The Speciation history of *Drosophila pseudoobscura* and Close Relatives: Inferences from DNA Sequence Variation at the Period Locus

Rong-Lin Wang[1] and Jody Hey

Department of Biological Sciences, Rutgers University, Piscataway, New Jersey 08855-1059

Manuscript received June 3, 1996
Accepted for publication August 3, 1996

ABSTRACT

Thirty-five period locus sequences from *Drosophila pseudoobscura* and its siblings species, *D. p. bogotana, D. persimilis,* and *D. miranda,* were studied. A large amount of variation was found within *D. pseudoobscura* and *D. persimilis,* consistent with histories of large effective population sizes. *D. p. bogotana,* however, has a severe reduction in diversity. Combined analysis of *per* with two other loci, in both *D. p. bogotana* and *D. pseudoobscura,* strongly suggest this reduction is due to recent directional selection at or near *per* within *D. p. bogotana.* Since *D. p. bogotana* is highly variable and shares variation with *D. pseudoobscura* at other loci, the low level of variation at *per* within *D. p. bogotana* can not be explained by a small effective population size or by speciation via founder event. Both *D. pseudoobscura* and *D. persimilis* have considerable intraspecific gene flow. A large portion of one *D. persimilis* sequence appears to have arisen via introgression from *D. pseudoobscura.* The time of this event appears to be well after the initial separation of these two species. The estimated times since speciation are one mya for *D. pseudoobscura* and *D. persimilis* and 2 mya since the formation of *D. miranda.*

EVOLUTIONARY biologists are concerned with both the population genetic mechanisms behind evolutionary processes and the reconstruction of phylogeny. The study of speciation is at the interface of these two areas of investigation. Knowledge of speciation is fundamental to our understanding of how the biological landscape grows and changes at a macrolevel. Yet it is the forces at the microlevel, including mutation, migration, random genetic drift, and natural selection, that lead to speciation.

Speciation, as a genetic process, can be studied from two different empirical perspectives. First are studies of the genetic basis of species differences and of reproductive barriers among species. These types of studies are focused on the number and identity of loci at which divergence may have, in some sense, contributed to speciation (WU and PALOPOLI 1994). While critically important, these analyses do not inform on the historical microlevel forces associated with speciation. A different perspective can be gained by studying patterns of DNA sequence variation, within and among closely related species, at loci that may or may not have been associated with species formation. This approach is essentially an extension of traditional population genetic questions to distant time periods, and it is made possible by a recent synergism between genealogical population genetic models and studies on DNA sequence variation (see *e.g.,* AVISE and BALL 1990; HEY 1994; TEMPLETON 1994).

Corresponding author: Jody Hey, Department of Biological Sciences, Rutgers University, Nelson Labs, P.O. Box 1059, Piscataway, NJ 08855-1059. E-mail: hey@mbcl.rutgers.edu

[1]*Present address:* Department of Plant Biology, University of Minnesota, 1445 Gortner Ave., St. Paul, MN 55108-1020.

One of the best studied species groups includes *Drosophila pseudoobscura* and its close relatives: *D. persimilis, D. miranda,* and subspecies *D. p. bogotana. D. pseudoobscura* and *D. persimilis* were once thought to be a single species but were later differentiated from each other based primarily on different chromosome arrangements and nearly complete sexual isolation (DOBZHANSKY and EPLING 1944). Females of both species discriminate against heterospecific males while males court females of either species (MAYR 1946). Crossed under laboratory conditions in either direction, they generate sterile males but fertile females. There is some chromosome inversion (DOBZHANSKY 1973) and mitochondrial DNA data (POWELL 1983; HALE and BECKENBACH 1985) that suggest some degree of gene flow between these two species.

Later came the discovery of *D. miranda* (DOBZHANSKY 1935) and *D. pseudoobscura bogotana* (DOBZHANSKY *et al.* 1963), an isolated population in Bogota, Columbia. Reproductive isolation between *D. miranda* and its sibling species is essentially complete (DOBZHANSKY and EPLING 1944) but that between *D. pseudoobscura* and *D. p. bogotana* is not: female *D. p. bogotana* crossed to male *D. pseudoobscura* give rise to sterile males and fertile females (PRAKASH 1972).

D. pseudoobscura has the widest range of geographic distribution within the group, extending from British Columbia, along the western third of the North American continent, to Mexico and Guatemala. *D. persimilis* and *D. miranda* are sympatric to *D. pseudoobscura* and found in the mountain ranges along the U.S. Pacific Coast. Morphological, chromosomal, and molecular evidence indicate that *D. pseudoobscura* is most closely re-

TABLE 1

A list of samples and their geographic origins

Species	NDSRC[a] lines	Sample numbers	Locations
D. pseudoobscura	0121.0	1	Tucson, AZ
	0121.1	2	Death Valley, CA
	0121.33	3	Kelowna, British Columbia
	0121.34	4	Cuernavaca (Mexico)
	0121.38	5	Mather, CA
	0121.41	6	Mather, CA
	0121.37	7	Mather, CA
	0121.32	8	Zimapan, Hidalgo (Mexico)
	0121.78	9	Dilley, TX
	0121.79	10	Port Coquitlam, British Columbia, Canada
	0121.83	11	Spray, OR
D. miranda	0101.4	22	Port Coquitlam, British Columbia
	0101.3	23	Port Coquitlam, British Columbia
	0101.5	24	Port Coquitlam, British Columbia
	0101.7	25	Port Coquitlam, British Columbia
D. persimilis	0111.0	40	Cold Creek, CA
	0111.2	41	Mt. San Jacinto, CA
	0111.17	42	Port Townsend, Washington
	0111.23	43	Victoria, British Columbia
	0111.24	44	McDonald Ranch, CA
	0111.1	45	Quesnal, Canada
	0111.18	46	McDonald Ranch, CA
	0111.27	47	Mt. San Jacinto, CA
	0111.29	48	Mt. San Jacinto, CA
	0111.35	49	Mt. San Jacinto, CA
	0111.38	50	Victoria, British Columbia
D. p. bogotana	0121.35	60	Bogota, Columbia
	0121.68	61	Bogota
	0121.69	62	Bogota
	0121.71	63	Bogota
	Oicata 9A	67	Bogota
	Sutatausa 4	70	Bogota
	Chiquinquira A	73	Bogota
	A Cacientes 4C	74	Bogota
	EL Recreo	77	Bogota

[a] National Drosophila Species Resources Center (Bowling Green, OH).

lated to *D. p. bogotana*, and together, they share a common ancestor with *D. persimilis* (GODDARD *et al.* 1990). *D. miranda* is an outgroup to all other three species.

