—
Keen E	9	1	—	Wildrose A	15	—	—
Pinon A	21	2	3*	Wildrose B	11	1	—
				Total	152	25	7

* Although the lethals Pinon A 169 and Pinon A 170, as well as Pinon A 170 and Pinon A 429, proved to be alleles, the cross Pinon A 169×Pinon A 429 gave a result showing that these lethals are not alleles. As shown in the previous paper, an appreciable portion of the lethal chromosomes are expected to contain two or more lethal genes of independent origin. Pinon A 170 may have been of this sort. It could also, of course, have been a deficiency covering more than a single locus.

thrice at Pinon A, once at Pinon B, once at Andreas B, and once at Wildrose A, but it was not met with at Keen.

Despite the smallness of the above figures, the situation is clearly that lethals which have attained high frequencies within the population of a given station tend to be found in the populations of other stations within the same locality. This is the opposite of what has been found to be the case in the Death Valley region. The contradiction, however, is explicable if the characteristics of the terrain on Mount San Jacinto and in Death Valley are taken into account. All the collecting stations on Mount San Jacinto are in a territory which supports a dense and continuous population of *D. pseudoobscura*, while the desert gaps which lie athwart the possible migration paths between the mountain ranges of the Death Valley region are obstacles for the movement of the flies.[3] Furthermore, the absolute

[3] Deserts are certainly not absolute barriers for fly migration. A number of collectors (DOBZHANSKY, MAINLAND, MAMPELL, EPLING, and others) have repeatedly taken *D. pseudoobscura* amid desert vegetation, and in spring following a rainy winter they may not be even rare. Nevertheless, the habitat in which permanent, large, and flourishing populations of this species are

distances between the collecting stations on Mount San Jacinto are smaller than those between the ranges in the Death Valley region. A mutant gene which, by chance or otherwise, has attained a high frequency on a mountain range in Death Valley may not spread to the adjacent ranges for a long time, but there is no comparable handicap for the diffusion of genes from population to population within a locality on Mount San Jacinto. For an understanding of the population structure in *D. pseudoobscura*, however, it is important to know that a mutant which is lethal when homozygous may not only increase in frequency in the immediate vicinity of the place of its origin but may spread to populations within a mile or more from that place.

MUTATION RATE

One of the statistics that is needed in attempting to analyze the situation in nature is the rate of occurrence of lethal mutations of the third chromosome of *Drosophila pseudoobscura*. The available data are recapitulated below (DOBZHANSKY and WRIGHT 1941).

TABLE 14

Data on rate of occurrence of lethal mutations in the third chromosome of Drosophila pseudoobscura.

SOURCE OF STOCK	KIND OF EXPERIMENT	CHROMOSOMES TESTED NO.	LETHAL MUTANTS NO.	%	SE	AVERAGES %	AVERAGES SE
Death Valley	direct	3580	9	0.25	0.08	0.297	0.047
Death Valley	cumulative	9892	31	0.31	0.06		
Mexico	cumulative	4177	15	0.36	0.09	0.325	0.065
Guatemala	cumulative	3522	10	0.28	0.09		
Total		21171	65			0.307	0.038

The results from Death Valley (two experiments), Mexico, and Guatemala are in excellent agreement ($n=3$, $\chi^2=0.81$, $P=0.8$). In the study of the situation in Death Valley the figure used was the average mutation rate from Death Valley stock—namely, 0.297 ± 0.047 percent. Since the rate has not been obtained from stock from San Jacinto, the best available estimate for this region seems to be that given by the grand total in table 14—namely, 0.307 ± 0.038 percent. The difference from the figure used for Death Valley is of little importance.

Determinations for Death Valley were made in two ways: (1) a direct determination of mutation rate per generation and (2) a determination

regularly found is unquestionably connected with tree vegetation. The flies can move more freely within the forest-covered portions of the isolated mountain ranges in the Death Valley region (DOBZHANSKY and QUEAL 1938a, 1938b) than across the desert gaps between these ranges.

from experiments in which mutations were allowed to accumulate for several generations before examination. The mutation rate obtained from the latter method should be lower if an appreciable portion of the mutations are strongly selected against as heterozygotes. Since the result was actually higher (though not significantly so), it was concluded that there is no important immediate elimination of lethal mutations due to this cause.

ESTIMATES OF POPULATION CONSTANTS

As brought out in the previous paper (DOBZHANSKY and WRIGHT 1941) it requires a rather large number of constants to give even a very rough description of the genetic situation in a natural population. The description of the breeding structure requires at least three.

N. The effective size of population. Roughly the harmonic mean of the numbers of mating individuals per generation throughout the year and hence much more closely related to the minimum than to the maximum number.

