Proc. Nat. Acad. Sci. USA
Vol. 68, No. 10, pp. 2480–2483, October 1971

Polymorphisms in Continental and Island Populations of *Drosophila willistoni**

(Colombia/West Indies/natural selection/chromosome inversions/isozymes/starch-gel electrophoresis)

FRANCISCO J. AYALA, JEFFREY R. POWELL, AND THEODOSIUS DOBZHANSKY

Department of Genetics, University of California, Davis, Calif. 95616

Contributed by Theodosius Dobzhansky, July 29, 1971

ABSTRACT **A comparative study of genic allozyme and chromosomal polymorphisms in four continental (South American) and six oceanic island (West Indies) populations of *Drosophila willistoni* has been made. The pattern of genic polymorphism is closely similar in all populations. Although regional and local differences in gene frequencies are found, generally the same alleles occur at high, intermediate, and low frequencies in all populations. An average individual is heterozygous at 18.4 and 16.2% of its loci in the continental and island populations, respectively. By contrast, chromosomal polymorphism is sharply reduced on the islands compared to most continental populations, and some chromosomal inversions are more frequent on some islands than on others. The observations are not compatible with the hypothesis that most of the gene variants are adaptively neutral. Balancing natural selection is responsible for most of the genic polymorphism in natural populations of *D. willistoni*.**

Recent studies, using the technique of gel electrophoresis, have disclosed a hitherto unsuspected abundance of protein polymorphisms in populations of organisms as diverse as man, mice, *Drosophila* flies, horseshoe crabs, wild oats, and barley (1–10). It is estimated that from 20 to 50% of gene loci are polymorphic, and that an average individual is heterozygous for 5–20% of its loci. Chromosomal polymorphisms, due to inversions of blocks of genes, have been known in natural populations of many species of *Drosophila* for several decades (for a review, see ref. 11). How are these polymorphisms maintained in the populations, and what is their evolutionary significance? Heterotic balancing selection, and perhaps other forms of balancing natural selection, is responsible for the maintenance of most of the inversion polymorphisms (11). What maintains enzyme and other protein polymorphisms is still an open question. One possibility is that balancing natural selection is here also the responsible agency. On the other hand, some genetic variants may be adaptively neutral, and their frequencies may fluctuate in populations owing to the process of random genetic drift. The hypothesis of adaptive neutrality would lead one to expect that the gene frequencies in geographically isolated populations should be uncorrelated. To test this hypothesis, we have examined samples of four continental and six isolated island populations of *Drosophila willistoni*. Gene allele frequencies at 24 loci that produce electrophoretically detectable variants, as well as chromosomal inversions, have been recorded.

* This is paper no. 2 in a series: "Enzyme variability in the *Drosophila willistoni* group." The preceding paper is ref. 5.

MATERIAL AND METHODS

Drosophila willistoni is a species found from southern Florida and tropical Mexico, through the West Indies and Central America, to tropical South America, as far as southern Brazil and northern Argentina. Population samples were taken in December 1970 at four localities east of the Andes in Colombia (Guayabero, Puerto Lopez, Tame, and Betoyes), and in February 1971 on six islands of the Windward Group in the Lesser Antilles (Grenada, Carriacou, Bequia, St. Vincent, St. Lucia, and Martinique). A body of comparative data is available for protein variants (7) and chromosomal inversions (16–18) in many other, chiefly continental, localities.

