Enzyme Variability in the *Drosophila willistoni* Group

VIII. Genetic differentiation and reproductive isolation between two subspecies

FRANCISCO J. AYALA AND MARTIN L. TRACEY

HOW much genetic differentiation occurs between populations as they evolve into different species? The first step in the process of geographic speciation is the genetic differentiation of allopatric populations. This genetic differentiation must proceed sufficiently far for the second step to occur, namely for the development of complete reproductive isolation between the populations if and when they geographically again come into contact. Without sufficient genetic differentiation previous to becoming sympatric, natural selection might not lead to the development of reproductive isolating mechanisms, and the two populations would then fuse into one single gene pool.

Subspecies of the same species represent what may be the first stage of geographic speciation. When populations of two different subspecies come into contact through part of their distributions, natural selection will favor the development of premating isolating mechanisms between them, if progenies from intersubspecific crosses have lower genetic fitness than progenies from intrasubspecific crosses.

The proportion of genes that are different in two populations can be estimated using the techniques of gel electrophoresis and enzyme assay. We report here our studies of two subspecies, *Drosophila willistoni willistoni* and *D. w. quechua,* showing that 1) incipient postmating reproductive isolating mechanisms have developed between the subspecies, and thus natural selection would favor intrasubspecific matings of the two subspecies were they to become sympatric; and 2) a substantial amount of genetic differentiation has occurred between the subspecies.

Materials and Methods

The sample of *D. w. quechua* used in this study was collected near Lima, Peru, by Professor Danko Brncic in the spring of 1972. The two populations of *D. w. willistoni* used for the study of reproductive isolation came from: Guayabero, on the left bank of the Guayabero River in the southwest of the Cordillera de la Magdalena, Colombia; and Manaus, on the Reserva Ducke, near the confluence of the Rio Negro and the Amazon. For the study of enzyme variation, many other populations described elsewhere[2,3] were sampled.

To study sexual isolation we used "observation chambers." Twelve virgin females and twelve males of each of two strains, separately aged for at least five days, were introduced into a chamber where matings were directly observed. The distal end of the wing was clipped in one or the other strain; each strain was marked in half of the replicate experiments of a given cross. Matings in the chambers were recorded as they occurred for a total of three hours. Most matings took place during the first hour of observation.

Enzyme variation was studied following standard techniques for starch gel electrophoresis and enzyme assay. Our procedures have been described elsewhere[2,4].

Results

Reproductive isolation

Sexual isolation between the subspecies was studied by testing the Lima strain of *D. w. quechua* with each of two strains of *D. w. willistoni,* one from Colombia and the other from northern Brazil. Sexual isolation was also studied between the two strains of *D. w. willistoni.* The results are summarized in Table I. A coefficient of sexual isolation, I, is calculated as follows:

$$I = (x_{AA} + x_{BB} - x_{AB} - x_{BA})/N \qquad (1)$$

where x_{AA}, x_{BB}, x_{AB}, and x_{BA} stand for the numbers of matings of $♀A \times A♂$, $♀B \times B♂$, $♀A \times B♂$, and $♀B \times A♂$, respectively; and $N = x_{AA} + x_{BB} + x_{AB} + x_{BA}$. The value of I can range from $+1$ (all matings are between individuals of the same strain) to -1 (all matings are between individuals of differ-

The authors are associated with the Department of Genetics, University of California, Davis, California 95616.
This work was supported by NSF Grant GB 30895. We are grateful to Mrs. Olga Pavlovsky and Miss Lori G. Barr for excellent technical assistance; and to Professor Danko Brncic for the collection of *Drosophila* flies from Lima. Professor Th. Dobzhansky read the manuscript and made useful suggestions.

ent strains); $I = 0$ when matings are at random. The variance of I is[8]:

$$\sigma_I^2 = (1 - I^2)/N \qquad (2)$$

The isolation coefficients have been calculated in two ways. First, by pooling the data from all replicate chambers, and obtaining a single I for each combination of strains. Second, by calculating I in each individual chamber, and obtaining the mean and standard error of I for each set of replicate chambers. These two procedures yield qualitatively identical results in our experiments (see Table I).

