UC Irvine

UC Irvine Previously Published Works

Title

Enzyme Variability in the Drosophila willistoni Group, I. Genetic Differentiation Among Sibling Species

Permalink

https://escholarship.org/uc/item/3jn205nb

Journal

Proceedings of the National Academy of Sciences of the United States of America, 67(1)

ISSN

0027-8424

Authors

Ayala, Francisco J
Mourão, Celso A
Pérez-Salas, Santiago
et al.

Publication Date

1970-09-01

DOI

10.1073/pnas.67.1.225

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Proceedings of the National Academy of Sciences
Vol. 67, No. 1, pp. 225–232, September 1970

Enzyme Variability in the *Drosophila willistoni* Group, I. Genetic Differentiation Among Sibling Species

Francisco J. Ayala, Celso A. Mourão†, Santiago Pérez-Salas†, Rollin Richmond†, and Theodosius Dobzhansky

THE ROCKEFELLER UNIVERSITY, NEW YORK, N. Y. 10021

Communicated June 18, 1970

Abstract. We have studied by gel electrophoresis the variability of 14 structural genes in four sibling species, *Drosophila willistoni*, *D. paulistorum*, *D. equinoxialis*, and *D. tropicalis*. Samples of about 30 populations from different parts of the distribution areas of each species were examined. Genetic variants are found at every locus; 67% of the loci are polymorphic, having two or more alleles, the rarer of which has a frequency of 5% or higher.

The gene frequencies are fairly uniform over the distribution area of each species, but considerably different in different species. It is estimated that individuals which belong to the different species differ on the average in somewhat more than one half of their gene loci. The morphological similarity of the four sibling species contrasts with the extensive diversity in their genetic materials.

One of the basic problems of evolutionary genetics, in a sense the cardinal problem, is to ascertain what proportions of gene loci are altered in the evolutionary processes, and particularly in the process of speciation. In outbreeding sexual organisms, a species is an array of Mendelian populations, among which gene exchange can occur without impediments other than geographical separation. Different species are arrays of populations reproductively isolated from each other. A mutational or other genetic change originating in a single or in a small group of individuals of a species can spread, impelled by natural selection, to the whole species. Because of reproductive isolation such a change cannot spread from one species to others, unless it arises in these species independently. Speciation is, then, a highly significant stage of evolutionary differentiation. Fully formed species are discrete and independent units of evolution.

The processes of speciation have been extensively studied and discussed.[1,2] A crucially important question is how much reorganization of the gene pool occurs during the process of species divergence. It has sometimes been claimed that species differ at only a small number of loci; other evolutionists held, on the contrary, that a considerable proportion of the gene pool is altered. (It should be remarked that this question is really separate from that of the number of genes directly involved in the formation of reproductive isolating mechanisms.) Methods of investigation developed in recent years permit a fresh approach to the problems in this field.

Related species have usually been studied to detect differences in their morphological and physiological traits, in ecologically significant parameters, in the

numbers of their chromosomes, and sometimes in the gene arrangements within the chromosomes. A serious limitation of this methodology is that it does not permit one to arrive at any well-founded conclusions regarding the proportions of the genes that have been altered in the process of speciation and the proportions which have remained unchanged. Other difficulties are also hard to overcome. Thus, sibling species, which appear to be common in certain groups of insects and of some other animals and plants, are indistinguishable or nearly so in their morphological traits. By morphological criteria, these species appear to be no more different than individuals, or closely related forms of the same species. And yet they are reproductively isolated. Does their morphological similarity indicate that most of their genes have remained unchanged?

Techniques for the separation of enzymes and other chemical constituents of living bodies by means of gel electrophoresis proved to be powerful tools for genetic studies.[3-5] We wish to report some of the results of the application of these tools to the study of the genetic differences between certain species of *Drosophila*. We have chosen to explore a group of species related to *D. willistoni*. This group contains at least twelve species endemic to the tropics of the New World. Their distribution extends from Mexico and southern Florida, through Central America and the Caribbean islands, to southern Brazil and Argentina. Six of the species are siblings, four of which are sympatric over the greater part of the geographic distribution of the group. One of the siblings, *D. paulistorum*, is, in turn, a complex of six semispecies or incipient species which are partially isolated reproductively from each other.[6] Five stages of increasing evolutionary divergence below the generic level, as estimated by classical methods of investigations, can, then, be studied: (1) more or less widely separated geographic populations of the same species; (2) semispecies or incipient species; (3) sibling species; (4) nonsibling species of the *willistoni* group; and (5) species of different groups of the genus *Drosophila*. We report here a preliminary analysis of our observations concerning mostly the stage (3). We have compared the four widely distributed and largely sympatric sibling species *D. willistoni*, *D. paulistorum*, *D. equinoxialis*, and *D. tropicalis*. The two remaining siblings, *D. insularis* and *D. pavlovskiana*, are excluded from this report since they are narrow endemics.