Together, *D. pseudoobscura* and its close relatives have been the subject of vigorous population genetic research for several decades (LEWONTIN *et al.* 1981). At the DNA level, much of the focus has been on *D. pseudoobscura* and on the relationship between *D. pseudoobscura* and *D. p. bogotana*. *D. pseudoobscura* is clearly a species with a large historical effective population size with ample gene flow across much of its range (RILEY *et al.* 1992; SCHAEFFER and MILLER 1992; VEUILLE and KING 1995). There is also strong evidence for the action of natural selection acting on third chromosome inversion polymorphisms (AQUADRO *et al.* 1991; POWELL 1992). *D. pseudoobscura* and *D. p. bogotana* share considerable DNA sequence variation at the Alcohol dehydrogenase locus (*Adh* either because of ongoing gene flow or recent divergence or both (SCHAEFFER and MILLER 1991; POWELL 1992). To date there has not been a nuclear gene DNA sequence study on variation in *D. persimilis*.

Our study has focused on a 1.5-kb portion of the period locus (*per*). *Per* is an essential component of the biological clock in *D. melanogaster* (KONOPKA and BENZER 1971; ROSBASH and HALL 1989; EDERY *et al.* 1994) and found to occur as a single copy gene on the *X* chromosome. *Per* is also involved in the determination of male courtship songs (KYRIACOU and HALL 1980, 1989; KYRIACOU 1990) which in turn affect the female response to male mating attempts (KYRIACOU and HALL 1982, 1984; KYRIACOU *et al.* 1992). *Per* thus appears to be a candidate for evolutionary changes in mating behavior that contribute to reproductive isolation and speciation. However, in the *melanogaster* species complex, there is no evidence that variation at *per* has contributed to reproductive isolation or to species divergence (KLIMAN and HEY 1993; RITCHIE and KYRIACOU 1994). Regardless of whether evolution at *per* contributes to re-

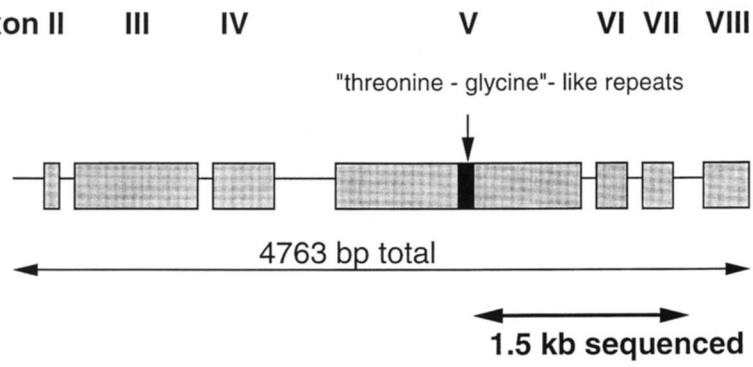

FIGURE 1.—Diagram of the period locus, *D. pseudoobscura*, based on COLOT *et al.* 1988.

productive isolation, a study of sequence variation within and among closely related species can yield valuable information on species divergence.

The primary objective of this study is to gain a genealogical perspective of the population genetic forces associated with the formation of the two sibling species, *D. pseudoobscura* and *D. persimilis,* as well as the geographically isolated subspecies, *D. p. bogotana.* Assessments of DNA sequence variation can reveal relative historical effective population sizes, and the patterns of shared variation may be used to assess historical levels of gene flow within and between species (HUDSON *et al.* 1992). Under some circumstances, this type of data can also reveal the action of positive directional and balancing selection (HUDSON *et al.* 1987). In addition, we hope also to estimate the divergence times among the species.

MATERIALS AND METHODS

Species and strains: A total of 35 strains were studied (Table 1). Most were acquired from the National Drosophila Species Resource Center (NDSRC, Bowling Green, OH). The rest were kindly provided by DANIEL WEINREICH.

Sequencing: DNA from individual male flies was extracted according to protocol 48 of ASHBURNER (1989). From each sample of genomic DNA, a section of *per* (either between positions 2849–4932 or 2993–4932, based on COLOT *et al.* 1988) was PCR-amplified. Additional DNA preparation and sequencing followed the method of KLIMAN and HEY (1993). Both strands were sequenced for each strain. A total of seven 20-bp long sequencing primers, spaced ~200 bp apart, were used on each strand. The final length of sequenced portion was ~1.5 kb, flanking the *D. pseudoobscura* equivalent of the Thr-Gly repeat and covering the exon 5-7 and intron 5-7 (Figure 1). All sequences have been submitted to GenBank (accession numbers L81261–L81295).

