ENZYME VARIABILITY IN THE *DROSOPHILA WILLISTONI* GROUP. IV. GENIC VARIATION IN NATURAL POPULATIONS OF *DROSOPHILA WILLISTONI*[1]

FRANCISCO J. AYALA, JEFFREY R. POWELL, MARTIN L. TRACEY, CELSO A. MOURÃO[2] AND SANTIAGO PÉREZ-SALAS[2]

Department of Genetics, University of California at Davis, Davis, California 95616

Manuscript received May 4, 1971
Revised copy received October 19, 1971

ABSTRACT

We describe allelic variation at 28 gene loci in natural populations of *D. willistoni*. Seventy samples were studied from localities extending from Mexico and Florida, through Central America, the West Indies, and tropical South America, down to South Brazil. At least several hundred, and often several thousand, genomes were sampled for each locus. We have discovered a great deal of genetic variation. On the average, 58% loci are polymorphic in a given population. (A locus is considered polymorphic when the frequency of the most common allele is no greater than 0.95). An individual fly is heterozygous, on the average, at 18.4% loci.——Concerning the pattern of the variation, the most remarkable finding is the similarity of the configuration of allelic frequencies from locality to locality throughout the distribution of the species. Our observations support the conclusion that balancing natural selection is the major factor responsible for the considerable genetic variation observed in *D. willistoni*.

THE *willistoni* group of Drosophila consists of at least twelve closely related species endemic to the tropics of the New World. Six of the species are siblings, morphologically nearly identical, distinguishable only by some slight differences in the male genitalia. Two of the siblings, *D. insularis* and *D. pavlovskiana*, are narrow endemics, the former in some islands of the Lesser Antilles and the latter in Guyana. Four other sibling species, namely, *D. willistoni*, *D. paulistorum*, *D. equinoxialis*, and *D. tropicalis* have wide and largely overlapping distributions throughout the New World tropics. Twenty years of intensive study have provided a large body of information concerning the evolutionary biology of these species. Their geographic distribution and relative abundance is reasonably well known (for a recent summary, see SPASSKY *et al.*, 1971). Information is also available about the reproductive affinities and incompatibilities within and between species, based mostly on studies of ethological (sexual) isolation and hybrid sterility (SPASSKY *et al.* 1971, and references therein). Studies

[1] Supported by NSF Grant GB 20694, AEC contract AT (30-1) 3096, and NIH Career Development Award K3GM 37265. This work was done while the authors were at Rockefeller University, New York 10021.

[2] Permanent addresses: Mourão—Faculdade de Filosofia, São José do Rio Prêto, São Paulo, Brazil; Pérez-Salas—Universidad Central de Venezuela, Caracas.

of the salivary gland chromosomes have uncovered a large number of inversion polymorphisms (DOBZHANSKY 1950; DA CUNHA, BRNCIC and SALZANO 1953; KASTRITSIS 1969, and references therein). The geographic distribution and frequencies of the chromosomal polymorphisms, and the relationship between the chromosomal polymorphisms and ecologically important environmental parameters have also been the subject of intensive studies (DOBZHANSKY 1950; DA CUNHA *et al* 1959, and references therein).

Techniques for the separation of enzymes and other proteins by means of gel electrophoresis have proved to be powerful tools for the study of the genetics of natural populations (LEWONTIN and HUBBY 1966; JOHNSON *et al.* 1966; HARRIS 1969; SELANDER and YANG 1969; KOJIMA, GILLESPIE and TOBARI 1970). Using these techniques we have undertaken a study of the genetic variation in the sibling species of *D. willistoni.* These species are excellent and challenging evolutionary materials. We think that studies of their variation at the single gene level, may provide valuable insights for an understanding of evolutionary problems. Our goal is to measure the *amount* of genetic differentiation within and between the species; to describe the *pattern* of genetic variation throughout the geographic distribution of each species; to relate the variation of the species to each other and to the environment; and to understand the evolutionary mechanisms which produce and maintain the patterns of genetic variation. We have published a preliminary report on the amount of genetic differentiation among the sibling species (AYALA *et al.* 1970). The present paper contains a description of the genetic variation in *D. willistoni.* Twenty-eight randomly selected gene loci have been studied in 70 samples from natural populations.

MATERIALS AND METHODS

Populations studied: The samples studied cover most of the distribution range of *D. willistoni.* They extend from Los Molinos, Sinaloa, Mexico, about 25°N and 107°W, to Reuter, in southern Brazil, about 30°S and 50°W. The distance between Los Molinos and Reuter is about 9,000 km, as the crow flies—a span of 55 degrees of latitude and 57 degrees of longitude. The localities surveyed are shown in Figure 1. Triangles indicate those localities from which one, or only a few strains were available for the enzyme survey. Circles indicate those localities from which many wild genomes were studied. A listing of the populations follows. The numerals refer to the numbers given in Figure 1. Additional details concerning most of the collections can be found in SPASSKY *et al.* (1971).

1, Sinaloa, Mexico—one strain descended from a single female collected in Los Molinos; 2, Tehuacan, Mexico; 3, Quirigua, Guatemala; 4, Lancetilla, Honduras; 5, Tortuguero beach, Costa Rica—two collections about 5 km apart; 6, Jaque, Panama; 7, Lake Placid, Florida; 8, Grand Cayman Island; 9, Mayaguez, Puerto Rico; 10, Monserrat, Virgin Islands; 11, Santa Marta (Las Nubes), Colombia; 12, Piojo, Colombia; 13, Turbo, Colombia; 14, Teresita, Colombia; 15, Malaga, Colombia; 16, La Macarena, Colombia; 17, Valparaiso, Colombia—two collections a few days apart; 18, Mitu, Colombia—three collections about one km apart from each other; 19, Leticia, Colombia—three collections: two about one km apart from each other in Leticia, and one collection in Marco, Brazil, about 3 km from the Leticia collections; 20, Trinidad—seven collections, one in each of the following places, a few km apart from each other: St. Patrick, St. Isabel, Arima Valley, Simla Valley, Blanchisseuse Road, Mora Forest, St. Pablo; 21, Caripe, Venezuela; 22, southeast of Caracas, Venezuela—three collections, a few km apart: Burguillo, Guatopo, Marrero; 23, Rancho Grande, Venezuela; 24, Puerto Ayacucho, Venezuela—two collections a

FIGURE 1.—Localities from which samples were studied for genetically determined enzyme variation. Triangles: localities from which one, or only a few strains were available; circles: localities from which many genomes were studied. See the section on MATERIALS AND METHODS for the names of localities.

few km apart; 25, Ocamo, Venezuela; 26, Perija (Machiques), Venezuela; 27, Pichilingue, Ecuador; 28, Santo Domingo de los Colorados, Ecuador; 29, Lima, Peru; 30, Kanuku Mountains, Guyana—two collections about two months apart; 31, Tracajatuba, Amapa, Brazil; 32, Macapa, Brazil; 33, Belem, Brazil; 34, Santarem, Brazil; 35, Manaus, Brazil; 36, Tapuruquara, on the Rio Negro, Brazil—two collections about 10 km apart; 37, Içana, on the Rio Negro, Brazil; 38, Tefe, Brazil; 39, Cruzeiro do Sul, Brazil; 40, Japiim, Brazil; 41, Imperatriz, Brazil; 42, Palma, Goiaz, Brazil; 43, Monjolinho, Brazil; 44, Sao Jose do Rio Preto, Brazil—four collections about a month apart from each other, from November, 1968, through January, 1969; 45, Mirassol, Brazil, 15 km east of Sao Jose do Rio Preto—six collections: five about one month apart from September 1968, through January, 1969, and a sixth collection in March, 1970; 46, Riberao Preto, Brazil; 47, Iguassu Falls, in the border between Brazil, Uruguay, and Argentina; 48, Reuter, Brazil.

Gel Preparation: Two kinds of starch are used: Sigma (from Sigma Chemical Co., St. Louis, Mo.) and Electrostarch (from Otto Hiller Electrostarch Co., Madison, Wis.). A 12.5% (w/v) solution of starch in gel buffer is heated with vigorous swirling to near boiling point in a side-arm filter flask. The starch solution is degassed with an aspirator and poured in a plexiglass mold

19.5 × 17.5 × 1 cm (approximate volume, 400 ml). The gel is covered with saran wrap and allowed to cool overnight.

Gel and Electrode Buffers: Three buffer systems are used: (A), gel buffer: 76 mM Tris and 5 mM citric acid, pH 8.65; electrode buffer: 300 mM boric acid and 60 mM NaOH (POULIK 1957). (B), gel and electrode buffer: 87 mM Tris, 8.7 mM boric acid, 1 mM EDTA, and 1 mM β-NAD+, pH 9.0. (C), gel buffer: 9 mM Tris, 3 mM, citric acid, pH 7.0; electrode buffer: 135 mM Tris, and 45 mM citric acid.