There are more homogamic than heterogamic matings in every combination of two strains, but the coefficient of sexual isolation is significantly greater than zero only between the Guayabero and Lima strains. Sexual isolation is not greater in one of the intersubspecific combinations (Manaus × Lima) than in the intrasubspecific combination (Manaus × Guayabero). An analysis of variance shows that there is not significant heterogeneity in the amount of sexual isolation between the three combinations of two strains. The mean isolation coefficient for all combinations of strains is statistically significant (0.182 ± 0.041, $t = 4.49$, $P < 0.001$); the isolation coefficient is also significantly positive for all combinations of strains if the data from all chambers are pooled ($I = 0.176 \pm 0.039$). Thus, our results indicate that incipient sexual isolation exists between geographic strains of *D. willistoni,* but that the isolation is not greater between strains of different subspecies than between strains of the same subspecies. A similar situation exists in *D. equinoxialis;* there is some incipient sexual isolation between geographic strains of the species, but that isolation is not greater between strains of different subspecies than between strains of the same subspecies[4].

The fertility or sterility of hybrids is tested as follows. About 10 virgin females of one strain and 10 males of another strain are placed together in a culture for about 10 days. From 30 to 50 F_1 flies are then placed in a culture. If F_2 progenies are produced, both male and female F_1 progenies are scored as fertile. If F_2 progenies are not produced, F_1 females are backcrossed to each parental strain separately to ascertain their fertility. The results can be briefly summarized. Crosses with *D. w. willistoni* as the female parent and *D. w. quechua* as the male parent yield fertile males and females. Crosses with *D. w. quechua* as the female parent produce fertile males and females when crossed with *D. w. willistoni* strains from western Ecuador, northern Central America, or Mexico. *D. w. quechua* females crossed with *D. w. willistoni* males from Colombia, Venezuela, Trinidad, or Brazil, produce fertile females and sterile males (see also Dobzhansky[6]). Thus incipient reproductive isolation exists between *D. w. quechua* and *D. w. willistoni* from continental South America. Whether *D. willistoni* strains from Ecuador west of the Andes, and from Central America and Mexico belong to the subspecies *quechua* remains to be ascertained; some recent results of DeToledo[5] are inconclusive.

Genetic variation

Table III summarizes the allelic variation found in the two subspecies at 30 loci coding for enzymes. The *D. w. willistoni* data are taken for the most part from Ayala *et al.*[2, 3], although we have included here

Table I. Sexual isolation between strains of *Drosophila willistoni willistoni* and *D. w. quechua*

A	B	No. of chambers	♀A×A♂	♀B×B♂	♀A×B♂	♀B×A♂	Total matings	% homogametic matings	$I \pm$ SD	$\bar{I} \pm$ SE
Manaus	Lima	12	64	67	51	58	240	54.6	.092±.064	.105±.088
Guayabero	Lima	16	92	100	56	55	303	63.4	.267±.055*	.270±.045*
Manaus	Guayabero	6	30	31	22	27	110	55.5	.109±.095	.104±.060
Total		34	186	198	129	140	653	58.8	.176±.039*	.182±.041*

I=coefficient of sexual isolation calculated by pooling the data from all chambers
$\bar{I}$=mean coefficient of sexual isolation calculated from the isolation coefficients in the individual chambers
SD=standard deviation; SE=standard error of the mean
* Statistically significant, $P<0.001$

Table II. Summary of genetic variation in natural populations of two subspecies of *Drosophila willistoni*

Subspecies	Loci studied	Genes sampled per locus	Polymorphic loci per population (1)	(2)	Heterozygous loci per individual
D. w. willistoni	30	3478±553	.607±.057	.865±.036	.197±.032
D. w. quechua	25	80±6	.440	.680	.174±.040

* A locus is considered polymorphic in a given population 1) when the frequency of the most common allele is < 0.95; 2) when the frequency of the second most common allele is > 0.01

some previously unpublished results; the data for *D. w. quechua* are new. Table III shows for each locus the sample size, the allelic frequencies, and the expected frequency of heterozygous individuals on the assumption of Hardy-Weinberg equilibrium. We have studied only the Lima population of *D. w. quechua,* but up to 80 local populations of *D. w. willistoni.* The frequency of heterozygous individuals for the latter subspecies is given by the mean and standard error for all local populations studied.