F. The inbreeding coefficient. This measures the departure from random mating within the population in question. With gene frequency q the zygotic frequencies are

$$[(1-F)(1-q)^2+F(1-q)]AA+2(1-F)q(1-q)Aa+[(1-F)q^2+Fq]aa.$$

This is modified by selection but not much if the amount of selective elimination is small.

m. The migration index. This measures the extent to which there is replacement in each generation by immigrants drawn from a population sufficiently large that each lethal is at its equilibrium frequency.

There was some ambiguity in the symbols used in the earlier paper to represent the properties of the lethal genes. Changes are made here to avoid this. Each lethal gene in the given population has a certain frequency (q) resulting from its rate of origin by mutation (v), its rate of elimination in homozygotes, the selection against it when heterozygous (s), and the accumulation of random deviations due to the accidents of sampling. We assume that the homozygous semilethals leave so few descendants that they may be grouped with the complete lethals with no appreciable error.

With regard to the accidents of sampling, let q_i be the frequency of a lethal belonging to the class characterized by particular values of s and v —namely, s_i and v_i. In a small population the genes of this class should exhibit in the long run a distribution, $\phi(q_i)$, characterized by a certain mean, $\bar{q}_i$, and a certain variance $\sigma_{q_i}^2$. The actual mean frequency for all lethals ($\bar{q}$) should equal the mean of these class means except for second order fluctuations. The actual variance of the frequencies of all lethals,

σ_q^2 (σ_{qt}^2 in the previous paper) is analyzable into two components. One of these is the mean of the values of σ_{qi}^2 for all classes, which will be represented by $\sigma_{q(c)}^2$ (represented by σ_q^2 in the previous paper). The other is the variance of class means due to differences in s_i and v_i, which will be represented here by $\sigma_{q(d)}^2$ (represented by $\sigma_{\bar{q}}^2$ in the previous paper).

n Number of loci in third chromosome subject to lethal mutation.

$\bar{v}$ Mean mutation rate per generation per locus.

$\bar{s}$ Mean selective disadvantage of heterozygotes.

$\bar{q}$ Mean frequency of lethals.

σ_q^2 Variance of frequencies of lethals.

$\sigma_{q(c)}^2$ Component of σ_q^2 due to cumulative effects of accidents of sampling.

$\sigma_{q(d)}^2$ Component of σ_q^2 due to differences among lethals in v and s.

p Chance of allelism of two lethal genes taken at random from the specified population (p_{sta} from "stations," p_{loc} from "localities").

p_∞ Chance of allelism of two lethal genes taken at random from a population sufficiently large that each is at its equilibrium frequency.

The 12 constants listed above are not all independent of each other. Equation (1) below holds by definition, and equations (2) and (3) were demonstrated in the previous paper (p. 38). Equation (4) follows from the other three.

$$\sigma_q^2 = \sigma_{q(c)}^2 + \sigma_{q(d)}^2 \tag{1}$$

$$p = n(\bar{q}^2 + \sigma_q^2)/(n\bar{q})^2 \tag{2}$$

$$p_\infty = n(\bar{q}^2 + \sigma_{q(d)}^2)/(n\bar{q})^2 \tag{3}$$

$$n\sigma_{q(c)}^2 = (p - p_\infty)(n\bar{q})^2. \tag{4}$$

Other relations depend on whether v and s are assumed to be the same for all genes, to vary only moderately about their mean, or to vary greatly. As in the previous paper, we shall begin with the assumption that there is only moderate variation of v and s, if any.

None of the above quantities are observed directly. The principal observed quantities are as follows:

V The rate of lethal mutation per generation per *chromosome.*

Q The frequency of lethal bearing third *chromosomes.*

P The chance of allelism of two random lethal *chromosomes* from the specified population (P_{sta} from "stations," P_{loc} from "localities").

P_∞ The chance of allelism of two random lethal *chromosomes* from a population sufficiently large that independent mutant origin is reasonably certain.

The conditions at the three localities on San Jacinto are so different that it is desirable to attempt separate analysis, although the numbers are

hardly adequate except perhaps at Keen. Analysis will also be carried through for the average of these localities and will be compared with revised figures for Death Valley (1937 only) and as far as possible with those from Mexico and Guatemala. The frequencies in all cases are based on the chromosomes other than Standard, instead of on all chromosomes, as in the previous study. The collecting areas at Death Valley were made from areas roughly comparable to smaller localities, Pinon and Andreas, at San Jacinto. The following figures are those which will be used (see table 4, 8, 12, and 14).