Sample Preparation: Single flies are ground in small wells in a plastic (delrin) block with 0.05 ml of gel buffer or distilled water. This crude homogenate is absorbed with one or two, 1 × 0.4 cm, pieces of filter paper (Whatman #1). The paper pieces containing the samples are inserted along a slot in the gel made about 5 cm from the edge. The gel is placed horizontally over two buffer trays with electrodes. Sponge wicks are used to establish contact between the gel and the electrode buffer in the trays. This set-up is placed in a refrigerator at 4°C and connected with a Heathkit (model 1P-17) power supply.

After the run is completed, the two ends of the gel which were in contact with the sponge wicks are cut off. The rest of the gel is sliced horizontally into five 2 mm-thick slices with help of a taut wire and 2 mm-thick guides. The top slice is discarded since surface effects make it unusable. The slices can be used for four different enzyme assays. The crude fly homogenate is sometimes absorbed with two pieces of filter paper, and each sample placed in a different gel. Thus, as many as eight enzyme assays can be obtained from an individual fly.

Enzyme Assays: Many assays are modified from SHAW and KOEN (1968), and BREWER (1970). *Acid phosphatase* (Acph) Sigma strach, buffer system A, run for 4 hr at 20 v/cm. The gel is soaked for two hours in 0.5 M boric acid. After removal of the boric acid, the following stain is used: 500 mg PVP (polyvinyl pyrrolidone), 100 mg Na-α-naphthyl acid phosphate, 100 mg Fast Blue BB salt, in 100 ml 0.125 M acetate buffer, pH 5.0. *Adenylate kinase* (Adk): Sigma starch, buffer system C, 4 hr at 20 v/cm. Strain: 90 mg glucose, 20 mg $MgCl_2$, 25 mg TPN+ (=NADP+), 30 mg NBT (nitro blue tetrazolium), 80 units glucose-6-phosphate dehydrogenase, and 160 units hexokinase in 100 ml 0.05 M Tris-HCl buffer, pH 7.1. After incubating for one hour, add 3 mg PMS (phenazine methosulfate). *Alcohol dehydrogenase* (Adh): electrostarch, buffer B, 6 hr at 25 v/cm. Stain: 4.5 ml isopropanol, 25 mg β-NAD+, 20 mg NBT in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. After 1½ hr add 5 mg PMS and keep in dark. *Aldolase* (Ald): electrostarch, buffer C, 4 hr at 20 v/cm. Stain: 550 mg fructose-1,6-diphosphate (tetrasodium salt), 50 mg β-NAD+, 30 mg NBT, 100 units glyceraldehyde-3-phosphate dehydrogenase in 100 ml 0.05 M Tris-HCl buffer, pH 7.1. After one hr add 3 mg PMS. *Alkaline phosphatase* (Aph): Sigma starch, buffer A, 4 hr at 20 v/cm. Stain: 500 mg PVP, 100 mg Fast Blue BB, 100 mg α-naphthyl acid phosphate, 60 mg $MgCl_2$, 60 mg $MnCl_2$, 2 g NaCl in 100 ml 0.05 M Tris-HCl buffer, pH 8.6 *Esterase* (Est): Sigma starch, buffer A, 4 hr at 20 v/cm. Incubate for half hour in 0.5 M boric acid. Stain: 1.5 ml 2% α-napthyl-acetate solution (1:1 water: aceton is used as solvent), 60 mg Fast Garnet GBC salt in 100 ml 0.1 M phosphate buffer, pH 6.5. *Glyceraldehyde-3-phosphate dehydrogenase* (G3pdh): Electrostarch, buffer C, 4 hr at 20 v/cm. Stain: incubate at 37°C, 550 mg fructose-1, 6-diphosphate (tetrasodium salt) and 100 units of aldolase in 100 ml Tris-HCl buffer, pH 7.1. After half hr add 50 mg β-NAD+, 30 mg NBT, 150 mg Na_3AsO_4, and pour over gel. After 1½ hr add 4 mg PMS. *α-Glycerophosphate dehydrogenase* (αGpdh). Electrostarch, buffer B, 6 hr at 25 v/cm. Stain: 25 mg β-NAD+, 20 mg NBT, 180 mg EDTA, 800 mg α-glycerophosphate in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. After 1½ hr add 5 mg PMS and keep in the dark. *Hexokinase* (Hk): Sigma starch, buffer C, 4 hr at 20 v/cm. Stain: 90 mg glucose, 20 mg $MgCl_2$, 25 mg ATP, 25 mg TPN+, 20 mg NBT, 80 units glucose-6-phosphate dehydrogenase, in 100 ml 0.05 M Tris-HCl buffer, pH 7.1. After 1½ hr add 3 mg PMS. *Isocitrate dehydrogenase* (Idh): Electrostarch, buffer C, 4 hr at 20 v/cm. Stain: 150 mg sodium isocitrate, 20 mg TPN+, 20 mg NBT, 20 mg $MnCl_2$ in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. After one hr add 4 mg PMS. *Leucine amino peptidase* (Lap): Sigma starch, buffer A, 4 hr at 20 v/cm. Incubate in 0.5 M boric acid for ½ hr. Stain: 70 mg L-leucyl-β-naphthylamide HCl, 40 mg black K salt, 50 ml 0.2 M maleic anhydride 0.2 M NaOH solution, 10 ml 0.2M NaOH, and 40 ml H_2O; pH is 5.0. *Malate dehydrogenase* (Mdh): electrostarch, buffer B, 6 hr at 25 v/cm. Stain:

20 mg NBT, 25 mg β-NAD+, 50 mg malic acid in 100 ml 0.05 M Tris-HCl, buffer pH 8.5. After 1½ hr add 5 mg PMS and keep in dark. *Malic enzyme* (Me): This is TPN+-dependent malate dehydrogenase. Sigma starch, buffer C, 4 hr at 20 v/cm. Stain: 50 mg malic acid, 20 mg NBT, 25 mg TPN+, 25 mg $MgCl_2$, 5 mg PMS in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. *Octanol dehydrogenase* (Odh): Electrostarch, buffer B, 6 hr at 25 v/cm. Stain: 0.2 ml octanol, in 1 ml 95% ethanol, 25 mg β-NAD+, 20 mg NBT in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. After 8 hr add 5 mg PMS. *Phosphoglucomutase* (Pgm): Sigma starch, buffer C, 4 hr at 20 v/cm. Stain: 600 mg glucose-1-phosphate (disodium salt), 200 mg $MgCl_2$, 10 mg TPN+, 20 mg NBT, 80 units glucose-6-phosphate dehydrogenase in 100 ml 0.05 M Tris-HCl buffer, pH 7.1. After one hr add 3 mg PMS. *Tetrazolium oxidase* (To). Electrostarch, buffer B, 6 hr at 25 v/cm. Stain: 25 mg β-NAD+, 20 mg NBT, 5 mg PMS in 100 ml 0.05 M Tris-HCl buffer, pH 8.5. Expose to light until white bands appear on blue background. *Triosephosphate isomerase* (Tpi): Electrostarch, buffer C, 4 hr at 20 v/cm. Stain: 2 g α-glycerophosphate (sodium salt), 1.1 g sodium pyruvate, 50 mg β-NAD+, 200 units α-glycerophosphate dehydrogenase, 200 units lactic dehydrogenase in 50 ml 0.05 M Tris-HCl, pH 8.0. Incubate for 2 hr at 37°C. Take to pH 2.0 with 1 N HCl, then to pH 7.0 with 1 M Tris. Add 60 mg β-NAD+, 30 mg NBT, 250 mg Na_3AsO_4, 100 units glyceraldehyde-3-phosphate dehydrogenase and H_2O to 100 ml. Pour this solution over gel. After one hr add 4 mg PMS.

Gel Fixation: After the enzyme bands appear, the reaction is stopped by washing the gel with water and adding fixing solution: 45 parts methanol to 55 parts acetic acid solution (glacial acetic acid diluted 1:5 in H_2O). After 8 hr the gel can be wrapped and sealed with saran wrap paper for storage. Most gels keep their original appearance for many months. However, with stains using NBT the bands start fading after a few weeks.

RESULTS

In the section on MATERIALS AND METHODS we have indicated the abbreviations used to designate each enzyme. These same abbreviations written in italics are used to represent the genes coding for the enzymes. When several forms of the same enzyme exist, each controlled by a different gene locus, we add to the symbol of the enzyme a hyphenated numeral. The isozyme with the least anodal migration is called one, the next is two, and so on. Thus the *Est-4* locus codes for enzymes which generally migrate less towards the anode than the enzymes coded by *Est-5*, but more than those coded by *Est-3*. We have found allelic variants at all loci studied. One allele is given the arbitrary value of 1.00; an allele coding for an enzyme which migrates approximately 5 mm more towards the anode is called 1.05, and one whose enzyme migrates 5 mm less is called .95. Following generally accepted practice in genetic nomenclature, we write alleles as superscripts over the symbol representing the locus. $Est\text{-}5^{.95}$, $Est\text{-}5^{1.00}$, and $Est\text{-}5^{1.05}$ are three alleles at the *Est-5* locus.