The techniques of starch-gel electrophoresis for the detection of protein variants in our material are described elsewhere (7). The 24 loci coding for soluble enzymes studied are as follows: Esterases, five loci (*Est-2*, *Est-3*, *Est-4*, *Est-5*, *Est-7*); acid phosphatase (EC 3.1.3.2), two loci (*Acph-1*, *Acph-2*); TPN^+-dependent malate dehydrogenase (EC 1.1.1.40), two loci ((*Me-1*, *Me-2*); adenylate kinase (EC 2.7.4.3), two loci (*Adk-1*, *Adk-2*); hexokinase (EC 2.7.1.1), three loci (*Hk-1*, *Hk-2*, *Hk-3*); and one locus for each of the following: leucine aminopeptidase (EC 3.4.1.1) (*Lap-5*), alkaline phosphatase (EC 3.1.3.1) (*Aph-1*), alcohol dehydrogenase (EC 1.1.1.1) (*Adh*), malate dehydrogenase (EC 1.1.1.37) (*Mdh-2*), glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (*α-Gpdh*), isocitrate dehydrogenase (EC 1.1.14) (*Idh*), octanol dehydrogenase (EC 1.1.1.1) (*Odh*), triose phosphate isomerase (EC 1.2.1.9) (*Tpi-2*); tetrazolium oxidase (*To*), and phosphoglucomutase (EC 2.7.5.1) (*Pgm-1*). For study of the chromosomes, salivary glands of mature larvae were stained in acetic–orceine and freshly made squashes were examined, usually on the same day as they were prepared.

Population samples collected in the wild were brought to the laboratory, and the males were immediately used for electrophoresis. Females were placed in individual culture bottles and allowed to produce progeny. A single larva was used to study the chromosomes, and one F_1 adult was used to study each enzyme tested; each individual thus provides information about two genomes and two sets of chromosomes (except for the males, which carry sex-linked genes in a single dose).

Table 1 gives for 16 loci at each locality the number of genomes sampled, the allelic frequencies, and the proportion of heterozygous individuals expected on the assumption of Hardy–Weinberg equilibrium. At each locus one allele, usually the most common, has been arbitrarily designated 1.00. Other alleles are named with reference to that standard. For exam-

ple, an allele designated 0.96 codes for a protein that in our gels migrates 4 mm less than the standard toward the anode. The localities are listed from the southernmost to the northernmost; the first four are the continental and the last six are island populations.

Genetic variation has been found at every locus studied, at least in some localities. The degree of polymorphism varies considerably, however. At three loci (*Lap-5*, *Est-7*, and *Adk-1*) more than 50% of individuals are heterozygous. At the other extreme, less than 2% of individuals are heterozygous for the *Mdh-2* and *α-Gpdh* loci. At six other loci (*Est-2*, *Est-3*, *Est-4*, *Me-1*, *Tpi-2*, and *Hk-3*), the most common allele has a frequency of 0.9 or greater in all the localities sampled. These eight weakly polymorphic loci are omitted in Table 1.

The patterns of the enzyme polymorphisms are, in general, remarkably similar in all (continental as well as island) populations. The same alleles at a given locus have, as a rule, high, intermediate, and low frequencies in all the localities sampled. Nevertheless, the allelic frequencies are not identical in all populations. For example, the allele 1.03 of the *Lap-5* locus is the most common in the four continental populations in Table 1, and, in fact, throughout most of the distribution area of *D. willistoni* (7). In the six island populations, the allele 1.00 is the most frequent. At the *Est-7* locus, the average frequency of the allele 1.00 increases from 0.56 in the continental populations to 0.67 in the islands. At the *To* locus, allele 0.86 is rare or absent in Colombia, but common in the islands, being the most frequent allele on Bequia and St. Vincent. These two island populations differ from the continental ones also at the *Idh* locus: the allele 1.04 has frequencies of 0.13 and 0.25 on these islands, while it is rare or absent elsewhere.

The proportion of loci that are polymorphic can be used as a measure of the genetic heterogeneity in a given population. If we consider a locus polymorphic if the second-most-common allele has a frequency 0.01 or higher, the proportion of polymorphic loci is 82.4% in Colombia and 79.5% on the islands. If we adopt a criterion of polymorphism that the most common allele should have a frequency of 0.95 or lower, 54.2% of the loci are polymorphic in Colombia and 48.8% on the islands. The average proportion of loci at which an individual is heterozygous is another measure of genetic heterogeneity. This is obtained by averaging over all loci the proportion of heterozygous individuals. These values are given in the bottom line in Table 1, based on 20 loci that have been examined in every locality; the average turns out to be 18.4 ± 0.8% in the continental and 16.9 ± 0.6% in island populations. At most loci, the continental populations are slightly more heterogeneous than the island ones. Two notable exceptions are the *Idh* and *To* loci, which are more variable on the islands than on the continent.