Materials and Methods. Thirty or more natural populations of each of the four siblings have been sampled from various parts of their distribution areas. Most samples were collected in Brazil, Venezuela, Colombia, and Panama between January 1966 and March 1970. The wild-collected males were used for electrophoresis study upon their arrival in the New York laboratory. The females were allowed to produce progenies in individual cultures, and some of their daughters and sons were then tested. Some small samples kept in stock in the laboratory but collected prior to 1966 were also tested.

Fourteen structural genes coding for enzymes were studied for each species, as follows. Alcohol dehydrogenase (*Adh*), one locus; α-glycerophosphate dehydrogenase (*α-Gpdh*), one locus; malate dehydrogenase (*Mdh*), two loci; tetrazolium oxidase (*To*), one locus; esterase (*Est*), six loci; alkaline phosphatase (*Aph*), two loci; leucine aminopeptidase (*Lap*), one locus. Standard procedures[4,7] for horizontal starch gel electrophoresis and isozyme assay were used, with minor modifications to suit our materials.

Results. A thousand or more genomes of wild flies of each species were

studied for each of about 14 loci. At every locus studied we have found at least two allelic variants—two or more bands of different electrophoretic mobility that segregated as Mendelian factors in simple genetic tests. That genetic variation was found at every locus not knowingly selected for variability is not surprising. Given the large size of the samples, even rare alleles are likely to be found occasionally. We consider a locus to be polymorphic only if at least two alleles occur, the rarer of them with a frequency of 1% or higher, and both found in at least two independent samples of natural populations. An alternative, and more stringent, criterion can also be used—a locus is considered polymorphic only if at least two alleles occur each with a frequency of 5% or higher. The results are summarized in Table 1. By the 1% criterion, 43 out of 52 (83%) of the loci

TABLE 1. *Polymorphic (P) and monomorphic (M) loci in four sibling species of Drosophila.*

Gene	*D. willistoni*	*D. paulistorum*	*D. equinoxialis*	*D. tropicalis*
α-Gpdh	M	M	P*	M
Mdh-1	P	P	P	P
Mdh-2	M	M	M	M
Adh	P*	M	P	P
To	P*	P*	P	P
Est-2	P	P	P	P
Est-3	P	†	P	†
Est-4	P	P	P	P
Est-5	P*	P	P*	P
Est-6	P	P	P	‡
Est-7	P	P	‡	P
Aph-1	P	P	P*	P*
Aph-2	P	P	P	P
Lap-5	P	M	P	P

* Polymorphic by the 1%, but not by the 5% criterion.
† Not studied.
‡ No activity can be detected.

studied are polymorphic. If the 5% criterion is used instead, 35 out of 52 (67%) loci are polymorphic. If these loci are assumed to be a representative, unbiased sample of the genome of the species, it is clear that these species are genetically extremely polymorphic. Some of the ways in which our estimates may be biased will be discussed below.

We have found that, for any one locus the pattern of genetic variation is reasonably constant throughout the whole range of any one species. If a locus is monomorphic in one local population, it is generally so in all other populations. Moreover, all local populations are fixed for the same allele. A locus polymorphic in one locality is generally polymorphic in all other localities. In such cases the most common allele in one locality is generally also the most common one throughout the species.

The situation is different when the sibling species are compared with each other. In Table 2 we give the relative frequencies of various alleles at each of several loci. In the table, +++ indicates the most frequent or predominant allele; ++ indicates that the frequency of the allele is 0.05 or higher; + indicates that the frequency of the allele is between 0.01 and 0.05; * indicates that the allele has been found but at frequencies lower than 0.01; finally, a dash

TABLE 2. *Allelic frequencies in eight genes. See text for explanation of symbols.*

