Proc. Nat. Acad. Sci. USA
Vol. 69, No. 5, pp. 1094–1096, May 1972

Allozymes as Diagnostic Characters of Sibling Species of *Drosophila*

(genetic variation/evolution/speciation)

FRANCISCO J. AYALA AND JEFFREY R. POWELL

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

Communicated by Th. Dobzhansky, February 22, 1972

ABSTRACT **Considerable enzyme variation that is genetically controlled exists within and between natural populations of *Drosophila*. Our studies of two groups of sibling species show that allozyme differences can be used as species-diagnostic characters. A locus is defined as diagnostic if an individual can be correctly assigned to one of two species with a probability of 99% or higher. Between 15 and 32% of the loci studied are diagnostic for any two of the sibling species. If several diagnostic loci are used, the species of any individual can be diagnosed with virtual certainty.**

Studies of enzyme differences controlled by alleles at a single locus (allozymes) provide information about genetic variation in natural populations. These studies make possible the measurement of the amount of genetic differentiation between populations. When the genetic constitutions of two populations at a single locus are compared, the two extreme possibilities are: (*a*) complete identity—the same genotypes occur with identical frequencies in the two populations, and (*b*) complete differentiation—the two populations do not share any genotypes. Of course, intermediate situations also occur.

Statistical methods have been devised to quantify the amount of genetic differentiation between populations (1). These methods measure in various ways the amount of overlap between the distributions of genotypic frequencies in two populations at each locus. The average over all loci studied provides an estimate of the amount of genetic differentiation between the populations. The same amount of genetic differentiation may arise either by many loci that have a moderate amount of differentiation, or by complete or nearly complete differentiation at only a few loci, with identity of the other loci studied. For taxonomic purposes, however, these two situations have different implications. A locus at which complete differentiation exists between two populations can be used to diagnose the population to which an individual belongs. A locus at which only partial differentiation occurs cannot be so used. Many loci at which two populations are partially different can be used for diagnostic purposes, but statistical manipulations are more complex.

Can allozyme differences be used as diagnostic characters of closely related species? Sibling species are so similar in their morphology that they can hardly or not at all be told apart by their external morphology. Our studies of two groups of *Drosophila* species show that in this genus allozyme differences are good diagnostic characters between closely related sibling species.

MATERIALS AND METHODS

The *Drosophila willistoni* group consists of at least six sibling species. Two siblings, *D. pavlovskiana* and *D. insularis* are narrow endemics, the former in Guyana and the latter on some islands of the lesser Antilles. Four other sibling species, *D. willistoni*, *D. tropicalis*, *D. equinoxialis*, and *D. paulistorum* have wide geographic distribution in the tropics of the New World. Their distributions overlap from Guatemala, through Central America and northern South America, down to central Peru and Brazil. We have studied samples of each species from 20 or more localities.

The second group includes the two siblings, *D. pseudoobscura* and *D. persimilis*. The distribution area of *D. persimilis* is included in that of *D. pseudoobscura*. Our samples of both came from Mather (in the Sierra Nevada, elevation 4800 ft) and McDonald Ranch (in the Coast Range, elevation about 1000 ft) in California. *D. miranda* and *D. azteca* occur in the same localities; they are related to, but are distinguishable from *D. pseudoobscura* and *D. persimilis*.

We have used techniques of starch gel electrophoresis and enzyme assay that are described in a previous publication (2), with minor modifications to suit our materials.

RESULTS AND DISCUSSION

We have studied genetic variation at 28 loci in each of the six sibling species of *D. willistoni* (2–7). In every species small but significant differences in allelic frequencies may occur from one locality to another. *D. paulistorum* consists of several semispecies or incipient species. Genetic differentiation among the semispecies is discussed elsewhere (5). For the present purposes, we ignore the differences among localities or semispecies. Allelic frequencies for each species are obtained by pooling the data from all the individuals studied.