Data analysis: DNA sequences were assembled and aligned visually. Intraspecific diversity was estimated by average pairwise differences per base pair (π) and by the number of polymorphic sites within a species. Interspecific divergence was assessed by the number of fixed and shared polymorphisms between species and the net average pairwise divergence (NEI 1987). Several tests were conducted to assess the history of natural selection at *per* (HUDSON *et al.* 1987; MCDONALD and KREITMAN 1991; TAJIMA 1989). Neighbor-joining trees (SAITOU and NEI 1987) and bootstrap estimates were constructed using the programs DNADIST, NEIGHBOR, SEQBOOT and CONSENSE in Phylip3.5 (FELSENSTEIN 1993). The minimum number of recombination events was assessed with the method of HUDSON and KAPLAN (1985), and an estimate of the population recombination parameter was obtained using the method of HUDSON (1987). The estimate of divergence times among the species is based on the substitution rates at synonymous sites (SHARP and LI 1989).

Outlier HKA test: The HKA test of natural selection is similar to a chi-square test and asks whether the relative levels of intraspecific polymorphism and interspecific divergence for a locus, are consistent across loci (HUDSON *et al.* 1987). The first step in applying the test is to complete a table that has, for example, as many rows as there are loci in the data set and three columns: one for variation in one species, one for variation in the second species, and one for the divergence between species. Like a chi-square test, the test proceeds by generating a companion table of values that are expected under a null model and then a third table of standardized discrepancies between observations and expectations (HUDSON *et al.* 1987). Rejection of the null hypothesis requires that the sum of standardized discrepancies between observations and expectations be greater than expected by chance. However, this may not be the null hypothesis of the greatest interest. If selection has occurred at just one locus in one species, then only one cell of the table of standardized discrepancies may have a large value, and the overall test may not detect the discrepancy. Like the chi-square test with multiple rows and columns, the test is most powerful for departures from the null model that affect multiple cells in the data matrix.

We have modified the HKA test to one that is sensitive to single cell departures from the null hypothesis. In this outlier test, the test statistic is the maximum standardized discrepancy (MSD) observed for any polymorphism value. If one of the species has experienced a recent selective sweep or a long standing balanced polymorphism at just one of the loci, then the standardized discrepancy for that observation will be high. This test statistic is then compared with a neutral distribution of the MSD that is generated by 5000 independent HKA coalescent simulations (HUDSON *et al.* 1987; HILTON *et al.* 1994). For each simulation, the MSD observed for any locus, for polymorphism in either species, is the MSD measure for that simulation and is recorded. If the observed MSD value is >95% of the entire frequency distribution of simulated MSD values, then the outlier test is significant.

RESULTS

Polymorphisms and divergence: Polymorphisms in 35 sequences are summarized in Figure 2. *Drosophila pseudoobscura* is the most variable and *D. p. bogotana* the least (Table 2). As expected under a hypothesis of reduced constraint, there is greater variation in synonymous sites than replacement sites and greater variation in introns than in exons for *D. pseudoobscura* and *D.*

```
position          1111 1122233333 3334444555 5555566666 6667777888 88889
position    3466890012 5727822355 6780125001 1244601135 7884457034 44560
position    9068661432 8358336805 1665682142 5158717884 7690680292 48069
S/R/I/D     RRRSSRSSSS SSRSRSSSSR RRSRSSSRRS SSSSRRSRRR SSSSSSSRII IIIII
amino acids NNDDGSPPLS FHTGGGNQLS GAAPTAAAAT PPRNAAQAIA PFYPPNLY-- -----
replacement GGN..A.... ..A.A....N AV.S...SP. ....PG.TMT .......F.. .....
Consensus   AAGCGTGCCG CTACGCCGCG GCCCCCAGGG TGCCGCGGAG GCCCCTCACC CTGCA
BOGOTA61    .....G.... .....T.... .....G.... .......... A.....T... ...G.
BOGOTA60    .....G.... .....T.... .......... ......NN.. A.....T... ...G.
BOGOTA62    .....G.... .....T.... .....G.... .......... A.....T... ...G.
BOGOTA63    .....G.... .....T.... .....G.... .......... A.....T... ...G.
BOGOTA67    .....G.... .....T.... .......... .......A.. A.....T... ...G.
BOGOTA70    .....G.... .....T.... .......... .......... A.....T... ...G.
BOGOTA73    .....G.... .....T.... .......... .......A.. A.....T... ...G.
BOGOTA74    .....G.... .....T.... .......... .......... A.....T... ...G.
BOGOTA77    .....G.... .....T.... .....G.... .......... A.....T... ...G.
PSEUDO1     .......... .....T.... ....T..... .......... .......... .....
PSEUDO2     .....G.... .....TT... ....T..... .......... ....T..... A....
PSEUDO3     .....G.... .....T.... ......T... .......... ......T..T ..C..
PSEUDO4     .......... .....T.... ....T..... .......... .......T.. A....
PSEUDO5     ..A....... .......... ......T..A ...T...... .T......T. T....
PSEUDO6     ..A....T.. .....T.... ....T..... ......NN.. ....T..... A....
PSEUDO7     .......T.. .....T.... ....T..... .......... ....T..... A....
PSEUDO8     .....G.... .......A.. ....T..... A......... A.....T... ...G.
PSEUDO9     .......... .......... .......... .........A ......T..T .A...
PSEUDO10    .....G.... .....T..T. .NT....... .......... .......... A....
PSEUDO11    .....G.... .......... .......... .......... ........T. T....
MIRANDA22   GG..C..... TCG....... .T....T.C. .AA.CG..G. .....C..T. ....T
MIRANDA23   GG..C....A TCGGC..... .T......C. .AA.CG..G. ...T.C..T. .....
MIRANDA24   GG..C....A TCGGC..... .T......C. .AA.CG..G. ...T.C..T. .....
MIRANDA25   GG..C....A TCGGC..... .T......C. .AA.CG..G. .....C..T. .....
PERSIMI40   GG.TC..... .C........ ......T... .......... ........T. T....
PERSIMI41   GG.TC..... .C........ ......T... ......A... ..T.....T. T....
PERSIMI42   GG.TC..... .C........ ......T... .........A ........T. T....
PERSIMI43   ....C..... .C........ ...T..T... .......... ........T. T....
PERSIMI44   ....C..... .C........ ......T... .......... ........T. T....
PERSIMI45   GG.TC..... .C........ C.....T... ......A... ........T. T....
PERSIMI46   GG.TC..... .C........ ......TT.. .......... ........T. T....
PERSIMI47   GG.TC.A... .C........ ......T... ......A... ........T. T....
PERSIMI48   GG.TC...T. .C.......A ......T... ......A... ........T. T....
PERSIMI49   GG.TC...T. .C.......A ......T... .......... ........T. T....
PERSIMI50   GG.TC..... .C........ C.....T... .......... ........T. T....
```