As shown in Table 2, chromosomal polymorphisms behave very differently from enzyme polymorphisms. With the exception of Trinidad, which is geologically a part of the continent of South America, island populations display little chromosomal polymorphism. Continental populations are much more variable, except those in ecologically or geographically marginal localities (16–19). Qualitatively, no chromosomal inversions are endemic to the islands, and the inversions found in the island populations are among the most frequent in continental populations. Furthermore, some island populations are strikingly different from others in the inver-

TABLE 1. *Allelic frequencies at 16 gene loci in populations of Drosophila willistoni*

Gene	Allele	Gua	Plo	Tam	Bet	Gre	Car	Beq	StV	StL	Mar
Lap-5	Sample size	248	402	192	180	266	306	354	258	280	264
	0.98	.12	.09	.12	.13	.04	.05	.04	.03	.03	.06
	1.00	.31	.30	.28	.26	.49	.49	.59	.59	.56	.53
	1.03	.51	.57	.55	.54	.42	.43	.35	.37	.39	.40
	1.05	.04	.04	.04	.06	.04	.03	.02	.01	.01	.01
	heterozygotes	.63	.59	.60	.62	.58	.57	.53	.51	.53	.56
Est-5:	Sample size	246	400	192	178	262	306	352	258	278	264
	0.95	.02	.04	.03	.02	.07	.04	.01	.02	.01	.01
	1.00	.98	.95	.95	.96	.93	.89	.96	.97	.91	.97
	1.05	.01	.01	.03	.02	.01	.06	.02	.01	.07	.01
	heterozygotes	.04	.10	.10	.08	.14	.19	.07	.06	.16	.06
Est-7:	Sample size	183	294	159	111	233	262	258	230	156	153
	0.96	.05	.02	.03	.06	.02	.01	.01	.01	.01	.05
	0.98	.14	.15	.15	.14	.13	.10	.09	.10	.06	.14
	1.00	.58	.58	.57	.53	.65	.65	.67	.70	.74	.63
	1.02	.20	.20	.23	.22	.17	.21	.21	.16	.16	.16
	1.05	.03	.05	.02	.05	.03	.03	.02	.03	.03	.01
	heterozygotes	.60	.56	.60	.65	.53	.52	.49	.48	.42	.55
Aph-1:	Sample size	181	300	155	125	233	264	256	214	159	154
	0.98	.02	.04	.07	.05	.11	.04	.03	.03	.01	.04
	1.00	.92	.91	.90	.94	.85	.94	.95	.96	.98	.94
	1.02	.04	.05	.03	.02	.03	.01	.02	.00	.01	.01
	heterozygotes	.16	.17	.18	.12	.26	.12	.10	.08	.04	.12
Acph-1:	Sample size	246	404	194	180	266	308	348	252	280	264
	0.94	.03	.02	.04	.03	.00	.01	.00	.00	.01	.00
	1.00	.88	.90	.87	.93	.99	.97	.99	1.0	.99	1.0
	1.04	.08	.08	.09	.03	.01	.02	.01	.00	.00	.00
	heterozygotes	.21	.18	.24	.13	.02	.06	.01	.00	.01	.00
Acph-2:	Sample size	18	22	72	14	136	190	210	128	40	44
	0.98	.06	.05	.07	.00	.02	.02	.02	.02	.00	.00
	1.00	.94	.95	.85	.86	.94	.92	.92	.88	.97	.93
	1.03	.00	.00	.07	.14	.04	.06	.05	.09	.03	.07
	heterozygotes	.11	.09	.27	.25	.11	.16	.14	.21	.05	.13
Adh:	Sample size	394	472	194	258	192	44	186	148	152	206
	0.98	.05	.11	.07	.04	.02	.04	.02	.02	.02	.04
	1.00	.94	.87	.93	.95	.96	.96	.97	.97	.98	.95