Table 1 gives the allelic frequencies at each of 12 loci in *D. willistoni*, *D. tropicalis*, *D. equinoxialis*, and *D. paulistorum*. Expected genotypic frequencies are calculated from the allelic frequencies by assumption of Hardy–Weinberg equilibrium. The proportion of overlap of the distribution of genotypic frequencies between any two populations can then be readily obtained for each locus. For example, *D. willistoni* and *D. equinoxialis* share alleles 1.00 and 1.04 at the *Me-1* locus. The expected frequencies of the genotypes that carry one or both alleles are:

	1.00/1.00	1.00/1.04	1.04/1.04
D. willistoni	0.9025	0.0760	0.0004
D. equinoxialis	0.000025	0.0099	0.9801

The overlap of the two species is, then, 0.000025 + 0.0099 + 0.0004 = 0.01035.

To estimate the probability of assigning correctly an individual to one of two species, we assume that the two species are equally common. We then assign individuals with each given genotype to the species in which that genotype has the higher frequency. Errors of attribution will be made with a frequency of half the amount of overlap of the genotypic distributions of the two species. For example, knowing their

TABLE 1. *Frequencies of alleles at twelve diagnostic loci in four species of Drosophila**

Gene	Alleles	*D. willistoni*	*D. tropicalis*	*D. equinoxialis*	*D. paulistorum*	Gene	Alleles	*D. willistoni*	*D. tropicalis*	*D. equinoxialis*	*D. paulistorum*
Lap-5	0.98	0.09	0.02	—†	—	*Mdh-2*	0.86	0.001	0.994	0.003	0.001
	1.00	0.29	0.19	—	—		0.94	0.02	0.005	0.994	0.993
	1.03	0.50	0.63	0.004	0.004		1.00	0.97	—	0.004	0.006
	1.05	0.09	0.15	0.21	0.08	*Me-1*	0.90	—	0.03	—	—
	1.07	0.007	0.01	0.71	0.86		0.94	—	0.91	—	0.004
	1.09	—	—	0.07	0.04		0.98	0.02	0.06	—	0.99
							1.00	0.95	—	0.005	0.005
Est-5	0.95	0.03	—	0.03	0.03		1.04	0.02	—	0.99	—
	1.00	0.96	—	0.94	0.84						
	1.05	0.01	—	0.02	0.13	*Tpi-2*	0.94	0.003	0.01	—	0.02
							1.00	0.98	0.98	0.02	0.98
Est-7	0.96	0.02	0.02	—	—		1.06	0.01	0.01	0.98	—
	0.98	0.16	0.11	—	—	*Pgm-1*	0.96	0.04	—	0.01	0.02
	1.00	0.54	0.62	—	0.002		1.00	0.87	0.01	0.35	0.94
	1.02	0.23	0.23	—	0.08		1.04	0.08	0.98	0.62	0.04
	1.05	0.05	0.03	—	0.78						
	1.07	0.003	0.001	—	0.09	*Adk-2*	0.96	0.01	0.05	—	—
							0.98	0.05	—	—	—
Aph-1	0.98	0.02	—	—	—		1.00	0.88	0.92	0.04	0.98
	1.00	0.84	0.05	0.02	0.01		1.02	0.05	—	—	—
	1.02	0.08	0.90	0.92	0.93		1.04	0.004	0.03	0.94	0.02
	1.04	0.06	0.04	0.06	0.03	*Hk-1*	0.96	0.04	0.02	0.08	—
							1.00	0.95	0.96	0.91	0.01
Acph-1	0.94	0.05	0.95	0.01	—		1.04	0.006	0.02	0.005	0.97
	1.00	0.92	0.03	0.17	—		1.08	—	0.001	0.002	0.02
	1.04	0.02	0.006	0.81	0.16						
	1.06	—	—	—	0.21	*Hk-3*	1.00	0.98	0.97	0.95	0.07
	1.08	—	—	0.004	0.62		1.04	0.006	0.01	0.04	0.92