A summary of the overall amount of genetic polymorphism is given in Table II. The more nearly exact measure of the amount of genetic variation in outcrossing, sexually reproducing organisms is the proportion of heterozygous loci per individual. We have estimated this parameter by averaging over all loci the proportion of heterozygous individuals at each locus. The frequency of heterozygous loci per individual is 19.7 ± 3.2 percent in *D. w. willistoni,* and 17.4 ± 4.0 percent in *D. w. quechua.* The difference in heterozygosity between the subspecies is not, however, statistically significant. At each of the 25 loci studied in both subspecies we have obtained their difference in the frequency of heterozygotes; the

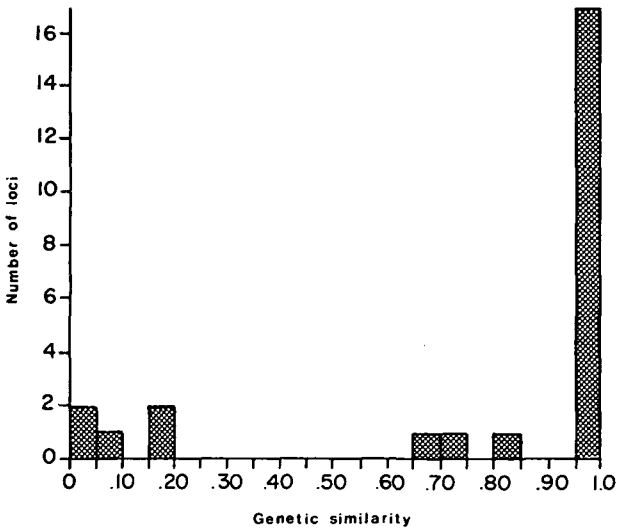

FIGURE 1—Histogram showing the number of loci within a given range of values of genetic similarity in the comparison between *Drosophila willistoni willistoni* and *D. w. quechua.*

Table III. Genetic variation at 30 loci in natural populations of two subspecies, *Drosophila willistoni willistoni* and *D. w. quechua*

Gene	Subspecies	Wild genomes sampled	Alleles* 1	2	3	4	5	Frequency of heterozygotes
Lap-5	*quechua*	114	.000	.035	.719	.246	.000	.421
	willistoni	10348	.008	.096	.348	.464	.082	.629±.007
Est-2	*quechua*	108	.231	.769	.000	.000		.356
	willistoni	7012	.003	.041	.941	.006		.117±.008
Est-3	*quechua*	48	.000	1.000	.000			.000
	willistoni	3914	.019	.948	.021			.108±.009
Est-4	*quechua*	114	.000	1.000	.000	.000		.000
	willistoni	9692	.004	.146	.838	.011		.270±.012
Est-5	*quechua*	108	.028	.972	.000			.054
	willistoni	10432	.029	.954	.016			.089±.006
Est-6	*willistoni*	2418	.785	.194	.008			.285±.030
Est-7	*quechua*	59	.847	.000	.153	.000	.000	.259
	willistoni	6819	.025	.147	.563	.211	.049	.601±.009
Aph-1	*quechua*	170	.000	.953	.012	.035		.091
	willistoni	6272	.030	.873	.056	.036		.136±.020
Acph-1	*quechua*	66	.000	.000	.894	.000	.106	.190
	willistoni	3066	.001	.023	.947	.028	.000	.102±.018
Acph-2	*willistoni*	847	.025	.917	.056			.151±.023
Ald	*willistoni*	404	.010	.032	.880	.068	.000	.182±.045
Adh	*quechua*	112	.000	.000	1.000	.000		.000
	willistoni	5916	.010	.034	.951	.001		.103±.015
Mdh-2	*quechua*	66	.000	.015	.470	.515	.000	.514
	willistoni	6680	.001	.000	.009	.982	.006	.040±.068
αGpdh	*quechua*	64	.000	.969	.000	.031		.061
	willistoni	6748	.003	.992	.004	.000		.021±.005
Idh	*quechua*	64	.000	1.000	.000			.000
	willistoni	3168	.004	.961	.029			.086±.023