Gene	Alleles	*D. willistoni*	*D. paulistorum*	*D. equinoxialis*	*D. tropicalis*
α-Gpdh	0.94	*	*	+	*
	1.00	+++	+++	+++	+++
	1.06	*	*	*	*
Mdh-2	0.86	*	*	*	+++
	0.94	*	+++	+++	*
	1.00	+++	*	+	*
Adh	−1.0	+	*	*	+
	0.9	+	*	++	+++
	1.0	+++	+++	+++	++
To	0.98	*	*	++	++
	1.00	+++	+++	+++	+++
	1.02	+	+	+	+
Est-2	0.98	++	+	*	++
	1.00	+++	+++	+++	+++
	1.02	+	++	++	++
Est-5	0.80	—	—	—	++
	0.85	—	—	—	+++
	0.96	+	+	+	—
	1.00	+++	+++	+++	—
	1.04	+	++	+	—
Lap-5	0.98	++	—	—	+
	1.00	++	—	—	++
	1.03	+++	—	—	+++
	1.05	++	*	++	++
	1.07	*	+++	+++	+
	1.09	—	*	+	—

indicates that the allele has not been found in that species. The alleles are identified by their anodal migration relative to an arbitrary standard. Not all the alleles found at each locus are included in the table. As seen in Table 2, some loci polymorphic in some species are monomorphic in others. For instance, *Adh* and *Lap-5* are monomorphic in *D. paulistorum* but polymorphic in all three other species. Two species monomorphic at one locus may be fixed for the same allele (for instance, *α-Gpdh*), but frequently they are fixed for different alleles (for instance, *Mdh-2*). Two species polymorphic at the same locus have occasionally very different distributions of allelic frequencies; for instance, *D. equinoxialis* and *D. tropicalis* at the *Est-5* and *Lap-5* loci.

We have indicated in Table 3 the situations encountered when the genetic structures of two species at the same locus are compared[17]. We have assigned a numerical "index of genetic dissimilarity" to each condition. This index ranges in value from zero to five. A value of zero is given when two individuals each belonging to a different species will not, on the average, be more different than two individuals of the same species. A value of five is given when individuals of one species are nearly always genetically different at that particular locus from individuals of the other species. Values from one to four are given to the intermediate situations. Two alternative values are possible in some cases; which one is given in a particular case depends on the amount of genetic overlap between the two species at the locus in question.

To illustrate how this index of genetic dissimilarity is determined we may consider two examples. Assuming Hardy-Weinberg equilibrium, the expected

TABLE 3. *Index of genetic dissimilarity when two species, S_1 and S_2, are compared at one gene locus.*

Situation	Index
1. Both species are monomorphic	
(*a*) Same allele fixed in both species	0
(*b*) Different allele fixed in each species	5
2. S_1 monomorphic, S_2 polymorphic	
(*a*) Allele fixed in S_1 is the most frequent in S_2	1 or 2
(*b*) Allele fixed in S_1 is not the most frequent in S_2	4 or 5
3. Both species polymorphic, with one predominant allele (frequency of 80% or higher)	
(*a*) Same allele is predominant in both species	0 or 1
(*b*) Different allele is predominant in each species	4 or 5
4. One or both species highly polymorphic	
(*a*) Essentially similar frequency distributions in both species	1 or 2
(*b*) Different but overlapping frequency distributions	3
(*c*) Essentially nonoverlapping frequency distributions	4 or 5

zygotic frequencies of the genotypes at the *Mdh-2* locus are given in Table 4 for three sibling species. In *D. paulistorum* and *D. equinoxialis*, 99% of individuals are homozygotes 0.94/0.94. The combined expected frequencies of genotypes 0.86/0.94, 0.86/1.00, and 0.94/1.00 is about 1%. An individual with any of these four genotypes is about as likely to belong to *D. paulistorum* as to *D. equinoxialis*. The genetic dissimilarity between these two species at this locus is zero. Consider now *D. willistoni*. The combined expected frequencies of the four genotypes just mentioned is about 1%. In particular, the expected frequency of the 0.94/0.94 genotype in *D. willistoni* is 10^{-4}. About 96% of the flies of this species are expected to be homozygotes 1.00/1.00, while only one in 100,000 of either *D. paulistorum* or *D. equinoxialis* will have that genotype. The genetic dissimilarity at this locus between *D. willistoni* and either *D. paulistorum* or *D. equinoxialis* is, then, essentially complete, and assigned the value five.