Symbols for the genes are: *Lap* = leucine aminopeptidase (EC 3.4.1.1); *Est* = esterase (EC 3.1.1.2); *Aph* = alkaline phosphatase (EC 3.1.3.1); *Acph* = acid phosphatase (EC 3.1.3.2); *Mdh* = malate dehydrogenase (EC 1.1.1.37); *Me* = malic enzyme (EC 1.1.1.40); *Tpi* = triose phosphate isomerase (EC 1.2.1.9); *Pgm* = phosphoglucomutase; *Adk* = adenylate kinase (EC 2.7.4.3); *Hk* = hexokinase (EC 2.7.1.1). * Several alleles that occur with low frequencies are not included; † A dash indicates that the allele has not been found in the species.

genotype at the *Me-1* locus, individuals belonging to either *D. willistoni* or *D. equinoxialis* will be assigned to the wrong species with a frequency of 0.010325/2, or 0.00516. The probability of correct diagnosis of the species is, then, 0.99484.

To choose loci that will be diagnostic, we must decide what probability of error we are willing to tolerate. We use two criteria, the second is more stringent than the first. A locus is defined as diagnostic: (*a*) when the probability of assigning an individual to the correct species is 99% or higher, or (*b*) when that probability is 99.9% or higher. By use of criterion (*a*), all loci in Table 1 are diagnostic between at least one pair of species.

We have estimated the proportions of the loci that are diagnostic between each pair of species. The results for the six sibling species of the *D. willistoni* group are given in Table 2. On the average, 25% of the 28 loci studied permit a correct decision as to which of two species an individual belongs with a probability of 99% or higher. 11% of the loci permit correct identification of the species of an individual with a probability of 99.9% or higher. It should, however, be noted that

TABLE 2. *Percentage of loci that are diagnostic between any two species**

	D. willistoni	*D. tropicalis*	*D. equinoxialis*	*D. paulistorum*	*D. pavlovskiana*	*D. insularis*	Average
D. willistoni		17.9	21.4	25.0	25.0	32.1	24.3
D. tropicalis	10.7		21.4	35.7	28.6	28.6	26.4
D. equinoxialis	7.1	10.7		14.3	25.0	28.6	22.1
D. paulistorum	7.1	10.7	10.7		14.3	32.1	24.3
D. pavlovskiana	10.7	10.7	14.3	3.6		32.1	25.0
D. insularis	14.3	10.7	14.3	10.7	14.3		30.7
Average	10.0	10.7	11.4	8.6	10.7	12.1	

* Above the diagonal: correct diagnosis is made with a probability of 99% or higher at each locus. Below the diagonal: the probability of correct diagnosis is 99.9% or higher at each locus.

TABLE 3. *Frequencies of alleles at four diagnostic loci in two species of Drosophila**

Gene	Alleles	*D. pseudoobscura*	*D. persimilis*
Est-5	0.92	0.01	—†
	0.96	0.08	—
	1.00	0.51	—
	1.02	0.01	—
	1.04	0.18	—
	1.06	0.08	—
	1.08	0.11	0.01
	1.10	0.01	0.05
	1.12	—	0.83
	1.14	—	0.08
	1.16	—	0.03
Me-1	0.96	0.05	0.02
	0.98	—	0.11
	0.99	—	0.78
	1.00	0.24	0.09
	1.01	0.48	—
	1.02	0.21	—
	1.03	0.02	—
Me-2	0.98	0.04	—
	1.00	0.57	—
	1.02	0.30	0.007
	1.04	0.08	0.86
	1.06	—	0.13
Hk-3	0.98	0.02	—
	1.00	0.20	—
	1.02	0.64	—
	1.04	0.14	0.90
	1.06	—	0.09

Symbols for the genes, and * and † are as in Table 1.

the estimates in Table 2 are correct only for females. Four of the loci in Table 1, namely *Est-7*, *Aph-1*, *Pgm-1*, and *Adk-2* are sex-linked, and are consequently in hemizygous condition in males. The proportions of the loci that are diagnostic for males are, then, somewhat lower than shown in Table 2.