We have studied 28 enzyme loci in *D. willistoni*. The frequency distributions for 25 loci are given in Tables 1–17. Relatively small samples were studied at three loci, *Acph-2*, *Odh-1*, and *Odh-2*. The data for these three loci are summarized in Table 19. Each enzyme reported in this paper appears to be coded for by a single locus. It is possible, however, for some enzyme to be coded for by two or more genes, only one of which exists in multiple allelic forms. The enzyme would, then, appear as controlled by a single gene. All alleles are codominant with the exception of "null" or "silent" alleles, which act as recessives. We have detected null alleles at only a few loci, mostly sex-linked, and always at very low

frequencies. The frequencies of null alleles, whenever found, are included in the "other alleles" column in the tables. Because of the recessive character of null alleles, it is possible that we may have sometimes failed to discover them or underestimated their frequency. The heterozygotes at some loci exhibit a "hybrid" band, indicating that the allozyme is at least a dimer. The formal genetics and chromosomal mapping of the enzyme loci included in the present study will be reported elsewhere.

Our sampling procedure is as follows. Upon arrival to our New York laboratory of the freshly collected samples, the males are immediately used for electrophoresis. Each female is placed in a separate culture bottle and one F_1 progeny is studied for each enzyme system. Thus we are studying two wild genomes for each individual collected in nature except for sex-linked loci which are carried by the wild males in a single dose. In a minority of samples, the progeny of the wild females was not studied until the F_2 or F_3 generation. We have also studied one individual of each strain in such cases, and still assumed that we are sampling two wild genomes. For these samples we are, then, underestimating the amount of genetic variation. From a few localities we had no freshly collected material available, but rather a single strain kept in the laboratory usually for several years. We have studied three to six individuals from each of these strains, and assumed that we are sampling two wild genomes. In two or three instances, three alleles were found segregating in a laboratory strain; we have assumed that three wild genomes were sampled in that strain at that locus.

In Tables 1–17 we give, for each specified locality, the number of genes sampled at each locus. Data from small samples (usually less than 10 genomes) have been pooled with the data from laboratory strains under the entry, "Other localities". The frequency at which each allele is found in each locality is given in the body of the Tables. The total number of genomes sampled and the allelic frequencies for the whole species are given in the last row. We have estimated the frequencies for the species by dividing the total number of times a given allele was found by the total number of genomes sampled for the whole species. With one exception (*Odh-2*) several hundred, and often several thousand, genes have been sampled at each locus. The last column of the Tables gives the proportion of individuals expected to be heterozygous in each locality at each locus assuming Hardy-Weinberg equilibrium. The agreement between the observed and the expected proportion of heterozygotes is generally reasonably good. When several samples from the same locality were available—whether samples taken at different times, or in different sites—the value given for expected heterozygosity is the unweighted average of the various samples. That is, we have estimated the expected heterozygosity for each sample of more than 20 genomes assuming Hardy-Weinberg equilibrium; then we have obtained a simple average of these heterozygosities. The degree of heterozygosity was in every case nearly identical for the various samples of a given locality. For the small samples pooled under "other localities," the heterozygosity has been calculated by averaging the expected heterozygosity of all the samples included in this category which consist of at least eight genomes.

We have found variant alleles at every locus studied, although not in every population. This is not surprising. Given a sufficiently large sample, variation must exist at every locus. But we have uncovered an enormous amount of genetic variation. The amount of variation varies considerably from locus to locus. In some loci, like *αGpdh*, *Idh*, *Odh-2*, *To*, *Tpi-2* and *Hk-3*, the most common allele reaches an average frequency of 0.98 over the whole species. The average proportion of heterozygous individuals at each locus ranges from a low of 2.7% at the *αGpdh* locus to a high of 64.2% at the *Lap-5* locus (Table 19). Within these values the proportion of heterozygous individuals forms nearly a continuum, although there is a gap between *Est-4* with 31.8% and *Adk-1* with 54.5% heterozygous individuals.

For convenience of discussion we may classify the 28 loci into four categories according to the degree of polymorphism. The actual boundaries between these categories are, of course, arbitrary. These categories going from more to less heterozygosity are as follows (see Table 19): (1) At three loci (*Lap-5*, *Est-7*, *Adk-1*), the proportion of heterozygotes is greater than 0.5. (2) At fifteen additional loci (*Est-2*, *Est-3*, *Est-4*, *Est-6*, *Aph-1*, *Acph-1*, *Acph-2*, *Ald*, *Adh*, *G3pdh*, *Odh-1*, *Me-2*, *Pgm-1*, *Adk-2*, and *Hk-2*) the proportion of heterozygotes is greater than 0.1. (3) Six more loci (*Est-5*, *Mdh-2*, *Idh*, *Odh-2*, *Me-1*, and *To*) are polymorphic enough to average 0.05 to 0.10 heterozygous individuals per population. (4) In only four out of 28 loci (*αGpdh*, *Tpi-2*, *Hk-1*, and *Hk-3*) the proportion of heterozygotes is lower, on the average, than 0.05.

That some gene loci are more variable than others was to be expected. What is remarkable is that all loci are strikingly similar in the pattern of the variation. Two generalizations can be made. First, and most notable, *at any given locus one and the same particular allele is* (with few and minor exceptions) *the most common allele throughout the whole distribution of the species*. Our samples of *D. willistoni* extend over a geographic area amounting to several million square kilometers. Yet, at each and every locus, the allele which is most frequent in, say, Costa Rica is usually also the most common in Trinidad, in south Brazil, and in every other population of the species. The second generalization is that *the gene frequencies are not uniform throughout the whole species*. The frequency of the most common allele at any given locus does not remain constant from one to another locality, but rather it fluctuates sometimes over wide limits. There are localities and regions where certain alleles exist at substantially higher frequencies than in other localities. There exist also locally indigenous alleles—present in one or a few localities but absent elsewhere.

A survey of Tables 1–17 shows that some heterogeneity in the frequency distribution of the alleles occurs at almost every locus. That is, local differences exist in highly polymorphic as well as in nearly monomorphic loci. To underline this point we will point out some instances of local differentiation in loci with various levels of polymorphism.

1. *Loci with more than fifty percent heterozygous individuals. Lap-5* is the most polymorphic of the loci we have studied. Within continental South America there is some clinal variation, with allele 1.03 generally increasing in frequency

TABLE 1

Frequency of alleles at the Lap-5 *locus of D. willistoni*

Locality	Genes sampled	Alleles						Proportion of individuals heterozygous
		.96	*.98*	*1.00*	*1.03*	*1.05*	*Other**	
5. Tortuguero	38	—	.16	.29	.53	.03	—	.614
6. Jaque	194	.03	.18	.34	.37	.06	.02 (2)	.709
9. P. Rico	12	—	—	.42	.50	.08	—	.569
14. Teresita	260	.03	.14	.32	.36	.14	.02 (2)	.728
16. La Macarena	78	—	.18	.30	.39	.14	—	.713
17. Valparaiso	82	—	.15	.26	.40	.19	—	.695
18. Mitu	160	.03	.09	.23	.41	.24	.01 (1)	.710
19. Leticia	64	.03	.13	.37	.39	.08	—	.639
20. Trinidad	1,072	.01	.06	.23	.56	.12	.02 (3)	.622
21. Caripe	76	.01	.05	.30	.53	.11	—	.617
22. SE of Caracas	316	.01	.11	.23	.52	.12	—	.646
24. Ayacucho	306	.003	.09	.34	.51	.07	—	.616
25. Ocamo	80	.01	.04	.39	.43	.14	—	.649
30. Guyana	606	.01	.10	.32	.48	.09	.002(1)	.636
32. Macapa	56	—	.11	.23	.63	.04	—	.543
33. Belem	74	—	.10	.22	.47	.22	—	.674
34. Santarem	492	.02	.14	.39	.43	.02	.01 (2)	.649
35. Manaus	462	.01	.11	.29	.53	.05	.01 (2)	.619
36. Tapuruquara	322	.02	.10	.39	.43	.06	—	.643
38. Tefe	172	—	.11	.34	.42	.13	—	.678
44. Rio Preto	762	.02	.10	.33	.45	.09	.003(1)	.632
45. Mirassol	1,806	.01	.07	.25	.57	.09	.002(3)	.618
Other localities	107	—	.13	.29	.49	.09	—	.629
Total	7,597	.013	.094	.291	.501	.095	.007(4)	

* The number of different alleles included in the "*other* alleles" category in each locality as well as in the whole species is given in parenthesis.

from west to east (Table 1). In Jaque and Teresita, in the northwest, the frequency of *Lap-5*$^{1.03}$ is 0.37 ± 0.03 and 0.36 ± 0.03 respectively. (The standard deviations are obtained from the binomial distribution). In Trinidad the frequency of allele 1.03 is 0.56 ± 0.02; in Macapa, in the northeast, is 0.63 ± 0.06, and in Mirassol, in the southeast, 0.57 ± 0.01. The regression of the percent frequency of allele 1.03 on degrees of longitude is −0.52, which is significantly different from zero ($t = 3.78$, $P < 0.01$). There is considerable heterogeneity in the frequency of other alleles at this locus. For instance, the frequency of allele 1.05 is 10 times as large in Mitu, Colombia (0.24 ± 0.03, sample size = 160) as in Santarem, Brazil (0.024 ± 0.006, sample size = 492).