* *Alleles. Lap-5:* 1=96, 2=98, 3=100, 4=103, 5=105. *Est-2:* 1=98, 2=100, 3=102, 4=104. *Est-3:* 1=98, 2=100, 3=102. *Est-4:* 1=98, 2=100, 3=102, 4=104. *Est-5:* 1=95, 2=100, 3=105. *Est-6:* 1=100, 2=104, 3=108. *Est-7:* 1=96, 2=98, 3=100, 4=102, 5=105. *Aph-1:* 1=98, 2=100, 3=102, 4=104. *Acph-1:* 1=84, 2=94, 3=100, 4=104, 5=106. *Acph-2:* 1=98, 2=100, 3=103. *Ald:* 1=96, 2=98, 3=100, 4=102, 5=104. *Adh:* 1=90, 2=98, 3=100, 4=106. *Mdh-2:* 1=86, 2=92, 3=94, 4=100, 5=106. *αGpdh:* 1=94, 2=100, 3=106, 4=112. *Idh:* 1=96, 2=100, 3=104.

mean of these differences over all loci, with its standard error, is 2.3 ± 3.6 percent.

We have measured the genetic differentiation between the subspecies by calculating their average genetic similarity, J, and genetic distance, D, per locus, according to described procedures [4,9]; these statistics are $J = 0.817$, $D = 0.202$. Thus, on the average, about 20.2 electrophoretically detectable substitutions for every 100 loci have taken place in the divergence between the subspecies. In contrast, little genetic variation has occurred between the local populations of *D. w. willistoni*. We have surveyed populations extending from Central America and the West Indies, through much of continental South America, down to south Brazil. Yet the average genetic similarity and distance between local populations of the *willistoni* subspecies are $J = 0.986 \pm 0.006$, and $D = 0.015 \pm 0.006$.

Table III shows that at most loci the two subspecies are essentially identical in their allelic variation, while at a few loci they are very different. Genetic differentiation between the two subspecies is for the most part due to the contribution of these few loci. Figure 1 is a histogram showing the distribution of loci within a given range of genetic similarity between the two subspecies. At 17 loci (68 percent of the total, the two subspecies are essentially identical, while at 5 loci (20 percent of the total) they are sharply different. At only 3 loci (12 percent) the genetic similarity falls between 0.20 and 0.95, although this range represents 75 percent of the total range. Using the method of Ayala and Powell[1], we have calculated the probability of correct diagnosis of the subspecies of a single individual of known genotype for each of the five loci with $J < 0.20$. These probabilities are 0.974, 0.989, 0.994, 0.999, and 0.99998 for *Est-7, Est-4, Odh-1, Est-2* and *Xdh,* respectively. Using jointly these five loci, the probability of incorrect identification of the subspecies of an individual of known genotype is 3.4×10^{-14}. It is clear that allozymes (enzymes coded by alleles at a given locus) can be used for the diagnosis of subspecies in *D. willistoni* as well as in *D. equinoxialis*[4]. It had been previously established that allozymes can be used as diagnostic characters of sibling species of *Drosophila*[1].