TABLE 4. *Expected genotypic frequencies at the Mdh-2 locus.*

	Genotypes					
Species	0.86/0.86	0.86/0.94	0.86/1.00	0.94/0.94	0 94/1.00	1.00/1.00
D. willistoni	10^{-6}	10^{-5}	0.004	10^{-4}	0.012	0.962
D. paulistorum	10^{-6}	0.002	10^{-5}	0.992	0.006	10^{-5}
D. equinoxialis	10^{-5}	0.006	10^{-5}	0.986	0.008	10^{-5}

Similar calculations can be made for the *Lap-5* locus. This gene is monomorphic in *D. paulistorum* but polymorphic in *D. equinoxialis*. The allele fixed in *D. paulistorum* is the most frequent one in *D. equinoxialis*. We have the situation 2*a* in Table 3, the index may have value of either one or two. More than 99% of *D. paulistorum* flies are expected to have the genotype 1.07/1.07; about 77% of *D. equinoxialis* flies will have this genotype. Now 23% of *D. equinoxialis* flies will have one of five other genotypes (1.05/1.05, 1.05/1.07, 1.05/1.09, 1.07/1.09, and 1.09/1.09); the combined frequencies of these five genotypes in *D. paulistorum* is less than 1%. At this locus the genetic dissimilarity between *D. paulistorum* and *D. equinoxialis* is given the value of one. If the proportion of *D. equinoxialis* flies having genotypes absent in *D. paulistorum* would have been closer to 40 than to 20%, their index of genetic dissimilarity at this locus would have been two rather than one.

Although there is a certain degree of arbitrariness and imprecision in our index of genetic dissimilarity, we think that it gives a fair idea of the likelihood that an individual of a certain genotype will belong to one or other species. When averaged over many loci, the index gives an approximate estimate of the proportion of loci at which individuals of two species are genetically different.

The average indices of genetic dissimilarity between any two species for all 14 loci studied are given in Table 5. The average genetic dissimilarity between any two species is, for all species, 3.0. On a scale ranging from zero to five, this value of 3.0 indicates that individuals of any one species are genetically different from individuals of another species on the average at about somewhat more than half of their gene loci.

It is worth noting that the index of dissimilarity between *D. paulistorum* and *D. equinoxialis* is only 2.0, the lowest of all values in Table 5. This is consistent

TABLE 5. *Index of genetic dissimilarity among four sibling species.*

Species	*D. paulistorum*	*D. equinoxialis*	*D. tropicalis*	Average
D. willistoni	2.8	3.2	3.2	3.08
D. paulistorum	...	2.0	3 2	2.67
D. equinoxialis	...	...	3.6	2.97
D. tropicalis	...	...	...	3.31

with results published earlier[8] and additional observations by one of us (Dobzhansky) indicating that these species are most similar to each other among the four siblings when morphological, chromosomal, and behavioral differences are considered.

Discussion. Some possible sources of error should be considered, which may be relevant to the evaluation of the significance of our findings. First, it is assumed that simple Mendelian segregation of alternative allozymes of perceptibly different electrophoretic mobility indicates that these allozymes are coded by alleles of a single structural gene. An enzyme may, however consist of two or more polypeptide chains coded by two or more genes. If one chain is variable and the other constant, we may be overestimating the amount of genetic diversity revealed by our observations. A second possible source of error may have the opposite effect—allozymes of similar electrophoretic mobility are scored as due to the same allele, although they may actually differ in amino acid substitutions not altering that mobility. The amount of genetic differentiation may then be underestimated. A third contingency to be considered is that the enzymes we have chosen to study may be controlled by genes on the average more or less variable than the rest of the genome. There are various reasons why this might be so. We are dealing only with structural, and not with regulatory, genes. Also we are studying enzymes that are water-soluble. Regulatory genes, or structural genes coding for insoluble enzymes, may be more or less variable than the soluble enzymes of our sample. Finally, we must be concerned with the relatively small sample of the genome that we are surveying. The particular enzymes that we have studied were chosen without knowing whether they were variable or not, but rather because techniques were available for their assay. Thus there was no conscious bias in selecting these particular en-

zymes. But it is clear that with a sample of 14 loci our quantitative estimates of the proportion of loci which are polymorphic may have a large error.

For all species the mean proportion of polymorphic loci per species is 83% by the 1% criterion, or 67% if the more stringent 5% criterion for polymorphism is used. This gives evidence that there is an enormous amount of genetic variation in natural populations of these *Drosophila* species. Evidence that a large proportion of enzyme loci are polymorphic has been found also in other species of *Drosophila*[3-4,9-11], in the house mouse[5,12], and in man[13-14]. Estimates of the amount of genetic polymorphism in these various organisms have been surprisingly consistent in that in all cases there is indication that 30% (or more) of the loci are polymorphic. We have found a higher proportion of polymorphic loci than in most of the previous studies, but this may or may not be biologically significant given the possible sources of error indicated above. The consistent finding that 30–60% of gene loci coding for soluble enzymes are polymorphic warrants, in any case, the general conclusion that genetic variation is a ubiquitous phenomenon in organisms as diverse as *Drosophila* flies, mice, and men.