The data in Table 1 show that correct identification of the species to which an individual belongs can be made even when the four species are considered simultaneously. One locus alone, *Me-1*, permits a decision as to which of the four species any individual belongs with a probability of 98%.

We have studied 26 loci in *D. pseudoobscura* and *D. persimilis*. Four loci, or 15% of the total, are diagnostic between these two species. Their allelic frequencies are given in Table 3. Only one locus, *Est-5*, permits correct diagnosis of the species with a probability higher than 99.9%, and this is only for females, since the locus is sex linked. Males can be diagnosed at this locus with a probability of 99%. The polypeptide coded for by *Est-5* exists in the flies in dimeric as well as in monomeric form. In our gels, the monomer migrates farther towards the anode than the dimer, and permits easier discrimination of allozymes. Allelic variation at this locus has been studied in *D. pseudoobscura* and in *D. persimilis* by Prakash, Lewontin, and Hubby (8) and by Prakash (9). They studied the dimeric form of the enzyme and used polyacrylamide gels and techniques different from ours. Since our techniques differ, we have made the following tentative identifications between the alleles in our table—before the equal sign—and theirs: 0.92 = 0.85, 0.96 = 0.95, 1.00 = 1.00 and 1.02, 1.02 = 1.03, 1.04 = 1.07, 1.06 = 1.09, 1.08 = 1.12, 1.10 = 1.16.

TABLE 4. *Percentage of loci that are diagnostic between any two species**

	D. pseudoobscura	*D. persimilis*	*D. miranda*	*D. azteca*
D. pseudoobscura		15.4	27.8	47.6
D. persimilis	3.8		22.2	52.4
D. miranda	16.7	16.7		38.9
D. azteca	33.3	28.6	33.3	

* As in Table 2.

Table 4 gives the proportion of loci permitting diagnosis between each pair of species in *D. pseudoobscura*, *D. persimilis*, *D. miranda*, and *D. azteca*. Diagnostic loci are defined as in the *D. willistoni* group. There are more diagnostic loci between *D. miranda* and *D. pseudoobscura* or *D. persimilis* than between *D. pseudoobscura* and *D. persimilis*. The highest proportions of diagnostic loci occur between *D. azteca* and any of the three other species. These findings agree well with what is known about the amount of morphological and chromosomal differentiation among these four species (10, 11).

We thank Professor R. C. Lewontin, who provided the stocks of *D. pseudoobscura* that made possible the identifications of alleles at the *Est-5* locus. Professor Th. Dobzhansky read the manuscript and provided valuable advice. This work was supported by NSF Grant GB 30895 and AEC contract AT(04-3)-34.

1. Hedrick, P. W. (1971) *Evolution* **25**, 276–280.
2. Ayala, F. J., Powell, J. R., Tracey, M. L., Mourão, C. A. & Pérez-Salas, S. (1972) *Genetics* **70**, 113–139.
3. Ayala, F. J., Mourao, C. A., Pérez-Salas, S., Richmond, R. & Dobzhansky, Th. (1970) *Proc. Nat. Acad. Sci. USA* **67**, 225–232.
4. Ayala, F. J., Powell, J. R. & Dobzhansky, Th. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2480–2483.
5. Richmond, R. C. (1972) *Genetics* **70**, 87–112.
6. Ayala, F. J., Powell, J. R. & Tracey, M. L. (1972) *Genet. Res.* **19**, in press.
7. Ayala, F. J., *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, in press.
8. Prakash, S., Lewontin, R. C. & Hubby, J. L. (1969) *Genetics* **61**, 841–858.
9. Prakash, S. (1969) *Proc. Nat. Acad. Sci. USA* **62**, 778–784.
10. Dobzhansky, Th. & Sokolov, D. (1939) *J. Hered.* **30**, 3–19.
11. Dobzhansky, Th. & Epling, C. (1944) *Carnegie Inst. Wash. Publ.* **554**, 1–46.