	Q	P_{sta}	P_{loc}
San Jacinto			
Keen	0.146 ±0.015	0.0245±0.0049	0.0080±0.0024
Pinon	0.122 ±0.017	0.0191±0.0052	0.0134±0.0050
Andreas	0.144 ±0.018	0.0171±0.0064	0.0051±0.0036
Average	0.1385±0.0097	0.0213±0.0032	0.0088±0.0020
Death Valley	0.165 ±0.014	—	0.0311±0.063
Mexico and Guatemala	0.300 ±0.042	—	—

V=0.00307±0.00038 (based on stocks from Death Valley, Mexico, and Guatemala)
P_∞=0.00413±0.00081 (based largely on Death Valley localities but partly on San Jacinto)

These figures relating to lethal chromosomes must be translated as well as possible into the corresponding expressions relating to lethal genes. As before, it is assumed that independently occurring lethals are distributed in the chromosomes according to the terms of the Poisson series $\exp[-n\bar{q}(1-l)]$ where l symbolizes a lethal. The possibility of simultaneous occurrence of two or more lethals (as from deficiencies) is ignored. The frequency of chromosomes lacking all lethals is thus taken to be $(1-Q)=\exp(-n\bar{q})$

$$n\bar{q} = -\log_e(1-Q). \tag{5}$$

It was shown in the previous paper that the chance of allelism of lethal chromosomes drawn from remote populations must be multiplied by $[Q/n\bar{q}]^2$ to give that for lethal genes.

$$p_\infty = P_\infty[Q/n\bar{q}]^2. \tag{6}$$

Within small populations, lethal chromosomes that prove to be allelic are most likely to trace to a common source and to be alike in all lethal genes. Thus p was assumed to be equal to P in the previous paper. However, there should be the same chance of separate origin as in the case of remote populations, and a small correction $(P_\infty - p_\infty)$ should be made.

$$p_{sta} = P_{sta} - (P_\infty - p_\infty) \tag{7}$$

$$p_{loc} = P_{loc} - (P_\infty - p_\infty). \tag{8}$$

No correction is necessary to relate mutation rate per locus to that per chromosome.

$$n\bar{v} = V. \tag{9}$$

Application of (5) gives the estimates of $n\bar{q}$ given below. Substitution of the values of P_∞, Q and $n\bar{q}$, in (6) gives 0.00345 as the estimate of p_∞ from the Death Valley data and 0.00356 from that from San Jacinto. We shall use the average , $p_\infty = 0.0035$.

The values of p_{sta} and p_{loc} may now be obtained by subtracting 0.0006 ($=P_\infty - p_\infty$) from the values of P_{sta} and P_{loc} (7, 8), respectively. These results permit estimates of $n\sigma_{q(c)}^2$ for stations and localities from the equation (4).

TABLE 15

Estimate of $n\bar{q}$ and of p and $n\sigma_{q(c)}^2$ for both stations and localities.

	$n\bar{q}$	STATION		LOCALITY	
		p_{sta}	$n\sigma_{q(c)}^2$	p_{loc}	$n\sigma_{q(c)}^2$
San Jacinto					
Keen	0.158	0.0239	0.000508	0.0074	0.000097
Pinon	0.130	0.0185	0.000254	0.0128	0.000157
Andreas	0.155	0.0165	0.000314	0.0045	0.000024
Average	0.149	0.0207	0.000382	0.0082	0.000104
Death Valley	0.181	—	—	0.0305	0.000883
Mexico and Guatemala	0.357	—	—	—	—

If it be assumed that v and s are constant, $\sigma_{q(d)}^2 = 0$ by definition. With this assumption, equation (3) reduces to the following, giving $n = 285$.

$$n = 1/p_\infty. \tag{10}$$

Substitution of this value in (9) gives $v = 1.077 \times 10^{-5}$, and substitution in table 15 gives the values of $\bar{q}$, and of $\sigma_{q(c)}$ for stations and localities, respectively, in table 16.

In a random breeding population, the rate of change of the frequency of *chromosomes* carrying one or more completely recessive lethals was given as follows.

$$\Delta Q = V(1 - Q) - \frac{PQ^2(1 - Q)}{1 - PQ^2}. \tag{11}$$

The frequency of lethal bearing chromosomes should fluctuate only slightly about the value of Q at which $\Delta Q = 0$, which is approximately $\sqrt{V/P}$. The actual value of Q is in all cases much smaller, which requires

TABLE 16

Estimates of $\bar{q}$, of $\sigma_{q(c)}$, and $(\bar{s}+F)$ for station and locality, and of the last also for region, all on the assumption that $\sigma_{q(d)}^2 = 0$.