At the *Est-7* locus, the frequency of the predominant allele, 1.00, ranges from 0.39 ± 0.08 in Macapa to 0.66 ± 0.03 in Santa Marta. Allele 1.02 has a frequency more than twice as large in Tapuruquara (0.38 ± 0.03, sample size = 246) than in La Macarena (0.14 ± 0.04, sample size = 86), Valparaiso (0.13 ± 0.03, sample size = 108), Mitu (0.14 ± 0.02, sample size = 224), and Leticia (0.12 ± 0.04, sample size = 82).

Allele 1.12 at the *Adk-1* locus reaches frequencies of 0.47 ± 0.06 in Guyana and in the small sample from south Brazil, while elsewhere it has a frequency of about 0.15. This locus provides an exception to our first generalization. Allele 1.06 is the most frequent at most, but not all, localities.

2. *Loci with 10 to 50 percent heterozygotes.* At the *Est-6* locus (Table 4) the frequency of allele 1.00 is higher, and that of allele 1.04 lower, in the western than in the central and eastern populations. The regression of the percent frequency of allele 1.00 on degrees of longtitude is 0.82, which is significantly different from zero ($t = 2.58$, $P < 0.05$). Allele 1.08 occurs in Santarem and Tapuruquara at a 0.02 frequency (sample size = 228 and 212, respectively) but it is absent or at very low frequencies in all other populations from which large samples were studied.

Considerable heterogeneity between populations exists at the *Adh* locus (Table 8). In Manaus and Tapuruquara this locus is nearly fixed for allele 1.00 (the frequencies are 0.992 ± 0.004 and 0.994 ± 0.004, respectively). In Trinidad the frequency of this allele is 0.89 ± 0.03. $Adh^{.98}$ has an average frequency smaller than 0.002 in the central Amazon region and south Brazil in spite of very large samples (2430 genomes in total), but has a frequency about 50 times greater, 0.09 ± 0.02, in Trinidad, and was repeatedly found in Jaque, Belem, and some other localities.

3. *Loci with five to ten percent heterozygous individuals.* Like *Adh*, *Mdh-2* has alleles which are endemic to certain localities (Table 8). Allele 1.06 reaches 0.05 ± 0.01 frequency in the southeast of Caracas, but is absent in Manaus (sample size = 490) and Santarem (sample size = 486). Locally endemic alleles exist also at the *Idh*, *Me-1* and *To* loci.

4. *Loci with less than five percent heterozygous individuals.* These loci are nearly fixed at many localities, but again we find some localities with indigenous polymorphisms. For instance, $\alpha Gpdh^{1.06}$ has a frequency of 0.05 ± 0.01 in Caracas, but is very rare or absent in the large samples from the Amazon region.

DISCUSSION

The results of our survey of allozyme variation in *D. willistoni* are summarized in Tables 18 and 19. In Table 18 we list all populations in which several loci were studied with an average sample of at least eight genomes per locus. For each locality we give the proportion of polymorphic loci, and the proportion of the genome at which an average individual is expected to be heterozygous. These values should be used advisedly. The number of loci studied varies considerably from population to population. The proportions of polymorphism and heterozygosity for a given population depend on what genes have been surveyed. For summary purposes we have listed in Table 18 all populations, but for comparisons among populations one can only use those populations in which a large number of loci, say 15 or more, have been studied.

Table 18 shows that the proportion of the genome heterozygous in an average individual varies only moderately from locality to locality. Of the 15 localities

TABLE 2

Frequencies of alleles at the Est-2 *and* Est-3 *loci of* D. willistoni

Locality	Est-2 Genes sampled	.98	1.00	1.02	1.04	Other*	Proportion of individuals heterozygous	Est-3 Genes sampled	.98	1.00	1.02	Other*	Proportion of individuals heterozygous
5. Tortuguero	36	—	.03	.97	—	—	.054	24	—	.96	.04	—	.080
20. Trinidad	74	—	.03	.96	.01	—	.079	—	—	—	—	—	—
21. Caripe	44	—	.05	.95	—	—	.087	—	—	—	—	—	—
22. SE of Caracas	248	.01	.06	.92	.01	—	.133	—	—	—	—	—	—
24. Ayacucho	150	—	.03	.96	.01	—	.077	—	—	—	—	—	—
25. Ocamo	26	—	.04	.92	.04	—	.145	—	—	—	—	—	—
30. Guyana	558	.01	.02	.95	.01	.005(2)	.123	242	.02	.95	.02	.004(1)	.096
33. Belem	20	—	.10	.90	—	—	.180	62	.02	.94	.03	.02 (1)	.123
34. Santarem	82	.01	.06	.91	.01	—	.159	12	—	1.00	—	—	.000
35. Manaus	204	—	.08	.91	.01	—	.171	84	.01	.93	.06	—	.134
36. Tapuruquara	106	.02	.06	.92	—	—	.132	42	.02	.93	.05	—	.131
38. Tefe	70	—	.09	.90	.01	—	.182	20	.05	.90	.05	—	.185
44. Rio Preto	760	.01	.03	.94	.02	—	.114	460	.03	.93	.02	.01 (1)	.104
45. Mirassol	1,916	.01	.04	.93	.02	.001(1)	.136	1,576	.03	.94	.03	.001(1)	.125
Other localities	94	.01	.07	.90	.01	—	.057	52	.02	.92	.06	—	.117
Total	4,342	.009	.041	.933	.017	.001(2)		2,574	.026	.939	.031	.004(2)	

* See Table 1.

TABLE 3

Frequencies of alleles at the Est-4 *and* Est-5 *loci of* D. willistoni

Locality	Est-4 Genes sampled	*.98*	*1.00*	*1.02*	*1.04*	*Other**	Proportion of individuals heterozygous	Est-5 Genes sampled	*.95*	*1.00*	*1.05*	*Other**	Proportion of individuals heterozygous
5. Tortuguero	40	—	.25	.75	—	—	.375	38	.03	.97	—	—	.051
6. Jaque	192	—	.21	.78	.01	—	.352	188	.01	.98	.01	—	.032
9. P. Rico	12	—	.33	.67	—	—	.444	12	—	1.0	—	—	.000
14. Teresita	230	.004	.15	.83	.01	—	.287	238	.02	.97	.01	—	.058
16. La Macarena	88	—	.14	.85	.01	—	.255	84	.01	.96	.02	—	.069
17. Valparaiso	108	.03	.25	.71	.01	—	.449	108	.07	.91	.02	—	.184
18. Mitu	218	.005	.15	.84	.005	—	.279	224	.04	.94	.02	—	.115
19. Leticia	82	—	.22	.78	—	—	.330	82	.01	.96	.02	—	.071
20. Trinidad	1,062	.01	.23	.75	.005	—	.322	984	.03	.96	.01	—	.094
21. Caripe	20	—	.15	.85	—	—	.255	60	.03	.95	.02	—	.096
22. SE of Caracas	136	.01	.29	.71	—	—	.390	212	.02	.96	.02	—	.078
23. Rancho Grande	—	—	—	—	—	—	—	40	.07	.88	.05	—	.226
24. Ayacucho	88	—	.10	.90	—	—	.180	390	.05	.95	—	—	.093
25. Ocamo	—	—	—	—	—	—	—	90	—	.99	.01	—	.022
30. Guyana	600	.005	.09	.90	.003	.002(1)	.170	572	.04	.95	.01	.01 (2)	.078
32. Macapa	34	—	.24	.71	.06	—	.443	48	.02	.96	.02	—	.081
33. Belem	74	—	.38	.61	.01	—	.487	58	.02	.98	—	—	.034
34. Santarem	472	.004	.29	.71	—	—	.416	444	.03	.96	.01	—	.079
35. Manaus	422	.002	.19	.79	.01	—	.333	408	.02	.97	.01	—	.062
36. Tapuruquara	328	.01	.20	.77	.01	.003(1)	.366	308	.02	.96	.02	—	.066
38. Tefe	176	.01	.20	.79	—	—	.336	174	.03	.96	.01	.01 (1)	.078
44. Rio Preto	680	.01	.13	.85	.001	—	.256	890	.05	.94	.01	—	.083
45. Mirassol	1,812	.02	.10	.82	.06	.001(2)	.300	1,976	.03	.96	.01	—	.086
Other localities	84	—	.19	.80	.01	—	.335	88	.05	.95	—	—	.068
Total	6,958	.011	.169	.801	.019	.0004(2)		7,696	.032	.957	.011	.0004(3)	

* See Table 1.