FIGURE 2.—Polymorphic sites of 35 samples from *D. pseudoobscura* (pseudo), *D. persimilis* (persimi), *D. p. bogotana* (bogota), and *D. miranda* (miranda). S/R/I/D, synonymous/replacment/insertion/deletion. Amino acids: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; S, serine; T, threonine; V, valine; Y, tyrosine.

persimilis. Measured by π, the level of intraspecific variation at *per* in *D. pseudoobscura* is lower than that found at other genes in the same species (Table 3). *D. persimilis* is somewhat less variable than *D. pseudoobscura* at *per*. *D. miranda* has less and *D. p. bogotana* has very little. *Per* locus for *D. pseudoobscura* and *D. persimilis* is comparable with *per* locus estimates from some of the *D. melanogaster* species complex. The estimate of π at *per* for *D. p. bogotana* is equivalent to that for *D. sechellia*, an island species in the *D. melanogaster* complex that is thought to have a very reduced effective population size (HEY and KLIMAN 1993). Most striking, however, is the finding that *D. p. bogotana* diversity at *per* is about one order of magnitude lower than those of other loci from the same subspecies. In contrast to the situation with *D. sechellia* in which low variation has been found at multiple loci (HEY and KLIMAN 1993; HILTON *et al.* 1994), low variation at per in *D. p. bogotana* does not reflect genome-wide levels of variation for this species.

Table 4 shows measures of interspecific divergence for each pair of species. Based on net divergence, it appears that *D. p. bogotana* and *D. pseudoobscura* are the most closely related taxa. However, these taxa differ by six fixed differences and share just a single polymorphism in our sample. In contrast, there are only two fixed differences but six shared polymorphisms between *D. pseudoobscura* and *D. persimilis*, a species pair that exhibits greater net pairwise divergence. There appear to be two reasons why *D. pseudoobscura* is more closely related to *D. persimilis* from the perspective of shared and fixed differences yet closer to *D. p. bogotana* from the viewpoint of net divergence. First, is that *D. p. bogotana* has little variation at *per*, and this has the effect of elevating fixed differences, and decreasing amounts of shared polymorphism. Second, is that most of the shared polymorphisms between *D. pseudoobscura* and *D. persimilis* are due to one particular sequence in the sample. The sequence from strain persimilis-40 accounts for five out of the six shared polymorphisms observed between the sibling species (Figure 2). When this strain is excluded from the comparison, the net divergence between the two species becomes 60% greater (Table 4). The pattern of variation suggest that persimilis-40 is a hybrid sequence: the first 926 bp

```
position      111111111111 1111111111 1111111111 1111111111 1111111111 111111
position    999000000000001 1111111222 3333333333 3333444444 4444444444 444555
position    299012235555772 3667899248 1555666678 8899011112 2224557888 999000
position    704578910367235 1264807376 0057234950 6934067890 1239791467 356458
S/R/I/D     ISRSSSSRSSSRRRS SIIIIIISSS SIIIIIIIII IIIIIIDDDD DDDIIIIIII IIIIII
aminoacids  -SQSENNVSNKKPPN A------GST D--------- ---------- ---------- ------
replacement..K....A...ESS. .......... .......... .......... .......... ......
Consensus   AGCGGACTCCAACCT ACACCAAATT TCGGTGAGCC CCTTTC---- ---TCTGGCG TCCCGT
BOGOTA61    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA60    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA62    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA63    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA67    ........T...... G......G.. .......A.. ..C....... ...C...... ...T..
BOGOTA70    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA73    ........T...... G......G.. .......A.. ..C....... ...C...... ...T..
BOGOTA74    ........T...... G......G.. .......A.. ..C....... ...C...... ......
BOGOTA77    ........T...... G......G.. .......A.. ..C....... ...C...... ......
PSEUDO1     G...........AG. .......... .......... .......... .......... ...T..
PSEUDO2     G.....TC....AG. .......... .......... .......... .......... ......
PSEUDO3     G...........AG. .......... .......... .......... .......... ......
PSEUDO4     ......TC....AG. .......... .......... .....A.... .......... ......
PSEUDO5     G.....TC....AG. .......... .......... .......... .......... ......
PSEUDO6     G.....TC....AG. .......... .......... .......... .......... ......
PSEUDO7     G.....TC....AG. .......... .......... .......... .......... ......
PSEUDO8     ......TC....AG. ...A..C... .A........ .......... .........A ......
PSEUDO9     G.....TC.TG.... .........A C.......... TT........ ......C... C...NN
PSEUDO10    ......TC....AG. .......... .......... .....A.... .......... ....NN
PSEUDO11    G.....TC....AG. .T......A. .......... .......... .....C..T. .....N
MIRANDA22   G..A..TC..G.... C.........A C.ATAAG.GT ...C..TTTA AGC.A..A.. .....C
MIRANDA23   G.A..GTC..G.... C.........A C.ATAAG.GT ...C..TTTA AGC.A..A.. .....C
MIRANDA24   G.A..GTC..G.... C.........A C.ATAAG.GT ...C..TTTA AGC.A..A.. .....C
MIRANDA25   G.A..GTC..G.... C.........A C.ATAAG.GT ...C..TTTA AGC.A..A.. .....C
PERSIMI40   G.....TC.TG...C C.T....... .......... .......... .......... ......
PERSIMI41   .C..A.......... .......... ......G... .......... .......... .AA.C.
PERSIMI42   .C..A.......... .......... ......G.GT ....G.TTTA AGC.A..A.. .....C
PERSIMI43   .C..A.......... ....AT.... ........T. ....G.TTTA ACC.A..A.. ......
PERSIMI44   .C..A.......... ....AT.... ........T. ....G.TTTA ACC.A..A.. ......
PERSIMI45   .C..A.......... .......... ......G.GT ....G.TTTA AGC.A..A.. .....C
PERSIMI46   .C..A.......... .......... ......G.GT ....G.TTTA AGC.A..A.. .....C
PERSIMI47   .C..A.......... .......... ......G.GT ....G.TTTA AGC.A..A.. .....C
PERSIMI48   .C..A.......... .......... .......... ....G.TTTA AGC.A..A.. ......
PERSIMI49   .C..A.......... .......... ......G.GT ....G.TTTA AGC....... ......
PERSIMI50   .C..A......G... .......... ......G.GT ....G.TTTA AGC.A..A.. .....C
```