	$\bar{q}$	STATION $\sigma_{q(c)}$	LOCALITY $\sigma_{q(c)}$	STATION $(\bar{s}+F)$	LOCALITY $(\bar{s}+F)$	REGION $(\bar{s}+F)$
San Jacinto						
Keen	0.000554	0.00134	0.00058	0.0158	0.0183	(0.0189)
Pinon	0.000457	0.00094	0.00074	0.0213	0.0220	(0.0230)
Andreas	0.000546	0.00105	0.00029	0.0173	0.0190	(0.0192)
Average	0.000523	0.00116	0.00061	0.0177	0.0194	0.0201
Death Valley	0.000635	—	0.00176	—	0.0116	0.0163
Mexico and Guatemala	0.001252	—	—	—	—	0.0074

that lethals be eliminated more rapidly than under the conditions specified (random mating, elimination only in homozygotes). An increased rate of elimination may be due either to consanguineous mating or to adverse selection in heterozygotes or both.

The rate of change in the frequency of lethal *genes* is as follows, omitting terms involving q^3 and $\bar{s}Fq$ (compare the more complete formula (13) in preceding paper):

$$\Delta q = \bar{v}(1 - q) - m(q - \bar{q}) - q(\bar{s} + F) - q^2(1 - 3\bar{s} - 2F). \qquad (12)$$

By putting $\int_0^1 \Delta q \phi(q) dq = 0$ to express the constancy of the mean frequency of lethal genes

$$(\bar{s} + F) = [\bar{v}(1 - \bar{q}) - (\sigma_{q(c)}^2 + \bar{q}^2)]/[\bar{q} - k(\sigma_{q(c)}^2 + \bar{q}^2)] \qquad (13)$$

where $k = 3$ if $F = 0$ and $k = 2$ if $\bar{s} = 0$ and in general k would be between 2 and 3. Estimates are given in table 16. The estimates of $(\bar{s}+F)$ for regions refer to regions sufficiently large that the frequency of each lethal would be at equilibrium. This is estimated by putting $\sigma_{q(c)}^2$ equal to zero in (13). These estimates are affected only slightly by the assumption made for n.

Joint estimates of N and m have been obtained as in the preceding paper by constructing the frequency distribution $f(q)(=\phi(q)/2N)$ for given $\bar{s}+F$, $\bar{v}$, n, and trial values of N and m.

$$f(q) = C\{1 - 2(\bar{s}+F)q - [1 - (k-1)(\bar{s}+F)]q^2\}^{2N} q^{4N[m\bar{q}+\bar{v}]-1}(1-q)^{4Nm[1-\bar{q}]-1} \qquad (14)$$

It makes no appreciable difference what value of n is chosen. The minimum, $n = 285$, was used. It also makes no appreciable difference whether k is taken as 2 or 3. An intermediate figure 2.5 was used. The frequency of the class of zero gene frequency was estimated from the formula

$$f(0) = f(1/2N)/4N(m\bar{q} + \bar{v}).$$

As an example, consider the case of the intra-station data from San

Jacinto. On substituting $n=285$, $\bar{s}+F=0.0177$, $\bar{v}=0.00001077$, $m=0.05$, $N=2500$ and constructing ordinates at appropriate intervals beginning with 0, 0.0002, 0.0004, etc., the mean was found to be 0.00050 (instead of theoretical 0.00052) and the standard deviation was 0.00086 instead of theoretical 0.00116. Because this estimate of $\sigma_{q(c)}$ is too small, a smaller value of N was tried—namely, $N=1000$—and ordinates were constructed at $q=0$, $q=0.0005$, $q=0.0010$, etc. Calculation yielded $\bar{q}=0.00051$ and $\sigma_{q(c)}=0.00134$. This value of $\sigma_{q(c)}$ is too large. Hence the true value of N (for $m=0.05$) must lie between 2500 and 1000. By interpolating between the logarithms of these values of N according to the differences in $\sigma_{q(c)}$, the estimate $N=1400$ was arrived at. Calculations have been made by this method for the cases listed below, giving N to the nearest hundred (table 17).

TABLE 17

Estimates of effective size of population (N) for various values of the immigration coefficient (m) obtained by construction of the frequency distribution of q for trial values of N and m.

	m=0	m=0.01	m=0.02	m=0.05
Station				
Keen	3,400	—	—	1,100
Av. San Jacinto	4,100	2,900	2,200	1,400
Locality				
Keen	21,000	—	—	—
Av. San Jacinto	16,700	11,600	—	—
Death Valley 1937	2,400	—	—	900

The mean and standard deviation could be obtained directly if Δq were linear (*cf.* WRIGHT 1937). In the present case, Δq is not linear, but a rough approximation can be obtained by balancing the tendency toward increase in the frequency of the lethal—namely, $(\bar{v}+m\bar{q})(1-q)$ by the linear expression,

$$-\left(\frac{\bar{v}}{\bar{q}}+m\right)(1-\bar{q})q$$

that gives the same mean, $\bar{q}$, as the correct expression.