TABLE 4

Frequencies of alleles at the Est-6 *and* Est-7 *loci of* D. willistoni

Locality	Est-6 Genes sampled	*1.00*	*1.04*	*1.08*	*Other**	Proportion of individuals *heterozygous*	Est-7 Genes sampled	*.96*	*.98*	*1.00*	*1.02*	*1.05*	*Other**	Proportion of individuals heterozygous
5. Tortuguero	8	1.00	—	—	—	.000	36	—	.06	.53	.39	.03	—	.566
6. Jaque	142	.94	.06	—	.01 (1)	.120	174	.03	.27	.46	.20	.03	—	.673
9. P. Rico	12	.83	.17	—	—	.278	12	—	.25	.50	.25	—	—	.625
11. Santa Marta	—	—	—	—	—	—	214	.02	.13	.66	.17	.01	.005(1)	.482
14. Teresita	158	.88	.11	—	.01 (1)	.213	216	.02	.20	.45	.30	.03	—	.669
16. La Macarena	84	.90	.10	—	—	.172	86	.07	.24	.50	.14	.03	.01 (1)	.666
17. Valparaiso	108	.90	.10	—	—	.144	108	.09	.19	.54	.13	.05	—	.605
18. Mitu	224	.88	.12	.004	.004(1)	.224	224	.06	.25	.50	.14	.05	.004(1)	.658
19. Leticia	82	.84	.16	—	—	.264	82	.04	.23	.59	.12	.02	—	.579
20. Trinidad	606	.87	.13	—	.005(1)	.191	910	.03	.20	.53	.20	.05	.002(2)	.628
30. Guyana	180	.52	.48	—	.006(1)	.475	442	.02	.19	.50	.22	.07	.002(1)	.637
32. Macapa	16	.75	.19	—	.06 (1)	.398	41	—	.12	.39	.29	.20	—	.709
33. Belem	52	.75	.19	—	.05 (2)	.399	61	—	.10	.64	.16	.10	—	.545
34. Santarem	228	.69	.28	.02	.01 (2)	.446	406	.01	.15	.55	.21	.07	—	.620
35. Manaus	158	.73	.22	—	.05 (2)	.413	358	.01	.09	.59	.22	.09	—	.585
36. Tapuruquara	212	.61	.35	.02	.01 (3)	.477	246	.004	.06	.45	.38	.11	.004(1)	.635
38. Tefe	92	.60	.36	—	.04 (2)	.512	147	.02	.12	.45	.31	.11	—	.679
45. Mirassol	—	—	—	—	—	—	920	.01	.13	.57	.26	.01	.02 (4)	.586
Other localities	56	.66	.27	.07	—	.219	97	—	.08	.60	.25	.06	.01 (1)	.533
Total	2,418	.783	.196	.006	.014(4)		4,780	.020	.161	.536	.227	.050	.005(5)	

* See Table 1.

TABLE 5

Frequencies of alleles at the Aph-1 *locus of* D. willistoni

Locality	Genes sampled	Alleles					Proportion of individuals heterozygous
		.98	*1.00*	*1.02*	*1.04*	*Other**	
6. Jaque	286	.01	.83	.09	.06	.01 (2)	.295
9. P. Rico	12	—	.75	.17	.08	—	.403
14. Teresita	243	.02	.88	.07	.03	—	.218
15. Malaga	76	.01	.80	.13	.05	—	.336
16. La Macarena	186	.02	.90	.04	.03	—	.181
17. Valparaiso	206	.01	.88	.04	.07	—	.222
18. Mitu	328	.01	.79	.12	.08	—	.350
19. Leticia	154	.03	.80	.10	.06	—	.340
20. Trinidad	962	.03	.79	.10	.07	.01 (2)	.315
30. Guyana	353	.04	.82	.05	.09	—	.313
32. Macapa	38	.05	.84	—	.11	—	.277
33. Belem	58	.03	.84	.02	.10	—	.274
34. Santarem	338	.05	.87	.02	.05	.01 (2)	.233
35. Manaus	326	.02	.89	.08	.01	.003 (1)	.196
36. Tapuruquara	374	.02	.89	.08	.01	—	.194
38. Tefe	212	.02	.85	.11	.03	—	.265
Other localities	79	—	.89	.08	.04	—	.172
Total	4,231	.02	.84	.08	.06	.003 (3)	

* See Table 1.

TABLE 6

Frequencies of alleles at the Acph-1 *locus of* D .willistoni

Locality	Genes sampled	Alleles				Proportion of individuals heterozygous
		.84	*.94*	*1.00*	*1.04*	
5. Tortuguero	36	—	.06	.92	.03	.156
30. Guyana	80	—	.06	.94	—	.117
Other localities	208	.010	.048	.918	.024	.121
Total	324	.006	.052	.923	.018	

TABLE 7

Frequencies of alleles at the Ald *locus of* D. willistoni

Locality	Genes sampled	Alleles						Proportion of individuals heterozygous
		.96	*.98*	*1.00*	*1.02*	*1.06*	*1.12*	
20. Trinidad	42	—	—	.98	.02	—	—	.043
30. Guyana	154	.03	.01	.88	.08	.01	—	.213
45. Mirassol	96	—	.04	.86	.07	.01	.01	.245
Other localities	112	.01	.08	.80	.10	.01	—	.089
Total	404	.012	.035	.866	.077	.007	.002	

TABLE 8

Frequencies of alleles at the Adh *and* Mdh-2 *loci of* D. willistoni

Locality	Adh						Mdh-2						
	Genes sampled	*.90*	*.98*	*1.00*	*1.06*	*Other**	Proportion of individuals heterozygous	Genes sampled	*.86*	*.94*	*1.00*	*1.06*	*Other**	Proportion of individuals heterozygous
6. Jaque	48	.05	.06	.90	—	—	.191	50	—	—	.98	.02	—	.039
9. P. Rico	10	—	—	1.00	—	—	.000	10	—	—	1.00	—	—	.000
14. Teresita	48	—	—	.98	—	.02 (1)	.041	48	—	—	1.00	—	—	.000
20. Trinidad	132	.02	.09	.89	—	—	.195	132	—	—	.99	.01	—	.010
21. Caripe	48	—	.02	.98	—	—	.041	48	—	—	1.00	—	—	.000
22. SE of Caracas	222	—	.03	.96	.01	—	.115	220	—	.02	.93	.05	.005(1)	.168
30. Guyana	508	.03	.002	.95	.01	.004(2)	.102	584	—	.02	.96	.02	—	.129
32. Macapa	54	.04	.02	.94	—	—	.106	54	—	—	1.00	—	—	.000
33. Belem	74	—	.08	.92	—	—	.149	74	.03	.03	.93	.01	—	.129
34. Santarem	498	.03	—	.97	—	.002(1)	.055	486	.004	.002	.99	—	.002(1)	.016
35. Manaus	470	.002	.006	.99	—	—	.017	490	—	.01	.99	—	—	.020
36. Tapuruquara	340	.006	—	.99	—	—	.010	340	—	.01	.99	.01	—	.024
38. Tefe	176	.05	.006	.94	—	.006(1)	.055	178	—	.03	.93	.02	.02 (2)	.095
45. Mirassol	946	.01	—	.98	.01	.001(1)	.046	1,024	.001	.03	.95	.01	.004(2)	.090
Other localities	94	.01	.08	.90	—	.02 (1)	.139	102	—	.03	.96	—	.01 (1)	.024
Total	3,670	.017	.012	.965	.002	.004(2)		3,840	.001	.018	.967	.011	.003(3)	

* See Table 1.

TABLE 9

Frequencies of alleles at the αGpdh *locus of* D. willistoni

Locality	Genes sampled	Alleles				Proportion of individuals heterozygous
		.94	*1.00*	*1.06*	*Other**	
5. Tortuguero	36	—	1.00	—	—	.000
6. Jaque	48	—	1.00	—	—	.000
9. P. Rico	10	—	1.00	—	—	.000
14. Teresita	48	—	1.00	—	—	.000
20. Trinidad	132	.02	.98	—	—	.032
21. Caripe	48	—	.98	.02	—	.041
22. SE of Caracas	222	.01	.95	.05	—	.105
30. Guyana	580	.01	.99	.003	.002(1)	.014
32. Macapa	54	—	1.00	—	—	.000
33. Belem	74	.01	.99	—	—	.027
34. Santarem	498	.002	.996	—	.002(1)	.008
35. Manaus	468	.01	.99	—	—	.021
36. Tapuruquara	340	.003	.99	.01	.003(1)	.037
38. Tefe	174	—	1.00	—	—	.000
45. Mirassol	1,102	.004	.98	.02	.004(2)	.046
Other localities	84	—	.99	.01	—	.090
Total	3,908	.005	.984	.009	.002(3)	

* See Table 1.

in which at least 15 loci were studied, Manaus has the lowest value, 0.144, and Malaga has the highest, 0.237. In 12 of the 15 populations, the proportion of heterozygous loci is 0.18–0.21. This contrasts strikingly with the considerable heterogeneity among localities observed in the amount of chromosomal polymorphism. DA CUNHA *et al.* (1959) found that the average number of heterozygous inversions per individual was in some localities 30 times greater than in other localities. The allozyme variation does not parallel the chromosomal variation.