FIGURE 2.—*Continued*

closely resemble other *D. persimilis* sequences, while the remainder are more similar to *D. pseudoobscura* sequences (Figure 2). This is graphically demonstrated in a sliding-window plot of the average pairwise divergence between persimilis-40 and the other sequences from *D. pseudoobscura* and *D. persimilis* (Figure 3). The persimilis-40 sequence has less divergence from *D. persimilis* sequences than it does from *D. pseudoobscura* sequences, in the first 1000 bp of the region. This pattern is reversed in the remainder of the sequence. In both regions, there are sections where persimilis-40 is equally divergent from both sibling species. As expected, neighbor-joining trees constructed for each of the two regions have different topologies (data not shown). For the first 926 bp, persimilis-40 is clustered with the *D. persimilis* group. For the latter portion, however, it is with *D. pseudoobscura*.

Natural selection: In recent years, several methods have been developed to test whether a pattern of DNA sequence variation has been shaped by natural selection (HUDSON *et al.* 1987; TAJIMA 1989; MCDONALD and KREITMAN 1991). All of these tests employ, as part of the null hypothesis, the neutral model of molecular evolution in which mutations are either deleterious (in which case they are removed from the population by natural selection) or neutral (KIMURA 1983). The HKA test is similar to a chi-square test and asks whether the relative levels of intraspecific polymorphism and interspecific divergences for a given locus are consistent across loci (HUDSON *et al.* 1987). We employed the HKA test by comparing the *per* data with sequence data from the alcohol dehydrogenase locus (SCHAEFFER and MILLER 1991) and the small ribosomal RNA locus of mitochondria (*SrRNA;* T. M. JENKINS, C. J. BASTEN, and W. W. ANDERSON, GenBank accession numbers X74849-75012; Table 5). In particular we focused on *D. pseudoobscura* and *D. p. bogotana,* the one species pair with data available from multiple loci (Table 3). We conducted tests with different numbers of loci and with and without the inclusion of intraspecific variation from both species (Table 5). Tests that include *per* locus data and just the *D. p. bogotana* intraspecific variation (but not the *D. pseudoobscura* intraspecific variation) are significant at the 5% level. Thus the pattern of low variation within *D. p. bogotana* at *per* but not at other loci is not consistent with the neutral model. However, the HKA tests that include polymorphism data from just *D. pseudoobscura,* or from both *D. pseudoobscura* and *D. p. bogotana,* are not significant (Table 5).

We also conducted outlier HKA tests (see MATERIALS

TABLE 2

The number of polymorphic sites within species

Species	Sample size	Syn[a]	Rep[b]	Intron[c]	Total	θ[d]	Tajima D[e]
Drosophila p. bogotana	9	1	1	1	3	1	0.60
D. pseudoobscura	11	22	7	20	49	16	−1.2
D. miranda	4	6	2	1	9	5	−0.49
D. persimilis	11	12	9	15	36	12	−0.76

[a] Synonymous changes in exons, total exon length is 1131 bp.
[b] Replacement change in exons.
[c] Total intron length is 345 bp.
[d] θ is estimated by $S/\Sigma(1/I)$, $I = 1, \ldots, N_r-1$, S, total number of polymorphic sites, n, sample size. (Watterson 1975).
[e] Tajima (1989).

and methods) with data from *D. pseudoobscura* and *D. p. bogotana*. In a test that included data from *per* and *Adh*, the MSD score occurred for *per* variation within *D. p. bogotana*. The observed value was 2.55, and the probability of observing a higher value was estimated by simulation to be 0.024. In an outlier test using *per*, *Adh*, and *SrRNA*, the MSD also occurred for *per* variation within *D. p. bogotana*. In this case the value was 2.513, and the probability of observing a higher value was estimated to be 0.034 (Table 6). In a test with just *Adh* and *SrRNA*, the MSD score was 0.36 for *SrRNA* in *D. p. bogotana*, and the probability of observing a higher value was estimated to be 0.429.

Table 2 shows values of Tajima's D (Tajima 1989). If population size has been changing or natural selection has altered the frequency distribution of polymorphisms from that expected under the null model, then D will be different from 0. None of the observed values of D are significantly different from 0 (Table 2).

Yet another kind of test asks whether the ratio of synonymous to replacement variation appears to be the same for both intraspecific polymorphisms and interspecific fixed differences (McDonald and Kreitman 1991). For most species pairs in this study, the McDonald-Kreitman test can not be carried out because of an absence of fixed replacement differences between species. For contrasts involving *D. miranda*, the McDonald-Kreitman test revealed no significant departures from the neutral model (data not shown).