If $\Delta q=-a(q-\bar{q})$, the deviation of gene frequency from the average changes in one generation from $(q-\bar{q})$ to $(1-a)(q-\bar{q})$. The mean sampling variance of $(q+\Delta q)$ is

$$\frac{1}{2N}\int_0^1 (q+\Delta q)(1-q-\Delta q)\phi(q)dq=\frac{1}{2N}\left[\bar{q}(1-\bar{q})-(1-a)^2\sigma_q^2\right]$$

At equilibrium

$$\sigma_q^2 = (1-a)^2\sigma_q^2 + \frac{1}{2N}[\bar{q}(1-\bar{q}) - (1-a)^2\sigma_q^2]$$

$$= \bar{q}(1-\bar{q})/\{2N[1-(1-a)^2] + (1-a)^2\}.$$

This includes the immediate sampling variance. But $\sigma_{q(c)}^2$ as determined by formula (4) is zero if $p = p_\infty$ and thus cannot include the immediate sampling variance. The above formula must accordingly be multiplied by $(1-a)^2$. In the present case $a = (m+\bar{v}/\bar{q})$ (to the degree of approximation used here). Solving for N:

$$N = \left[\frac{\bar{q}(1-\bar{q})}{\sigma_{q(c)}^2} - 1\right]\left(1 - m - \frac{\bar{v}}{\bar{q}}\right)^2 \Big/ 2\left[1 - \left(1 - m - \frac{\bar{v}}{\bar{q}}\right)^2\right] \quad (15)$$

The results agree sufficiently well for the present purposes with those obtained by trial and error from the formula for the distribution of gene frequencies. Some of the results are shown as the lower limits in table 19 and figure 2.

We must now consider the effect of variability in the values of v and s of different lethals.

Extreme heterogeneity with respect to the values of v does not appear to require much consideration. There is no indication in the present case that there are any loci that are giving rise to lethals with exceptionally high frequencies. There may be loci that give lethal mutations with extreme rarity, but if so, they would contribute little to either the mutations observed in the laboratory or found in nature. These can be ignored with restriction of n to loci that give lethal mutations with sufficient frequency to be of importance in the observations.

The greater chance of allelism within than between localities of San Jacinto (0.88% *vs.* 0.57%) might be due to differential selection, related to the marked ecological differences between the localities, instead of to limitation of numbers. This would not, however, account satisfactorily for the much greater chance of allelism within than between stations of the same locality at San Jacinto (2.13% *vs.* 0.88%) or within than between localities at Death Valley (3.11% *vs.* 0.41%) in which cases ecological differences are less pronounced.

The effect of extreme heterogeneity with respect to s (F being assumed to be zero at stations) was examined in the previous paper by postulating two classes of mutations, ones that are completely recessive and ones that are subject to very strong adverse selection as heterozygotes. It is possible for some of the mutations that occurred in the laboratory to be of the latter class, while most of those accumulating in natural populations are of the former class.

There is a third hypothetical class, lethal mutations that are subject to

favorable selection as heterozygotes and which therefore tend to reach exceptionally high frequencies in nature. Such lethals should be detected easily. Since there has been no indication of the occurrence of particular lethals at high frequencies, this possibility will be ignored.

As in the previous paper, we shall designate by subscript 1 the constants relating to completely recessive lethals—namely, n_1, $\bar{s}_1$ (=0), $\bar{q}_1$, etc., and by subscript 2 those relating to the class with strong adverse selection in the heterozygotes—namely, n_2, $\bar{s}_2$, $\bar{q}_2$, etc. There seems no reason for postulating different mean mutation rates for these classes. As in the previous paper, $\bar{q}_2$ is taken equal to v to give the most extreme possible case and $\bar{q}_2^2$ and $\sigma_{q(c)_2}^2$ are assumed to be negligibly small, permitting use of the following equations (*cf* p. 46, DOBZHANSKY and WRIGHT 1941).

$$n_1 v = p(n\bar{q})^2 \qquad (16)$$

$$n_2 v = V - n_1 v \qquad (17)$$

$$n_1 \bar{q}_1 = n\bar{q} - n_2 v \qquad (18)$$

$$n_1 \bar{q}_1^2 = p_\infty (n\bar{q})^2 \qquad (19)$$

$$n_1 \sigma^2_{q(c)_1} = n_1 v - n\bar{q}_1^2 \qquad (20)$$

TABLE 18

Estimates based on the assumption that there are two classes of lethals, n_1 for which $\bar{s}+F=0$ within stations, and n_2 for which s is very high.

	n_1	n_2	v	$\bar{q}_1$	$\sigma_{q(c)_1}$ STATION	$\sigma_{q(c)_1}$ LOCALITY
San Jacinto						
Keen	276	1148	2.16×10^{-6}	0.000563	0.00136	0.00059
Pinon	273	2401	1.15×10^{-6}	0.000467	0.00096	0.00076
Andreas	275	1842	1.45×10^{-6}	0.000555	0.00107	—
Average	275	1560	1.67×10^{-6}	0.000533	0.00118	0.00062
Death Valley	279	551	3.70×10^{-6}	0.000653	—	0.00181

In dealing with localities, it was assumed that n_1, n_2, v, and $\bar{q}_1$ were the same as for stations, which implies that F is not zero in this case but is approximately equal to the excess of $\bar{s}+F$ in localities over that at stations as found under the hypothesis of uniform $\bar{s}+F$.