We can turn now to the principal question of this investigation, namely how much genetic differentiation there is among the sibling species. *D. willistoni*, *D. equinoxialis*, *D. tropicalis*, and *D. paulistorum* are closely related species, distinguishable morphologically mainly by some slight differences in the male genitalia. The species have largely overlapping geographic distributions, and can be collected in the same baits, indicating that they are also ecologically not too dissimilar.[15] Yet we have found that the average index of genetic dissimilarity between any two sibling species is 3.0 on a scale ranging from 0 to 5. Individuals belonging to different species are genetically different from each other at about 50% of the genes sampled in our study. We have no reason to believe that the genes we have studied are either more or less differentiated on the average than other structural loci. Thousands of structural genes exist in these species, as in any other metazoan organisms. Our results indicate, then, that in spite of their morphological and ecological similarity and their evolutionary affinity these species have very different gene pools. Considerable genetic differentiation among closely related groups of organisms has been observed also in other *Drosophila* species[16] and in mice.[12]

We have taken a step towards answering the question that was asked in the introduction, namely whether a major reorganization of the gene pools occurs during the process of speciation, or whether only a few genetic changes may be sufficient to initiate the development of reproductive isolation between populations. Our results make the first alternative more likely than the second, since closely related species are shown to be genetically very different. One conclusion is in any case warranted by our results—slight morphological and/or ecological dissimilarity such as exists between sibling species cannot be taken as evidence of little genetic differentiation. *D. willistoni*, *D. paulistorum*, *D. equinoxialis*, and *D. tropicalis* are morphologically very similar but genetically very different.

* Supported by NSF grant GB-12562 (International Biological Program), AEC contract AT-(30-1), and PHS Career Development Award K3 GM37265.

† Permanent addresses: Mourão - Faculdade de Filosofia, São José do Rio Prêto, São Paulo Brazil; Pérez-Salas - Universidad Central de Venezuela, Caracas; Richmond - University of Indiana, Bloomington.

[1] Mayr, E., *Animal Species and Evolution* (Cambridge: Harvard University Press, 1963).

[2] Dobzhansky, Th., *Genetics of the Evolutionary Process* (New York: Columbia University Press, 1970).

[3] Lewontin, R. C., and J. L. Hubby, *Genetics*, **54**, 595 (1966).

[4] Johnson, F. M., C. G. Kanapi, R. H. Richardson, M. R. Wheeler, and W. S. Stone, *Proc. Nat. Acad. Sci. USA*, **56**, 119 (1966).

[5] Selander, R. K., and S. Y. Yang, *Genetics*, **63**, 653 (1969).

[6] Dobzhansky, Th., L. Ehrman, O. Pavlovsky, and B. Spassky, *Proc. Nat. Acad. Sci. USA*, **54**, 3 (1964).

[7] Shaw, C. R., and A. L. Koen, in *Chromatographic and Electrophoretic Techniques*, ed. I. Smith (New York: Interscience 1968), vol. 2, p. 325.

[8] Burla, H., A. B. daCunha, A. R. Cordeiro, Th. Dobzhansky, C. Malogolowkin, and C. Pavan, *Evolution*, **3**, 300 (1949).

[9] Prakash, S., R. C. Lewontin, and J. L. Hubby, *Genetics*, **61**, 841 (1969).

[10] O'Brien, S. J., and R. MacIntyre, *Amer. Naturalist*, **103**, 97 (1969).

[11] Stone, W. S., M. R. Wheeler, F. M. Johnson, and K-I. Kojima, *Proc. Nat. Acad. Sci. USA*, **59**, 102 (1968).

[12] Selander, R. K., W. G. Hunt, and S. Y. Yang, *Evolution*, **23**, 379 (1969).

[13] Harris, H., *Proc. Roy. Soc. London, Ser. B*, **164**, 298 (1966).

[14] Lewontin, R. C., *Amer. J. Human Genet.*, **19**, 681 (1967).

[15] Ayala, F. J., in *Essays in Evolution and Genetics in Honor of Theodosius Dobzhansky*, eds. M. K. Hecht and W. C. Steere (New York: Appleton-Century-Crofts, 1970), pp. 121–158.

[16] Hubby, L., and L. H. Throckmorton, *Amer. Naturalist*, **102**, 193 (1968).

[17] Prakash, S. *Proc. Nat. Acad. Sci. USA*, **62**, 778 (1969).