Recombination, gene tree estimates, and gene flow: We applied the counting method of Hudson and Kaplan (1985) and found evidence of a minimum of four recombination events at *per* in *D. pseudoobscura* and three in *D. persimilis*. We also applied the method of Hudson (1987) to estimate the population recombination parameter, $4Nc$ ($3Nc$ for sex-linked loci; c, the recombination rate per generation for the sequenced region). These estimates are 37.9 for *D. pseudoobscura* and

TABLE 3

Comparison of the *period* locus of *D. pseudoobscura* and the sibling species with some previous studies

Species	*Period* sex-linked	*Adh* autosomal	*Xdh* autosomal	*Est-5B* sex-linked	*SrRNA* mitochondrial
D. pseudoobscura	8.4	11.7[a]	12.0[b]	12.2[c]	22.6[d]
D. persimilis	5.8[e]	—	—	—	—
D. p. bogotana	0.9	8.4[a]	—	—	9.4[d]
D. miranda	3.2	—	—	—	—
D. melanogaster	6.2[f]	8.1[g]	—	—	—
D. simulans	11.5[f]	16.4[h]	—	—	—
D. mauritiana	11.8[f]	16.6[i]	—	—	—
D. sechellia	0.9[f]	—	—	—	—

π (per base pair, × 1000) Observed average pairwise differences among DNA sequences or nucleotide diversity (Nei 1987).
[a] Schaeffer and Miller (1991).
[b] Riley *et al.* (1992).
[c] Veuille and King (1995).
[d] T. M. Jenkins, C. J. Basten, and W. W. Anderson, GenBank accesson numbers X74849-75012.
[e] Sample persimilis-40 excluded.
[f] Kliman and Hey (1993).
[g] Kreitman (1983).
[h] Bodmer and Ashburner (1984).
[i] Cohn *et al.* (1984).

TABLE 4

Number of fixed polymorphisms and net average pairwise divergence

Species	1	2	3	4
1. *D. p. bogotana*		6 (1)	43 (0)	16 (0)
2. *D. pseudoobscura*	5.0		26 (1)	2/4[a] (6/1[a])
3. *D. miranda*	30.0	23.5		18 (0)
4. *D. persimilis*	12.3	9.7/15.6[a]	19.8	

[a] Persimilis-40 is excluded.

Number of shared polymorphisms are shown in parentheses. Net distance between *D. p. bogotana* and the other species are corrected for the low diversity of *D. p. bogotana* NEI (1987). Divergence values are in the lower triangle, × 1000.

23.7 for *D. persimilis*. However, these are rough estimates, as the method is not expected to be very accurate for sample sizes of <20 sequences (HUDSON 1987).

A neighbor-joining distance tree (SAITOU and NEI 1987) was constructed for the 35 samples and the confidence levels for various clades are indicated by bootstrapping values (Figures 4 and 5). With the noteworthy exceptions of the branch leading to the *D. p. bogotana* sequences and the branch leading to the *D. miranda* sequences, most of the branches on the gene tree estimate have low bootstrapping values (Figure 5). This pattern is expected if the sequences of *D. pseudoobscura* and *D. persimilis* have histories that include multiple recombination events. When the genealogical history for a locus has included recombination, then different portions of the locus have different gene trees, and an analysis that imposes a single tree estimate cannot be taken as an estimate of the true genealogy. In effect much of the tree in Figure 4, especially those portions involving sequences from *D. pseudoobscura* and *D. persimilis*, must be considered only as a summarization of the pattern of observed divergences among sequences and not as an estimate of the genealogy.

Among samples from *D. miranda* and *D. p. bogotana*, the sequences do cluster with their respective groups, and each of these groups are exclusive of sequences from other groups (Figure 4). However, apart from these groupings, the tree diagram does not suggest a sister taxa relationship among species. In particular, *D. p. bogotana* sequences do not form a sister group to *D. pseudoobscura* but rather appear to be more closely related to some sequences of *D. pseudoobscura* than to others. This pattern is consistent with a speciation model in which *D. p. bogotana* was recently formed by a small isolate from a much larger population of *D. pseudoobscura* that still segregates old lineages. However, it is also consistent with a recent selective sweep, at or near *per*, that was limited to *D. p. bogotana*. The relationship between *D. pseudoobscura* and *D. persimilis* is complicated by the *D. persimilis*-40 sequence. In Figure 4, persimilis-40 appears as a sister group to a group that includes all other sequences (excluding those of *D. miranda* and collectively the sequences of *D. persimilis* do not form a monophyletic group. However, persimilis-40 may be a hybrid sequence formed by gene flow and