The results from the different localities are not consistent, especially with respect to n_2 and v. It is certain, however, that the hypotheses that $\bar{s}+F$ is zero for stations and that $\bar{q}_2=v$ are too extreme. The analysis is carried through merely to obtain upper limits for the values of N associated with given values of m_1.

In the case of San Jacinto stations, N was estimated by constructing frequency distributions with various trial values, assuming m=0. This gave N=30,000. Otherwise estimates were obtained by using formula (15). It may be noted that in the case referred to above this gave N=30,200. Some of the results are shown in table 19 and figure 2 (upper limit).

TABLE 19

Rough lower and upper estimates of effective size of population (N) on various assumptions with respect to immigration coefficient (m).

	m=0	m=0.01	m=0.05	m=0.20
Stations	lower upper	lower upper	lower upper	lower upper
Keen	3,900– 20,000	2,500– 5,400	1,000–1,300	240– 260
Pinon	5,200– 52,000	3,600– 9,800	1,500–2,200	390– 430
Andreas	6,100– 46,000	4,000– 9,400	1,600–2,100	380– 420
Av. San Jacinto	3,600– 30,200	3,000– 7,100	1,200–1,700	300– 330
Localities				
Keen	20,200–104,000	13,200–28,000	5,500–6,800	1,300–1,380
Pinon	8,400– 82,000	5,800–16,000	2,500–3,500	630– 710
Av. San Jacinto	16,800–111,000	11,100–26,000	4,500–6,100	1,100–1,220
Death Valley	2,900– 8,700	1,800– 3,100	680– 820	160– 170

These figures depend largely on the difference between the observed values of P and P_∞. This difference exceeds twice the standard error in the case of stations, in the average of the San Jacinto localities and in the Death Valley localities but is somewhat less than this in the data from the localities Keen and Pinon and is wholly without significance in the case of Andreas.

It may be seen from table 19 and figure 2 that the estimated upper limit is as much as tenfold greater than the lower limit in some cases, if m is assumed to be zero. The ratio is much less for m = 0.01 and becomes unimportant, considering other sources of uncertainty for values of m above 0.05.

In interpreting these figures we may note the following very rough estimates of the areas sampled in each case.

	STATION (square meters)	LOCALITY (square meters)	RATIO
San Jacinto			
Keen	10,000	6,000,000	600
Pinon	10,000	300,000	30
Andreas	5,000	50,000	10
Death Valley	—	200,000	—

It may be assumed that m is less than 0.01 in the locality Keen, indicating a value of N of the order 20,000 to 30,000 since close approach to the upper limit can safely be ruled out. The fact that this estimate is based largely on the percentage of allelism of lethals from flies of different generations (caught at intervals up to six months) would tend to make it too large, but only slightly, since the rate of change of gene frequency per generation is small compared with the standard deviation. The area sampled was about two square miles.

On this basis the population sampled at a typical station would have an effective size of only about fifty, but, as may be seen from figure 2, this implies such a high value of m (about 0.50) that the only effective isolation is that of a considerably larger area.

The estimates for station population at Pinon and Andreas are similar to those at Keen. This suggests that the effective density of population is