In *D. willistoni*, the variation from population to population in the degree of chromosomal polymorphism follows certain rules (DA CUNHA *et al.* 1959; DOBZHANSKY 1963; AYALA *et al.* 1970). Populations of the central portion of the

TABLE 10

Frequencies of alleles at the G3pdh *locus of* D. willistoni

Locality	Genes sampled	Alleles			Proportion of individuals heterozygous
		.96	*1.00*	*1.05*	
6. Jaque	48	.10	.85	.04	.258
14. Teresita	46	—	.98	.02	.043
20. Trinidad	24	.08	.88	.04	.233
Other localities	72	.04	.92	.04	.059
Total	190	.053	.911	.037	

TABLE 11

Frequencies of alleles at the Idh *and* Me-1 *loci of* D. willistoni

Locality	*Idh*					*Me-1*				
	Genes sampled	*.96*	*1.00*	*1.04*	Proportion of individuals heterozygous	Genes sampled	*.98*	*1.00*	*1.04*	Proportion of individuals heterozygous
5. Tortuguero	36	—	.97	.03	.054	36	.11	.89	—	.197
6. Jaque	44	—	1.00	—	.000	40	.025	.975	—	.049
9. P. Rico	10	—	1.00	—	.000	12	—	1.00	—	.000
14. Teresita	48	—	.96	.04	.080	36	—	1.00	—	.000
20. Trinidad	118	.03	.94	.03	.094	96	.03	.97	—	.029
30. Guyana	354	.003	.98	.02	.039	282	.03	.94	.03	.108
45. Mirassol	96	—	.98	.02	.041	96	.01	.97	.02	.061
31–33. NE Brazil	46	.02	.96	.02	.076	40	.05	.95	—	.125
34–38. Amazonas	82	—	.99	.01	.019	80	.01	.99	—	.044
Other localities	72	—	1.00	—	.000	70	.04	.87	.09	.188
Total	906	.006	.976	.019		788	.029	.948	.023	

TABLE 12

Frequencies of alleles at the Me-2 *locus of* D. willistoni

Locality	Genes sampled	Alleles			Proportion of individuals heterozygous
		.96	*1.00*	*1.04*	
5. Tortuguero	36	—	.86	.14	.239
30. Guyana	72	—	.97	.03	.054
Other localities	160	.056	.675	.269	.308
Total	268	.034	.780	.187	

TABLE 13

Frequencies of alleles at the To *locus of* D. willistoni

Locality	Genes sampled	Alleles					Proportion of individuals heterozygous
		.98	*1.00*	*1.02*	*1.12*	*Other**	
5. Tortuguero	36	—	1.00	—	—	—	.000
30. Guyana	444	.005	.97	.01	.01	.005(2)	.037
31. Tracajatuba	20	—	.85	.05	.05	.05 (1)	.270
32. Macapa	44	—	.93	.02	.05	—	.129
33. Belem	80	.01	.96	—	.01	.01 (1)	.073
34. Santarem	244	—	.98	.02	—	—	.032
35. Manaus	118	—	.97	.03	—	—	.050
36. Tapuruquara	318	.003	.98	.02	—	.003(1)	.052
38. Tefe	145	—	.98	.01	—	.01 (2)	.041
45. Mirassol	1,038	.004	.995	.001	—	—	.010
Other localities	42	—	1.00	—	—	—	
Total	2,529	.003	.982	.009	.004	.002(3)	

* See Table 1.

TABLE 14

Frequencies of alleles at the Tpi-2 *locus of* D. willistoni

Locality	Genes sampled	Alleles			Proportion of individuals heterozygous
		.94	*1.00*	*1.06*	
5. Tortuguero	24	.04	.96	—	.080
6. Jaque	44	—	.98	.02	.044
14. Teresita	42	—	1.00	—	.000
20. Trinidad	94	—	.98	.02	.022
30. Guyana	208	.005	.99	.005	.015
31–33. NE Brazil	32	—	1.00	—	.000
34–38. Amazonas	72	—	1.00	—	.000
45. Mirassol	64	—	.92	.08	.144
Other localities	70	—	1.00	—	.000
Total	652	.003	.983	.014	

TABLE 15

Frequencies of alleles at the Adk-1 *and* Adk-2 *loci of* D. willistoni

Locality	*Adk-1* Genes sampled	*1.00*	*1.06*	*1.12*	*Other**	Proportion of individuals heterozygous	*Adk-2* Genes sampled	*.96*	*.98*	*1.00*	*1.02*	*Other**	Proportion of individuals heterozygous
5. Tortuguero	34	.21	.65	.15	—	.517	36	.03	.06	.78	.14	—	.372
6. Teresita	44	.34	.55	.11	—	.573	42	.02	.10	.88	—	—	.214
20. Trinidad	98	.22	.53	.18	.06 (2)	.609	66	—	.08	.91	.02	—	.170
30. Guyana	66	.17	.36	.47	—	.619	254	.02	.05	.83	.10	.004(1)	.330
31–33. NE Brazil	48	.33	.48	.15	.04 (1)	.565	44	—	.05	.95	—	—	.073
34–38. Amazonas	74	.20	.64	.14	.03 (1)	.520	72	—	.06	.94	—	—	.072
45. Mirassol	12	.08	.42	.50	—	.569	112	—	.03	.93	.04	—	.135
Other localities	54	.39	.37	.17	.07 (2)	.517	74	—	.03	.95	—	.03 (2)	.160
Total	430	.251	.505	.212	.033(3)		700	.009	.050	.884	.053	.004(3)	

* See Table 1.

TABLE 16

Frequencies of alleles at the Pgm-1 *and* Hk-1 *loci of* D. willistoni

Locality	Genes sampled	*Pgm-1* .96	*1.00*	*1.04*	*1.12*	Proportion of individuals heterozygous	Genes sampled	*Hk-1* .96	*1.00*	*Other**	Proportion of individuals heterozygous
6. Jaque	42	—	.95	.05	—	.091	42	.02	.98	—	.047
9. P. Rico	10	—	.90	.10	—	.180	12	—	1.00	—	.000
14. Teresita	36	—	.92	.08	—	.153	38	—	.97	.03 (1)	.051
20. Trinidad	90	.04	.92	.01	.02	.149	86	.02	.97	.01 (1)	.034
30. Guyana	190	.06	.79	.15	—	.343	132	.07	.92	.01 (1)	.141
31–33. NE Brazil	12	—	.83	.17	—	.320	34	.15	.82	.03 (1)	.262
34–38. Amazonas	60	.03	.93	.03	—	.090	80	—	1.00	—	.000
45. Mirassol	40	.05	.875	.075	—	.226	—	—	—	—	—
Other localities	12	—	1.00	—	—	.000	40	—	.98	.02 (1)	.000
Total	492	.039	.872	.085	.004		464	.037	.953	.011 (2)	

* See Table 1.

TABLE 17

Frequencies of alleles at the Hk-2 *and* Hk-3 *loci of* D. willistoni

Locality	*Hk-2*					*Hk-3*					
	Genes sampled	*.96*	*1.00*	*1.04*	*1.12*	Proportion of individuals heterozygous	Genes sampled	*.96*	*1.00*	*1.04*	*1.12*	Proportion of individuals heterozygous
6. Jaque	42	—	1.00	—	—	.000	42	—	1.00	—	—	.000
9. P. Rico	12	—	.92	.08	—	.153	12	.08	.83	.08	—	.292
14. Teresita	38	—	.97	—	.03	.051	38	.03	.97	—	—	.051
20. Trinidad	102	—	.97	.02	.01	.045	106	—	.99	—	.01	.012
30. Guyana	130	.02	.87	.08	.02	.236	142	.01	.96	.01	.01	.069
31–33. NE Brazil	40	—	.90	.03	.07	.181	38	—	1.00	—	—	.000
34–38. Amazonas	80	—	.79	.09	.12	.329	58	—	1.00	—	—	.000
Other localities	50	—	1.00	—	—	.000	50	—	1.00	—	—	.000
Total	494	.006	.913	.045	.036		486	.008	.982	.006	.004	

TABLE 18

Number of loci studied for each of 26 populations of D. willistoni, *proportion of those loci which are polymorphic in each population, and the proportion of the genome estimated to be heterozygous in an average individual*