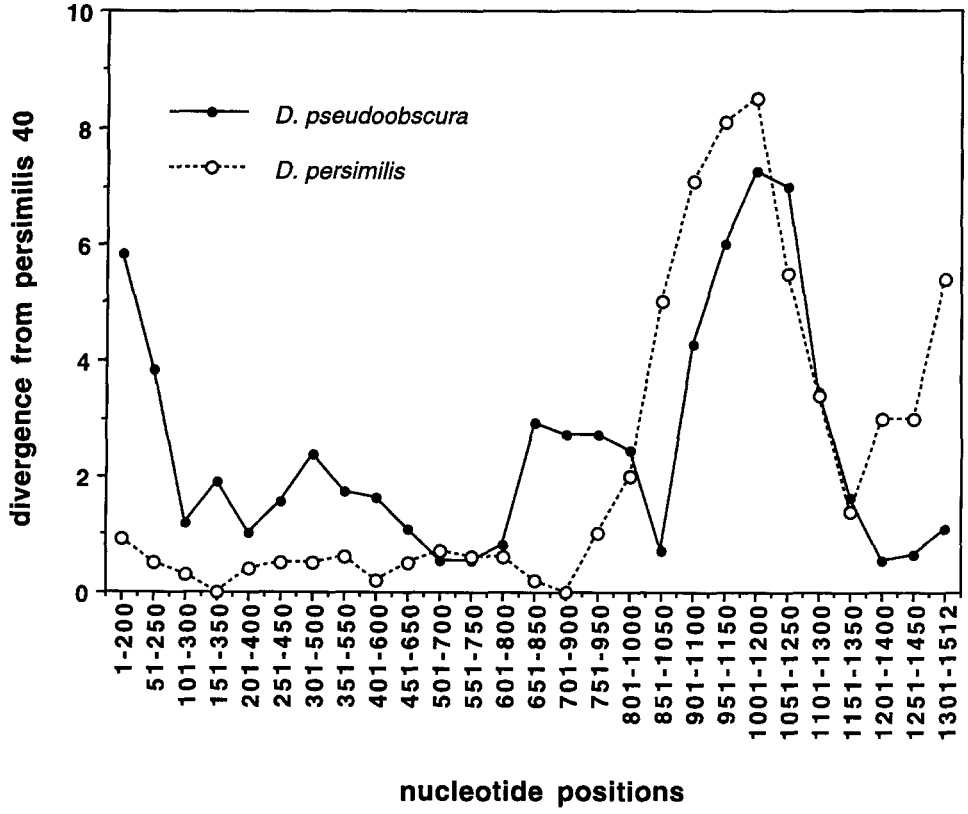

FIGURE 3.—A sliding window plot of average pairwise divergence of *D. persimilis* (excluding the sample persimilis-40) and *D. pseudoobscura* from persimilis-40. Window size is 200 nucleotides.

TABLE 5

The HKA test of natural selection for *D. pseudoobscura* and *D. p. bogotana*

Species	*Period*	*Adh*[a]	*SrRNA*
	Diversity and divergence[b]		
θ within *D. p. bogotana*	0.75	7.5	7.6
θ within *D. pseudoobscura*	11.2	11.9	24.6
interspecific divergence[c]	13.4	12.7	17.3

Test combinations	Loci included	χ^2	d.f.	Probability
D. p. bogotana only	*period, Adh*	4.85	1	0.028*
	period, Adh, SrRNA	6.15	2	0.046*
D. pseudoobscura only	*period, Adh*	0.04	1	0.845
	period, Adh, SrRNA	0.12	2	0.943
D. p. bogotana + *D. pseudoobscura*	*period, Adh*	3.39	2	0.184
	period, Adh, SrRNA	6.63	4	0.157

[a] Estimated from SCHAEFFER and MILLER (1991).
[b] Per basepair × 1000.
[c] Average pairwise differences between species.

recombination. With persimilis-40 excluded, the tree is more consistent with the established species phylogeny (GODDARD *et al.* 1990).

The tree estimate also suggests large amounts of gene flow within *D. pseudoobscura* and *D. persimilis* (Table 1, Figure 4) as there does not appear to be a tendency for samples to cluster by their geographic origin. The amount of gene flow, both inter- and intraspecific, can be estimated under a model of population subdivision using F_{ST} and the average pairwise differences within and between populations (WRIGHT 1951; SLATKIN 1991; HUDSON *et al.* 1992). The estimate of *Nm* between *D. pseudoobscura* and *D. p. bogotana* is 0.13 and that between *D. pseudoobscura* and *D. persimilis* 0.20 (0.17 without persimilis-40). For a rough estimate of intraspecific gene flow, the samples from *D. pseudoobscura* and *D. persimilis* were both split into a "northern" and a "southern" group. For *D. pseudoobscura,* the *Nm* between the northern (samples 3, 10, 11) and the southern (the rest) is 17.6. And for *D. persimilis,* the *Nm* for the northern (samples 42, 43, 45, 50) and the southern (the rest) is 16.1. In a very large study of the Adh locus in multiple populations of *D. pseudoobscura,* estimates of *Nm* varied from one to 12, depending on which populations were compared (SCHAEFFER and MILLER 1992). In general, a *Nm* value ≥ 1 leads to considerable homogeneity among populations while population structuring and divergence can occur when *Nm* is <1 (WRIGHT 1940). These estimates of *Nm* also assume an equilibrium between a steady level of gene flow and divergence. This assumption may be appropriate for the within species estimates. For the interspecific estimates, however, current levels of shared variation may be remnants from the time of speciation, in which case the *Nm* estimates could be much greater than current levels of gene flow between species.

TABLE 6

Three locus HKA test with *D. pseudoobscura* and *D. p. bogotana*

	D. pseudoobscura		*D. p. bogotana*		
Locus	*S*	*n*	*S*	*n*	Divergence
period					
Observed	48	11	3	9	20
Expected[a]	41.3		16.6		12.8
Deviation[b]	0.129		2.513*		0.413
Adh					
Observed[c]	114	10	66	8	37
Expected	127.5		50.6		38.7
Deviation	0.056		0.379		0.003
SrRNA					
Observed[d]	43	15	10	7	9
Expected	36.2		11.8		14.3
Deviation	0.200		0.070		0.32

S, number of polymorphic sites; *n*, number of DNA sequences in the sample. * Statistically significant in an outlier HKA test, $P = 0.034$; see text.
[a] Expected values are calculated using expressions (1) and (3) of HUDSON *et al.* (1987).
[b] Deviations are calculated using expressions (2) and (4) and the expression for χ^2 in HUDSON *et al.* (1987).
[c] Data from SCHAEFFER and MILLER (1991).
[d] Data from T. M. JENKINS, C. J. BASTEN, and W. W. ANDERSON, GenBank accesson numbers X74849–75012.