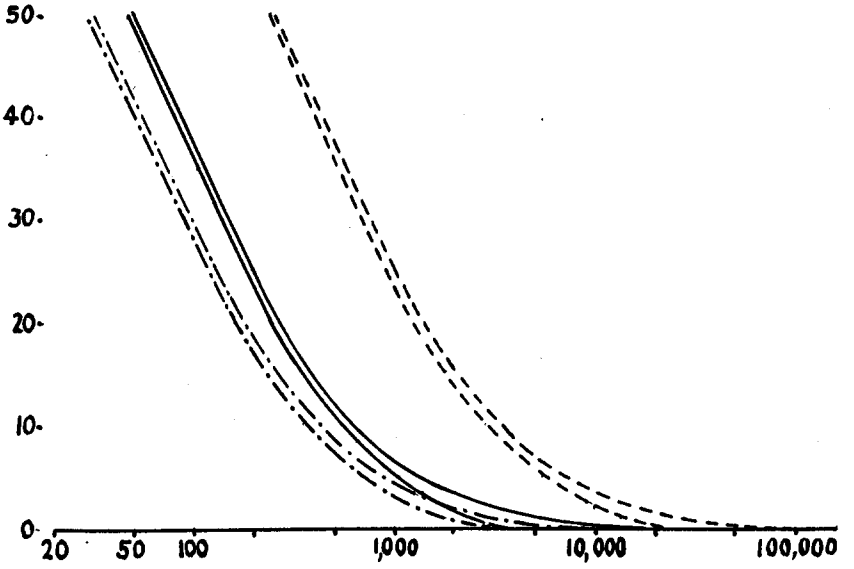

FIGURE 2.—The joint estimates of N, effective size of population (abscissas), and 100 m, replacement by immigration by representatives of the species (ordinates), indicated by the greater frequency of allelism of lethals from the same station or locality than of lethals from different localities. The solid lines are the upper and lower estimates for Keen stations. The broken lines are the corresponding estimates for the whole locality, Keen. The lines given by dash and dot are the estimates for Death Valley localities. These limits are those given by the extreme assumptions with respect to inbreeding and selection in heterozygotes and do not take account of the uncertainties of the figures due to small numbers.

about the same in all of these localities in spite of the differences in conditions. It is possible, however, that there are compensatory differences in N and m.

The value of m should be much larger for the locality Pinon than for the locality Keen because of the smaller area of the former. Accepting this, the estimation of N for Pinon comes out considerably smaller than for Keen and need not be more than about 30 times that of a station in Pinon as expected.

In Death Valley the effective size of population of localities appears at first sight to be smaller than at the San Jacinto stations in spite of their larger area. It is probable, however, that m is considerably smaller, permitting N to be a few times larger. Indeed, it may seem that m should be negligibly small in reference to the relatively much more isolated mountain

forests of Death Valley. It must be recalled, however, that each Death Valley collection was from only a small portion of a mountain forest. It is possible that the population of each of these forests was large enough (10^6 or more) for each lethal gene to be at approximately its equilibrium frequency in a sample representative of the whole forest. If this is the case, the isolation of the forests from each other would have no effect on the frequencies of the lethals. Thus, while m is probably considerably less than for a San Jacinto station, it should not be negligibly small. Another point that may be noted is that the Death Valley collections from a single locality were made at one time and therefore the estimates of N for given m applies more strictly to a single generation.

It is desirable at this point to consider the interpretation of the statistic $\bar{s}+F$. The upper limits of N considered above were on the hypothesis that F is zero (for stations) and that s is zero for most of the genes that accumulate in nature, but very high for most of the mutations observed in the laboratory, leading to their prompt elimination. This hypothesis was apposed by the absence of any evidence of such elimination in experiments in which mutations were allowed to accumulate and also by the similarity of the proportions of semilethals in nature and among the laboratory mutations. It led, moreover, to inconsistent estimates (of n_2 and v).

The hypothesis on which the lower limits of N were based—namely that s is constant (or nearly so) for all genes—led to certain estimates of $(\bar{s}+F)$. This was much smaller for Mexico and Guatemala (0.0074) than for San Jacinto (0.0201) or Death Valley (0.0163) (*cf.* table 16), reflecting the highly significant difference in lethal frequencies (table 4). This difference could be accounted for either by increased severity of selection against heterozygotes in the northern region or by more inbreeding. At first sight, the latter does not seem plausible since $(\bar{s}+F)$ for stations is as high as 0.0177 at San Jacinto, and it is obvious from the discussion above that there can be no isolation of subgroups within stations. It is possible, however, that there is enough tendency to brother-sister mating at times of the year when emerging broods are fewest to make a substantial contribution to F. The value of F for brother-sister mating is 0.25. The value of $(\bar{s}+F)$ for a typical San Jacinto station (0.0177) would require that about 7 percent of the matings be between brother and sister, if all other matings are random. It may be that there is no appreciable amount of brother-sister mating in the southern region where breeding is probably continuous throughout the year and that the value of $(\bar{s}+F)$ there measures wholly selection against heterozygotes. If the same selection applies in the northern region, it would require in addition about 4 percent of brother-sister matings to account for lower frequency of lethals. The data do not permit anything more definite than this survey of alternatives.