Locality	Number of loci studied	Average sample size per locus	Proportion of loci polymorphic*		Proportion of genome heterozygous per individual
			Criterion 1	Criterion 2	
5. Tortuguero	18	32	.555	.833	.205
6. Jaque	18	93	.444	.778	.184
9. Puerto Rico	15	11	.533	.533	.196
14. Teresita	23	91	.478	.826	.189
15. Malaga	19	15	.684	.684	.237
16. La Macarena	9	86	.889	1.00	.324
17. Valparaiso	9	92	.778	1.00	.293
18. Mitu	8	175	.875	.875	.305
19. Leticia	6	91	.833	1.00	.375
20. Trinidad	23	305	.478	.957	.177
21. Caripe	9	87	.556	.889	.215
22. SE of Caracas	7	225	.714	1.00	.229
23. Rancho Grande	3	25	.667	.667	.276
24. Ayacucho	4	234	.750	1.00	.242
25. Ocamo	3	65	.667	1.00	.272
30. Guyana	23	313	.696	1.00	.195
31. Tracajatuba	22	16	.682	.727	.190
32. Macapa	18	29	.556	.611	.200
33. Belem	20	44	.650	.800	.214
34. Santarem	19	218	.474	.632	.189
35. Manaus	24	166	.417	.583	.144
36. Tapuruquara	21	136	.476	.810	.193
38. Tefe	21	87	.571	.667	.209
44. Rio Preto	5	710	1.00	1.00	.241
45. Mirassol	17	860	.588	.941	.194
46. Riberao Preto	4	8	.500	.500	.180

* Criterion 1: the frequency of the most common allele is $\leq .95$;
Criterion 2: the frequency of the most common allele is $\leq .99$.

distribution area are more polymorphic than peripheral populations. Island populations and those of distributional pockets are less polymorphic than continental populations. For instance, the average numbers of heterozygous inversions per female are: Puerto Rico, 1.2; Trinidad, 2.3; Panama, 6.2; northern Colombia, 6.7; Caripe, Venezuela, 5.8. It is clear that the allozyme polymorphisms do not obey the same rules as the chromosomal polymorphisms. The proportion of heterozygous loci per individual is 0.177 in Trinidad, somewhat lower than 0.184, the average for the whole species. For Puerto Rico, an island smaller than Trinidad, the value is 0.196. Manaus in the central Amazonian forests has the lowest proportion of heterozygosity, but in the same region Tefe and Tapuruquara have values well within the norm. The Macapa and Tracajatuba collections were made in small patches of forest (about one square mile in the case of Macapa) isolated

TABLE 19

Proportion of populations polymorphic and the average proportion of individuals estimated to be heterozygous at each of 28 loci of D. willistoni

Gene	Genes sampled	Samples	Proportion of populations which are polymorphic*		Average proportion of individuals heterozygous at each locus*
			Criterion	Criterion 2	
Lap-5	7,597	70	1.00	1.00	.642±.008
Est-2	4,342	51	.733	1.00	.130±.010
Est-3	2,574	35	.875	1.00	.121±.012
Est-4	6,958	68	1.00	1.00	.318±.015
Est-5	7,696	71	.304	1.00	.086±.006
Est-6	2,418	48	1.00	1.00	.285±.030
Est-7	4,780	57	1.00	1.00	.622±.009
Aph-1	4,231	55	1.00	1.00	.262±.018
Acph-1	324	14	.857	1.00	.125±.013
Acph-2	268	26	.833	1.00	.145±.037
Ald	404	41	.571	.714	.135±.048
Adh	3,670	46	.467	.867	.106±.021
Mdh-2	3,840	48	.267	.667	.055±.013
αGpdh	3,908	47	.067	.533	.027±.008
Idh	906	53	.125	.750	.062±.018
G3pdh	190	27	.500	1.00	.148±.053
Odh-1	226	20	.714	.857	.140±.035
Odh-2	40	4	.333	.333	.060±.024
Me-1	788	51	.250	.750	.087±.024
Me-2	268	5	.600	1.00	.244±.091
To	2,529	12	.111	.778	.065±.021
Tpi-2	652	49	.125	.600	.040±.016
Pgm-1	492	18	.857	1.00	.180±.034
Adk-1	430	30	1.00	1.00	.545±.025
Adk-2	700	50	1.00	1.00	.247±.035
Hk-1	464	36	.167	.667	.049±.021
Hk-2	494	39	.500	.833	.187±.076
Hk-3	486	38	.000	.750	.031±.014
Mean proportion of loci polymorphic per population:					
		Criterion 1	0.581		
		Criterion 2	0.861		
Estimated proportion of genome heterozygous in an average individual 0.184					

* Only samples with at least 20 genes have been used for these estimates, except for *Ald*, *G3pdh*, *Odh-1* and *Odh-2* for which all samples of 10 or more genes have been used. Criteria 1 and 2 as in Table 18.

by many tens of kilometers from any other forests. Yet the proportion of heterozygosity per individual, 0.200 and 0.190 respectively, is well within the range observed in central populations.

In Table 19 we give for each of 28 loci, the number of genes and of populations sampled; the proportion of populations which are polymorphic and the proportion of individuals which are heterozygous. The average over all loci of the proportion of polymorphic populations estimates the proportion of polymorphic loci per population. We have used two criteria of polymorphism, depending on

whether the frequency of the most common allele in a population is (1) no greater than 0.95 or (2) no greater than 0.99. Fifty-eight point one percent loci are polymorphic by the first criterion, and 86.1% by the second criterion.

The average over all loci of the proportion of individuals heterozygous per locus estimates the proportion of genes at which an individual is heterozygous on the average. This statistic is, for *D. willistoni*, 0.184. In *D. paulistorum*, a sibling species of *D. willistoni*, RICHMOND (1972) has estimated that an individual is heterozygous at 0.21 of its loci. The amount of heterozygosity per individual observed in *D. willistoni* is on the higher side of the values observed for other Drosophila species, which range from 0.08 to 0.25. Estimates have been recently obtained, mostly on the basis of gel electrophoresis studies, of the amount of genetic variation in other animals, like man, field and house mouse, and horseshoe crab. The proportion of heterozygous loci per individual in these diverse organisms falls between 0.05 and 0.16. (Summary in SELANDER *et al.* 1970). It is necessary to point out that these estimates are only gross approximations, since the biochemical techniques used, as well as the sampling procedures, are subject to various biases (LEWONTIN and HUBBY 1966; AYALA *et al.* 1970) which may affect the estimates in an unpredictable manner as to direction and amount. Yet the estimates are probably correct as to the order of magnitude. It is, then, remarkable that the amount of genetic variation in animals as different as arthropods and mammals should be so similar. It will be most interesting to find out whether other sexually reproducing animals contain similarly large amounts of genetic variation.

What are the biological processes responsible for the considerable genetic variability present in these organisms? Any hypothesis concerning the mechanisms which account for the high levels of genetic variation observed in *D. willistoni* must explain the facts encompassed by the two generalizations stated above (see RESULTS). It must explain the remarkable similarity in the configuration of the allelic frequencies which obtains throughout the geographic distribution of the species, without ignoring the fact that differences between localities do exist.

It has been recently suggested that most of the genetic variability existing in natural populations, particularly that variability which is observed at the molecular level may be adaptively neutral (KIMURA 1968; KING and JUKES 1969). The proponents of this hypothesis agree that a large fraction of all new mutations are adaptively harmful. These mutants are eliminated or kept at very low frequencies by natural selection. They argue, however, that a substantial fraction of all new mutants are selectively equivalent. The frequencies of these adaptively neutral alleles are governed not by natural selection, but by the random process of sampling errors which occur every generation in any finite population. This process has been termed "non-Darwinian evolution" (KING and JUKES 1969), or more appropriately, "evolution by random walk" (DOBZHANSKY 1970).

The theory of evolution by random walk leads to certain predictions as to the *amount* of neutral genetic variation expected in a given population and as to the *pattern* of this variation. The theory can, therefore, be tested by checking these

predictions against the amounts and patterns of genetic variation observed in natural populations. The amount of neutral genetic variation expected in a population can be predicted in terms of two parameters, namely the effective size of the population and the rate of mutation to neutral alleles. Unfortunately, these quantities are in most cases unknown even as to the order of magnitude. Calculations of the expected amount of neutral genetic variation have, therefore, an enormous latitude which severely restricts their value. If N is the effective size of a population, and u is the rate of mutation to neutral alleles per generation, the effective number of neutral alleles, n, segregating in a population is (KIMURA and CROW 1964):

$$n = 4Nu + 1 \quad (1)$$

It is clear that by postulating that N and u are of the appropriate magnitude *any* given number of segregating alleles can be accounted for by equation (1). If the effective size of the population is approximately one tenth of the reciprocal of the mutation rate, the effective number of alleles per locus will be 1.4 and the heterozygosity about 29 percent. This is approximately what is observed at many loci of *D. willistoni.*

The *pattern* of the variation, however, does not agree with the predictions of the theory. If as postulated by the theory of random walk allelic frequencies are the result of sampling errors, the allelic frequencies should be completely uncorrelated among populations. That is, *different neutral alleles should occur in different populations*; or whenever the same alleles are found in two or more different populations *they should occur in frequencies completely uncorrelated.* This expectation stands in sharp contrast with the facts. As shown in Tables 1 through 17 at every locus studied there is a remarkable similarity in the configuration of allelic frequencies from population to population throughout the whole species.