Table III. Continued

Gene	Subspecies	Wild genomes sampled	Alleles* 1	2	3	4	5	Frequency of heterozygotes
G3pdh	*quechua*	64	.062	.937	.000			.117
	willistoni	190	.055	.908	.035			.148±.053
Odh-1	*quechua*	64	.016	.000	.984			.031
	willistoni	1088	.039	.882	.071			.180±.019
Odh-2	*willistoni*	40	.975	.025				.060±.024
Me-1	*quechua*	64	.000	1.000				.000
	willistoni	2882	.028	.959	.006			.079±.013
Me-2	*quechua*	64	.000	.031	.437	.469	.062	.584
	willistoni	1368	.017	.000	.829	.000	.154	.235±.044
Xdh	*quechua*	64	.547	.422	.000	.000	.000	.522
	willistoni	2320	.000	.007	.114	.468	.322	.648±.014
Ao-1	*willistoni*	186	.011	.091	.833	.038	.016	.294±.051
To	*quechua*	64	.000	1.000	.000	.000		.000
	willistoni	4949	.080	.898	.008	.006		.131±.029
Tpi-2	*quechua*	64	.000	1.000	.000			.000
	willistoni	2478	.003	.984	.013			.041±.010
Pgm-1	*quechua*	64	.000	.859	.141	.000		.242
	willistoni	2636	.018	.891	.088	.001		.186±.025
Adk-1	*quechua*	64	.141	.687	.141	.031		.487
	willistoni	2150	.275	.561	.148	.000		.527±.019
Adk-2	*quechua*	64	.016	.000	.984	.000	.000	.031
	willistoni	2580	.011	.032	.926	.013	.011	.171±.023
Hk-1	*quechua*	64	.000	.031	.969	.000		.061
	willistoni	1620	.043	.000	.937	.013		.099±.016
Hk-2	*quechua*	64	.000	.812	.047	.141		.318
	willistoni	2228	.008	.915	.039	.033		.138±.032
Hk-3	*quechua*	64	.000	1.000	.000	.000		.000
	willistoni	2060	.005	.979	.009	.001		.039±.009

* *G3pdh:* 1=96, 2=100, 3=105. *Odh-1:* 1=96, 2=100, 3=104. *Odh-2:* 1=100, 2=106. *Me-1:* 1=98, 2=100, 3=104. *Me-2:* 1=96, 2=98, 3=100, 4=102, 5=104. *Xdh:* 1=95, 2=97, 3=98, 4=100, 5=101. *Ao-1:* 1=95, 2=98, 3=100, 4=102, 5=105. *To:* 1=86, 2=100, 3=102, 4=112. *Tpi-2:* 1=94, 2=100, 3=106. *Pgm-1:* 1=96, 2=100, 3=104, 4=112. *Adk-1:* 1=100, 2=106, 3=112, 4=118. *Adk-2:* 1=96, 2=98, 3=100, 4=102, 5=104. *Hk-1:* 1=96, 2=98, 3=100, 4=104. *Hk-2:* 1=96, 2=100, 3=104, 4=112. *Hk-3:* 1=96, 2=100, 3=104, 4=112. Some rare alleles have been omitted from the table.

Discussion

The two subspecies, *D. w. willistoni* and *D. w. quechua* show incipient reproductive isolation in the form of sterility of hybrid crosses when *quechua* is the female parent, and the male parent is *willistoni* from South America east of the Andes. If populations of the two subspecies were to come into contact in some locality or region, natural selection would promote the development of genetic systems favoring intrasubspecific crosses. Individuals involved in intersubspecific crosses would have lower fitness than those mating within the subspecies, owing to the sterility of some hybrid males. Whether the natural selection would lead to the development of premating isolating mechanisms, and eventually to the formation of two species, remains conjectural. In any case, the two subspecies represent a stage of evolutionary divergence that is antecedent to speciation. This degree of evolutionary divergence has been achieved with genetic divergence at substantial proportion of the loci. On the average, 20 electrophoretically detectable allelic substitutions for every 100 loci have taken place between *D. w. willistoni* and *D. w. quechua*. The total number of allelic substitutions between the subspecies is probably substantially greater, since it is estimated that only between one-quarter and one-third of all allelic substitutions are detectable by electrophoretic techniques.