Although the lethal mutations have little direct bearing on the evolution of the species, the breeding structure of the population which is revealed by this study must be taken into account in considering possible mechanisms of evolutionary change. Unfortunately estimates of the immigration coefficient, m, of a given territory cannot be directly carried over from one character to another. The ideal case for analysis in terms of N and m is one in which a largely isolated colony with an effective population N receives migrants to the extent m, the migrants being genuinely representative of the species as a whole. In reality, a colony receives its migrants not from the entire species at random but mainly from adjacent localities. Populations of adjacent localities, however, tend to resemble each other to a greater extent than populations drawn from the entire distribution area of the species. For some genes the array of neighboring localities may show a gene frequency that approaches that in the species as a whole; for other genes the same array may be markedly different from the species at large. Thus, the mean frequency of each lethal gene should, as noted above, be close to its equilibrium frequency in a population with an effective size of the order 10^6 which, therefore, should be representative of the whole species. On the other hand, the population of *Drosophila pseudoobscura* is highly differentiated geographically with respect to the relative frequencies of various gene arrangements in the third chromosome (DOBZHANSKY and STURTEVANT 1938; also unpublished data). All the populations on Mount San Jacinto and in Death Valley region resemble each other with respect to these gene arrangements to a much greater extent than they do the populations of, for example, Texas or Mexico. The effective value of the immigration coefficient is therefore very much smaller for the gene arrangements than indicated for lethals.

SUMMARY

Samples of the population of race A of *Drosophila pseudoobscura* were obtained at three widely different localities on Mount San Jacinto in California. At one (Keen) the population minimum is in winter, at another (Pinon) there are two minima, winter and summer, at the third (Andreas) the minimum is in summer. Other data were obtained from a locality near Death Valley.

Among 1292 wild third chromosomes from flies caught in San Jacinto, 13.85 percent contained lethals and semilethals. This compares with 17.0 percent in 857 Death Valley chromosomes (including earlier data) and 30.0 percent in 120 chromosomes from Mexico and Guatemala. These exclude tests of a particular chromosome arrangement (Standard) in which the percentages were slightly less, probably because of technical difficulties.

There were no significant differences among the three "localities" on

San Jacinto or among the "stations" included in these localities. Collections were made throughout the year, but no significant differences were found in this respect.

These results contrast with the marked differences in the frequencies of different chromosome arrangements in different regions or among the three localities of San Jacinto (but not among the stations of the same locality).

The average percentage of allelism of lethals from populations so remote that common origin is largely excluded (different regions or localities) was 0.413 percent with a standard error of 0.081, including all available data. The percentage of allelism of lethals from different stations of the same locality was significantly higher (0.88 ± 0.20). It was still higher from flies from the same station (2.13 ± 0.32). It was somewhat higher in both cases for flies captured simultaneously than at intervals of one to six months apart, but the differences are not significant. These figures are somewhat lower than the figures previously published for flies from the same locality in Death Valley (3.11 ± 0.63).

An attempt was made to make deductions in regard to the breeding structure of the population and the properties of the lethal genes. The minimum number of loci capable of lethal mutation in the third chromosome is estimated at 285, corresponding to which is a mean mutation rate of 1.077×10^{-5} per locus per generation, a mean frequency per lethal of 0.00052 with a standard deviation of 0.00116 in "stations" and 0.00061 in "localities."

The frequency of lethals is much less than would be expected of completely recessive lethals subject to the observed mutation rate and with random mating. The data do not permit separation of the effects of inbreeding (coefficient F) and of selection against heterozygotes (coefficient $\bar{s}$) but indicate a joint estimate $(\bar{s}+F) = 0.0177$ for stations, 0.0194 for localities, and 0.0201 for San Jacinto as a whole, the latter comparable to 0.0163 for Death Valley and 0.0074 for Mexico and Guatemala.

The higher percentage of allelism within stations and localities than at greater distances indicates that the effective size of population (N) in these areas is limited, but estimates can only be made contingent on assumptions with respect to the degree of isolation (immigration coefficient m). Effective N of the largest locality (Keen) area (about two square miles) is estimated at 2 or 3×10^4, assuming m much less than 1 percent. A typical station (diameter 100 yards) is estimated to have an effective N of only about fifty but with substantially no isolation from other stations in the locality.

LITERATURE CITED

DOBZHANSKY, TH., 1939 Genetics of natural populations. IV. Mexican and Guatemalan populations of *Drosophila pseudoobscura*. Genetics **24**: 391–412.

DOBZHANSKY, TH., and M. L. QUEAL, 1938a Genetics of natural populations. I. Chromosome variation in populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. Genetics **23**: 239–251.

DOBZHANSKY, TH., and M. L. QUEAL, 1938b Genetics of natural populations. II. Genic variation in populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. Genetics **23**: 463–484.

DOBZHANSKY, TH., and A. N. STURTEVANT, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23**: 28–64.

DOBZHANSKY, TH., and S. WRIGHT, 1941 Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. Genetics **26**: 23–51.

KOLLER, P. CH., 1939 Genetics of natural populations. III. Gene arrangements in populations of *Drosophila pseudoobscura* from contiguous localities. Genetics **24**: 22–33.

WRIGHT, S., 1937 The distribution of gene frequencies in populations. Proc. Nat. Acad. Sci. **23**: 307–320.