	heterozygotes	.11	.23	.13	.10	.07	.09	.05	.05	.04	.10
Idh:	Sample size	188	256	142	142	198	220	374	216	238	288
	1.00	1.0	.98	.99	.99	.93	1.0	.79	.75	.97	.94
	1.04	.00	.02	.01	.01	.01	.00	.13	.25	.02	.05
	heterozygotes	.00	.03	.01	.03	.12	.00	.35	.37	.06	.11
Odh-1:	Sample size	180	--	36	120	152	102	146	--	54	72
	0.96	.04	--	.06	.03	.03	.05	.03	--	.04	.04
	1.00	.85	--	.92	.88	.88	.91	.90	--	.91	.81
	1.04	.09	--	.03	.08	.08	.03	.06	--	.06	.14
	heterozygotes	.27	--	.16	.23	.22	.16	.18	--	.17	.33
Me-2:	Sample size	124	78	24	84	120	88	150	176	178	78
	1.00	.60	.91	.29	.80	.92	.98	.93	.90	.96	.88
	1.04	.39	.05	.63	.18	.06	.02	.04	.08	.02	.09
	heterozygotes	.49	.17	.52	.33	.16	.04	.13	.18	.09	.21
To:	Sample size	329	378	161	208	190	220	294	223	170	247
	0.86	.01	.01	.00	.00	.05	.21	.54	.56	.14	.19
	1.00	.98	.99	.97	.99	.94	.79	.46	.44	.85	.81
	heterozygotes	.04	.02	.05	.01	.11	.33	.50	.49	.25	.31
Pgm-1:	Sample size	190	188	138	134	198	218	384	174	234	286
	0.96	.03	.03	.06	.01	.02	.01	.00	.02	.00	.00
	1.00	.95	.88	.91	.98	.97	.98	.91	.79	.70	.77
	1.04	.03	.08	.06	.01	.02	.01	.09	.19	.30	.23
	heterozygotes	.10	.22	.16	.04	.06	.04	.17	.34	.42	.36
Adk-1:	Sample size	190	196	36	98	184	86	208	208	236	278
	1.00	.27	.32	.36	.41	.34	.16	.18	.39	.29	.42
	1.06	.64	.56	.56	.52	.60	.78	.78	.48	.59	.51
	1.12	.09	.10	.08	.07	.05	.06	.04	.13	.11	.06
	heterozygotes	.51	.57	.55	.56	.52	.36	.36	.60	.55	.56
Adk-2:	Sample size	152	230	128	96	198	132	208	214	238	284
	0.96	.02	.01	.05	.01	.04	.02	.01	.01	.01	.02
	1.00	.95	.88	.81	.92	.94	.97	.99	.99	.99	.96
	1.04	.03	.08	.06	.04	.02	.02	.00	.00	.00	.01
	heterozygotes	.10	.22	.33	.16	.11	.06	.02	.01	.02	.08
Hk-1:	Sample size	192	210	138	136	--	132	--	88	60	200
	0.96	.03	.02	.03	.02	--	.04	--	.03	.07	.09
	1.00	.96	.87	.97	.96	--	.93	--	.86	.92	.90
	1.04	.01	.10	.00	.02	--	.03	--	.09	.02	.01
	heterozygotes	.07	.23	.06	.07	--	.13	--	.24	.16	.18
Hk-2:	Sample size	186	200	142	118	66	220	194	180	192	236
	0.96	.01	.04	.01	.02	.02	.01	.02	.01	.03	.01
	1.00	.96	.88	.92	.94	.94	.97	.94	.98	.93	.99
	1.04	.01	.07	.05	.01	.03	.02	.04	.01	.04	.01
	heterozygotes	.07	.22	.16	.11	.12	.06	.11	.03	.13	.02
Proportion of heterozygous loci per individual		.172	.176	.210	.176	.156	.142	.161	.184	.152	.175
Standard error		.044	.042	.044	.043	.038	.036	.038	.043	.039	.041

Eight weakly polymorphic loci and some rare alleles are omitted. Continental populations: Gua = Guayabero, Plo = Puerto Lopez, Tam = Tame, Bet = Betoyes. Island populations: Gre = Grenada, Car = Carriacou, Beq = Bequia, St. Vincent, StL = St. Lucia, Mar = Martinique.