KIMURA and OHTA (1971) have suggested a conceivable escape from this difficulty. If a certain amount of migration occurs between neighboring populations, the species may effectively approximate a single panmictic population. Local populations would effectively represent samples of a single population. Similar allelic frequencies would then occur in all populations interconnected by migration. The proportion of polymorphic loci per population and of heterozygosity per individual observed in man, Drosophila and mouse and the similarity of the pattern of variation can be explained according to KIMURA and OHTA (1971) if the following two conditions obtain:

$$Nm > 4 \quad (2)$$

where N is the effective size of the *local* population and m is the effective rate of migration per generation between neighboring populations; and

$$4N_e u \approx 0.1 \quad (3)$$

where N_e is the effective size of the *species or subspecies*, and u is the mutation rate as in (1). If (3) is correct, the average number of segregating alleles per locus will be 1.1. This value is below the actual average of the effective number of alleles per locus segregating in *D. willistoni.* Is equation (3) likely to be approximately correct? It is difficult to estimate even approximately the effective size of the species, *D. willistoni.* The geographic distribution of the species encom-

passes several million square kilometers. Throughout this enormous territory, *D. willistoni* is often the most abundant drosophilid in tropical forests. A low estimate of the number of breeding *D. willistoni* flies in any given generation is 10^9 (AYALA 1972). The mutation rate to neutral alleles probably lies between 10^{-5} and 10^{-7}. Taking the lower estimate, 10^{-7}, of the mutation rate and substituting it in (3) together with the minimum estimate, 10^9, of N_e, we obtain

$$4 N_e u = 400,$$

which is 4000 times greater than the value suggested by KIMURA and OHTA. If there were enough migration in *D. willistoni* so that the species would approximate a single panmictic population we would expect at each locus several hundred alleles each in very low frequency. We find instead a small number of alleles at high and intermediate frequencies.

The recourse to migration to explain the similarity of allelic frequencies from population to population encounters other serious difficulties (AYALA 1972). We have found that populations geographically isolated from the rest of the species nevertheless exhibit configurations of allelic frequencies remarkably similar to each other and to the rest of the species. Two obvious examples are the Puerto Rico and Macapa populations. Our samples from Puerto Rico are small, but they are sufficient to show that at each of the 15 loci studied this population has a similar pattern of genetic variation as that found in the rest of the species. The Macapa population lives in a small patch of forest (about one square mile) surrounded by several thousand square kilometers of semidesertic savanna. The nearest forest is more than one hundred kilometers away.

AYALA, POWELL and DOBZHANSKY (1971) have recently studied allelic variation at 24 enzyme loci of *D. willistoni* in six small oceanic islands, Martinique, St. Lucia, St. Vincent, Bequia, Carriacou, and Grenada. Those islands belong to the Windward group of the Lesser Antilles. Bequia and Carriacou have an area about 10 square kilometers. The islands are off the hurricane belt and are separated from each other and from the mainland by sizeable bodies of water. Differences in allelic frequencies among the islands and between the islands and the rest of the species do exist. But generally at every locus studied the same alleles are found at high, intermediate, and low frequencies in these islands that are found in the continental populations.

That *D. willistoni* does not approximate a single panmictic population is further confirmed by the study of the chromosomal polymorphisms. In contrast with the enzyme polymorphisms, the chromosomal polymorphisms are very different from one locality to another. As stated above, the amount of chromosomal polymorphism is much greater in some than in other populations. Moreover, there are qualitative differences. At least 50 distinct inversions have been recorded in the chromosomes of *D. willistoni* (DA CUNHA and DOBZHANSKY 1954; DOBZHANSKY 1957; DA CUNHA *et al.* 1959). Only three of these inversions are species wide, while most are endemic to a particular locality or region. AYALA, POWELL and DOBZHANSKY (1971) found that *D. willistoni* flies from the six small Caribbean islands are chromosomally very different from continental populations of the species.

In conclusion, our observations are incompatible with the notion that most of the enzyme variation observed in *D. willistoni* is adaptively neutral. Our results are consistent with the hypothesis that widespread balancing selection is the main factor controlling genetic variation in natural populations of *D. willistoni*. Further studies will be required to ascertain the relative roles played by heterotic, diversifying, frequency dependent and other forms of balancing selection. That diversifying selection is responsible for some of the variation has been shown by POWELL (1971).

LITERATURE CITED

AYALA, F. J., 1972 Darwinian versus non-Darwinian evolution in natural populations of Drosophila. Proc. Sixth Berkeley Symp. Math. Stat. Prob., in press.

AYALA, F. J., C. A. MOURÃO, S. PÉREZ-SALAS, R. RICHMOND and TH. DOBZHANSKY, 1970 Enzyme variability in the *Drosophila willistoni* group, I. Genetic differentiation among sibling species. Proc. Natl. Acad. Sci. U.S. **67**: 225–232.

AYALA, F. J., J. R. POWELL and TH. DOBZHANSKY, 1971 Enzyme variability in the *Drosophila willistoni* group, II. Polymorphisms in continental and island populations of *Drosophila willistoni*. Proc. Natl. Acad. Sci. U.S. **68**: 2480–2483.

BREWER, G. J., 1970 *An Introduction to Isozyme Techniques*. Academic Press, New York.

DA CUNHA, A. B., D. BRNCIC and F. M. SALZANO, 1953 A comparative study of chromosomal polymorphism in certain South American species of Drosophila. Heredity **7**: 193–202.

DA CUNHA, A. B., and TH. DOBZHANSKY, 1954 A further study of chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. Evolution **8**: 119–134.

DA CUNHA, A. B., TH. DOBZHANSKY, O. PAVLOVSKY and B. SPASSKY, 1959 Genetics of natural populations. XXVIII. Supplementary data on the chromosomal polymorphism in *Drosophila willistoni* in relation to its environment. Evolution **13**: 389–404.

DOBZHANSKY, TH., 1950 The chromosomes of *Drosophila willistoni*. J. Heredity **41**: 156–158. ——, 1957 Genetics of natural populations, XXVI. Evolution **11**: 280–293. ——, 1963 Species in Drosophila. Proc. Linn. Soc. Lond. **174**: 1–12. ——, 1970 *Genetics of the Evolutionary Process*, Columbia University, New York.

HARRIS, H., 1970 *The Principles of Human Biochemical Genetics*. Elsevier, New York.

JOHNSON, F. M., C. G. KANAPI, R. H. RICHARDSON, M. R. WHEELER and W. S. STONE, 1966 An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. Proc. Natl. Acad. Sci. U.S. **56**: 119–125.

KASTRITSIS, C. D., 1969 A cytological study on some recently collected strains of *Drosophila paulistorum*. Evolution **23**: 663–675.

KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11**: 247–269.

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

KIMURA, M. and T. OHTA, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229**: 467–469.

KING, J. L. and T. H. JUKES, 1969 Non-Darwinian evolution. Science **164**: 788–798.

KOJIMA, K., J. GILLESPIE and Y. N. TOBARI, 1970 A profile of Drosophila species' enzymes assayed by electrophoresis. I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. Biochem. Genet. **4**: 627–637.

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

POULIK, M. D., 1957 Starch gel electrophoresis in a discontinuous system of buffers. Nature **180**: 1477–1479.

POWELL, J. R., 1971 Genetic polymorphisms in varied environments. *Science* **174**: 1035–1036.

PRAKASH, S., R. C. LEWONTIN and J. L. HUBBY, 1969 A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. Genetics **61**: 841–858.

RICHMOND, R., 1972 Enzyme variability in the *Drosophila willistoni* group. III. Amounts of variability in the superspecies, *D. paulistorum*. Genetics **70**: 87–112.

SELANDER, R. K. and S. Y. YANG, 1969 Protein polymorphism and genic heterozygosity in a wild population of the house mouse (*Mus musculus*). Genetics **63**: 653–667.

SELANDER, R. K., S. Y. YANG, R. C. LEWONTIN and W. E. JOHNSON, 1970 Genetic variation in the horseshoe crab (*Limulus polyphemus*), a phylogenetic "relic". Evolution **24**: 402–414.

SHAW, C. R. and A. L. KOEN, 1968 Starch gel zone electrophoresis of enzymes. In: *Chromatographic and Electrophoretic Techniques*. Vol. 2, 2nd ed. Edited by I. SMITH, John Wiley, New York.

SPASSKY, B., R. C. RICHMOND, S. PÉREZ-SALAS, O. PAVLOVSKY, C. A. MOURÃO, A. S. HUNTER, H. HOENIGSBERG, TH. DOBZHANSKY and F. J. AYALA, 1971 Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. Evolution **25**: 129–143.