The genetic distance between two other subspecies, *D. equinoxialis equinoxialis* and *D. e. caribbensis,* which are closely related to, and largely sympatric with *D. willistoni*[4] is $D = 0.255 \pm 0.008$. Selander *et al.*[10] have studied genetic variation in two subspecies of the house mouse, *Mus musculus musculus* and *M. m. domesticus;* using their data, the mean genetic distance between these subspecies can be calculated as 0.171 ± 0.009. The mean genetic distance between populations of two subspecies of salamanders, *Taricha torosa torosa* and *T. t. sierrae,* is $D = 0.168 \pm 0.042$[7]. Although the number of subspecies adequately studied is very small, the presently available evidence indicates that a substantial amount of genetic differentiation occurs in the formation of subspecies. It remains unknown how much additional genetic differentiation is likely to take place before complete reproductive isolation, and thus speciation, is achieved. The genetic distance between closely related species is frequently as great as or greater than 0.50[4], but it is not known how much of the differentiation occurred previous to the achievement of the species grade, and how much took place after the species were formed.

Summary

Incipient reproductive isolation, in the form of hybrid sterility, exists between the subspecies *Drosophila willistoni willistoni,* and *D. w. quechua.* Genetic variation has been studied in both subspecies using the techniques of starch gel electrophoresis. Depending on the criterion of polymorphism chosen, the proportion of polymorphic loci per population is 60.7 or 86.5 percent in *D. w. willistoni,* and 44.0 or 68.0 percent in *D. w. quechua.* The average proportion of heterozygous loci per individual is 19.4 percent in *D. w. willistoni,* and 17.4 percent in *D. w. quechua.* Thus, a very large amount of genetic variation occurs in these subspecies.

We have estimated that, on the average, 20.2 electrophoretically detectible allelic substitutions for every 100 loci have occurred between the subspecies, but only 1.5 per 100 loci between geographic populations of the same subspecies, on the average. The formation of incipiently reproductively isolated subspecies is accompanied by a substantial amount of genetic differentiation.

Literature Cited

1. Ayala, F. J. and J. R. Powell. Allozymes as diagnostic characters of sibling species of *Drosophila. Proc. Natl. Acad. Sci.* 69:1094–1096, 1972.

2. ———, ———, and Th. Dobzhansky. Polymorphisms in continental and island populations of *Drosophila willistoni. Proc. Natl. Acad. Sci.* 68:2480–2483. 1971.

3. ———, ———, M. L. Tracey, C. A. Mourão, and S. Pérez-Salas. Enzyme variability in the *Drosophila willistoni* group. IV. Genetic variation in natural populations of *Drosophila willistoni. Genetics* 70:113–139. 1972.

4. ———, M. L. Tracey, L. G. Barr, and J. G. Ehrenfeld. Genetic and reproductive differentiation of the subspecies, *Drosophila equinoxialis caribbensis.* Submitted to *Evolution.*

5. De Toledo Cardoso, J. Estudo da esterilidade entre diferentes linhagens de *Drosophila willistoni. Rev. Brasil. Biol.* 31:145–151. 1971.

6. Dobzhansky, Th. "Wild" and "domestic" species of *Drosophila. In* The Genetics of Colonizing Species. J. Baker and G. L. Stebbins, Eds. Academic Press, New York. p. 533–546. 1965.

7. Hedgecock, D. and F. J. Ayala. Evolutionary divergence in the genus *Taricha* (Salamandridae). *Copeia* in press.

8. Malogolowkin-Cohen, C., A. Solima-Simmons, and H. Levene. A study of sexual isolation between certain strains of *Drosophila paulistorum. Evolution* 19:95–103. 1965.

9. Nei, M. Genetic distance between populations. *Am. Nat.* 106:283–292. 1972.

10. Selander, R. K., W. G. Hunt, and S. Y. Yang. Protein polymorphisms and genic heterozygosity in two European subspecies of the house mouse. *Evolution* 23: 379–390. 1969.