TABLE 2. *Mean numbers of heterozygous autosomal inversions per individual*

Island populations		Continental populations	
Trinidad	2.58 ± 0.29	Goyas, Brazil	6.56 ± 0.31
Barbados	2.05 ± 0.29	Maranhão, Brazil	3.24 ± 0.21
St. Lucia	1.40 ± 0.16	Belem, Brazil	2.75 ± 0.17
Grenada	0.84 ± 0.12	Içana, Brazil	4.65 ± 0.28
St. Vincent	0.76 ± 0.13	St. Marta, Colombia	5.45 ± 0.49
Bequia	0.68 ± 0.09	Bucaramanga, Colombia	5.09 ± 0.35
Martinique	0.56 ± 0.10	Panama	4.73 ± 0.51
Carriacou	0.48 ± 0.14	Costa Rica	4.38 ± 0.44
St. Kitts	0.23 ± 0.08	Salvador	2.06 ± 0.26

sions they contain. Thus, inversion *E* in the right limb of the second chromosome (19) has a frequency of 40% on Bequia, but it has not been encountered on Martinique or St. Vincent. Inversion *F* in the left limb of the second chromosome showed a frequency of 24% on St. Lucia, but was not found on St. Vincent, Martinique, or Grenada. Inversion *A* in the third chromosome has not been found on St. Vincent, Martinique, Bequia, or Carriacou, but reaches a 40% frequency on St. Lucia. Despite the small numbers of the chromosome complements studied (25 per island), these differences are probably significant.

DISCUSSION

An enormous amount of genetic variation, both at the molecular and at the chromosomal levels, has been discovered in *Drosophila willistoni*. That the chromosomal variation is maintained by balancing natural selection is fairly well established, but this hypothesis has not been proved for the molecular-level variation. A hypothesis that much of the molecular-level variation is adaptively neutral is favored by certain authors (12–14). The frequencies in populations of selectively neutral variants are subject to random genetic drift. Differences between species at the molecular level, according to this hypothesis, are mostly the result of random processes. This hypothesis leads to some predictions that can, fortunately, be tested against observable facts.

The number of neutral alleles that can be maintained in a given population is, approximately, $4Nu + 1$, where N is the effective size of the population and u is the mutation rate to neutral alleles (12). If N and u are of the appropriate magnitude, any number of alleles can exist in a given population. However, if we assume, as is reasonable, that the rate of mutation at a given locus is of the same order of magnitude in all populations, differences in population size should result in differences in the number of alleles. We find that, on the contrary, the number of alleles is about the same in all populations studied.

An even more serious difficulty with the neutrality hypothesis is as follows. If we postulate effective population sizes much larger than the reciprocal of the mutation rate, we should find a large number of alleles. For instance, if the population size is 10 times as large as the reciprocal of the mutation rate, about 41 alleles should be segregating in the population. If the size of the population is approximately the reciprocal of the mutation rate or smaller, we should find five or fewer alleles in each population, *but different alleles in different populations*. In any case, whether population size is large or small, the allelic frequencies should not be the same in all populations if the frequencies are governed by random processes. These expectations stand in sharp contrast with our findings. The same alleles appear in most populations, and they occur with frequencies that are highly correlated. The hypothesis of selective neutrality evidently cannot account for the observed pattern of genetic variation.

It has also been suggested that if there is a substantial amount of migration between neighboring populations, the species may effectively approximate a single panmictic population (13). At first sight, this is an attractive hypothesis because it would explain, even with selective neutrality, the similarity of allelic frequencies in different populations. One difficulty, however, is that if the allelic variants are selectively neutral there is no easy way of explaining regional and local differences as have been observed in the present, as well as in other, studies (2, 6, 7, 10, 15). Why should two populations, such as Puerto Lopez and Tame, have very similar allelic frequencies at some loci, e.g., *Idh* and *Lap-5*, but different frequencies at *Me-2* and *Acph-2*?

The similarities of gene frequencies between continental and island populations cannot be accounted for by migration. We have chosen six oceanic islands that were not connected with each other or with the continent of South America in geological history. Their *Drosophila* inhabitants must have reached them by accidental transport of small numbers of founders. Substantial differences in the chromosomal polymorphisms between the islands and the continent, and between different islands, attest to their geographic isolation. It is natural selection that is the main factor controlling the genetic variation in natural populations of *D. willistoni*. Further studies are, of course, required to ascertain the relative roles played by heterotic, diversifying, frequency-dependent, and other forms of balancing selection (11, 15).

Comparison of the genic and chromosomal polymorphisms in the populations studied is very instructive. At least 50 distinct inverted sections have been recorded in the chromosomes of *D. willistoni* (16–19). The extent of chromosomal polymorphism varies greatly in different populations, being reduced on oceanic islands as compared to the continent and continental islands (e.g., Trinidad, Table 2). Founder effect is responsible for the meager inversion polymorphism on oceanic islands. There are good reasons to think that each chromosomal inversion is monophyletic, i.e., arose only once, even though it may now occur throughout the distribution area of the species. Gene variants detected by electrophoretic techniques may, on the contrary, arise repeatedly by mutation. No new inversions have been found on the islands; the inversions found there are those introduced by the founders transported from the continent or from other islands. No reliable estimates of the mutation frequencies giving rise to enzyme variants are at present available. The populations of *D. willistoni*, even on the smaller islands, are not small; they are probably in the millions. Assuming mutation rates as low as 10^{-7} per generation, it is reasonable for us to infer that a given allozyme mutant arises at least once every few years on every island. If selectively favored, it may be incorporated into the population, even though it was not introduced by the original founders.

This work was supported by NSF grant GB-20694 (International Biological Program), AEC contract AT (30-1) 3096, and USPHS Career-Development Award K3GM 37265, and was done while we were at Rockefeller University, New York City.

1. Lewontin, R. C., and J. L. Hubby, *Genetics*, **54**, 595 (1966).
2. Prakash, S., R. C. Lewontin, and J. L. Hubby, *Genetics*, **61**, 841 (1969).
3. Selander, R. K., W. G. Hunt, and S. Y. Yang, *Evolution*, **23**, 379 (1969).
4. Selander, R. K., S. Y. Yang, R. C. Lewontin, and W. E. Johnson, *Evolution*, **24**, 402 (1970).
5. Ayala, F. J., C. A. Mourao, S. Perez-Salas, R. Richmond, and Th. Dobzhansky, *Proc. Nat. Acad. Sci. USA*, **67**, 225 (1970).
6. Richmond, R., *Genetics*, in press.
7. Ayala, F. J., J. R. Powell, M. L. Tracey, C. A. Mourao, and S. Perez-Salas, *Genetics*, in press.
8. Stone, W. S., M. R. Wheeler, F. M. Johnson, and K. Kojima, *Proc. Nat. Acad. Sci. USA*, **59**, 102 (1968).
9. O'Brien, S. J., and R. MačIntyre, *Amer. Natur.*, **103**, 97 (1969).
10. Allard, R. W., and A. L. Kahler, *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, in press.
11. Dobzhansky, Th., *Genetics of the Evolutionary Process* (Columbia University Press, New York, 1970).
12. Kimura, M., and J. F. Crow, *Genetics*, **49**, 725 (1964).
13. Kimura, M., and T. Ohta, *Nature*, **229**, 467 (1971).
14. Kimura, M., and T. Ohta, *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, in press.
15. Ayala, F. J., *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, in press.
16. DaCunha, B., and Th. Dobzhansky, *Evolution*, **8**, 119 (1954).
17. Dobzhansky, Th., *Evolution*, **11**, 280 (1957).
18. DaCunha, B., Th. Dobzhansky, O. Pavlovsky, and B. Spassky, *Evolution*, **13**, 389 (1959).
19. DaCunha, B., H. Burla, and Th. Dobzhansky, *Evolution*, **4**, 212 (1950).