Time of introgression: One sequence, persimilis-40, contains a stretch of 586 bases (from position 927 to the 3′ end, Figures 2 and 3) that seems typical of *D. pseudoobscura* sequences and is divergent with respect to *D. persimilis* sequences. To estimate the time of introgression, we assume that the pattern arose via introgression of a single sequence from *D. pseudoobscura* into *D. persimilis,* followed by one or more recombination events. Two different approaches to estimating the time of introgression were considered. In the following: *I* is

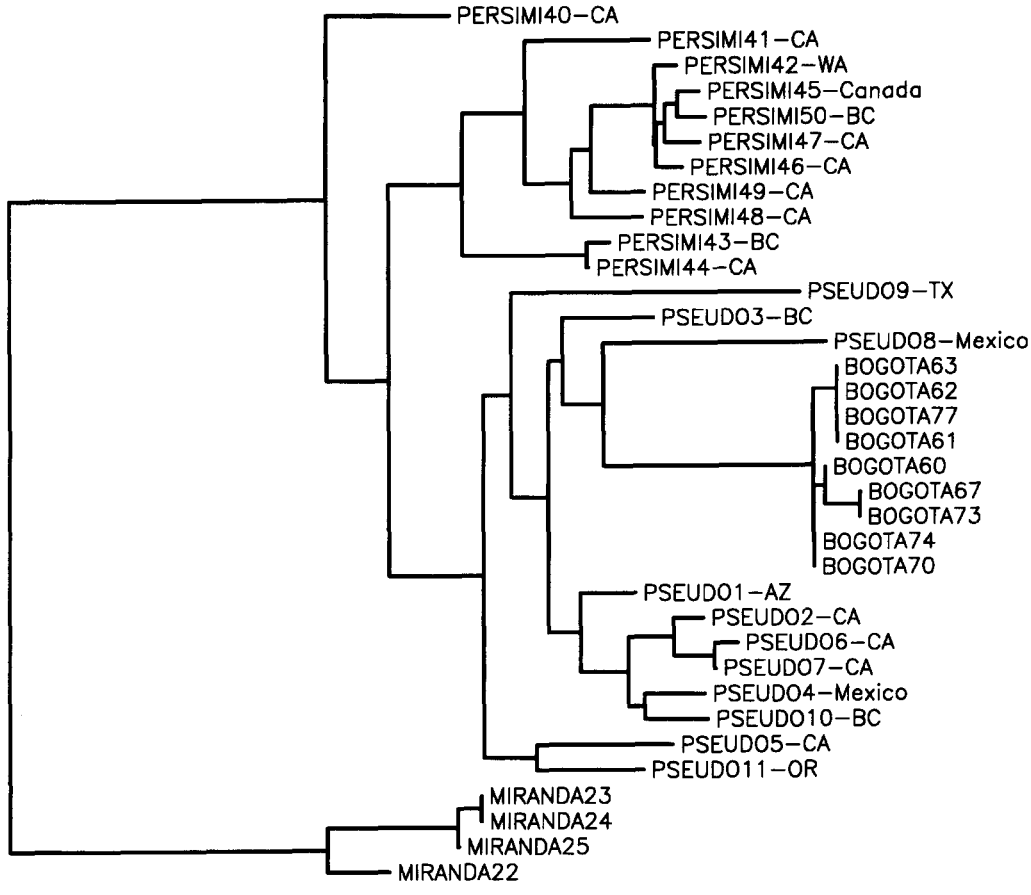

FIGURE 4.—A neighbor-joining distance tree. The sample numbers are followed by their geographic locations: CA, California; WA, Washington; BC, British Columbia; TX, Texas; AZ, Arizona; OR, Oregon. The scale and the units of individual branch lengths can be assessed by comparison with the branch that leads to PERSIMI40, which corresponds to 0.0035 substitutions per base pair.

the time of introgression in units of generations; N_r is the effective population size of *D. persimilis*; and N_s is the effective population size of *D. pseudoobscura*.

The first method is based on the divergence of the 586-bp region in persimilis-40 from *D. pseudoobscura* sequences. Let D be the average number of differences between persimilis-40 and *D. pseudoobscura* sequences that is expected to be observed for the introgressed portion. Under a simple neutral model, in which the amount of variation within *D. pseudoobscura* at the time of introgression is the same as is currently present,

$$D = 2Iu + 3Nu. \quad (1)$$

In other words, D is equal to the divergence that has accumulated since introgression plus divergence that was present at the time of introgression. For the 586-bp region, D is estimated to be 8.82, and $3N_s u$ is estimated to be 4.87. Then from (1), Iu is equal to 1.975. Since 4.87 mutations accumulate in $3N_s$ generations, the estimated time of introgression is $1.2N_s$ generations (*i.e.*, $1.975 * 3N_s/4.87$).

Introgression time can also be estimated using a model of recombination. Given that a region of 586 bases (out of 1512) has been left intact following one or more recombination events, it is possible to generate a maximum likelihood estimate of the number of recombination events since the time of the introgression of the *D. pseudoobscura* sequence. Assuming that our attention has fallen on the largest nonrecombining fragment of the introgressed sequence (a conservative assumption, since it is possible that a larger portion may have occurred in a larger sample of sequences), then use can be made of an expression in FELLER (1971, expression 9.9, p. 28) for the probability distribution of the length, a, of the longest unbroken portion of an interval that is randomly divided into a number of subintervals. The derivative of FELLER's (1971) expression (9.9), taken with respect to a, is the probability density of a and is a function of the number of random breaks. In the present context, setting a equal to 0.387 (586/1512) yields a maximum likelihood estimate of five recombination events since the time of introgression. Thus the interval between the expected times for five and six recombination events can be used as an estimate of the time since introgression. The population recombination rate for *D. persimilis*, $3N_r c$, was estimated to be 23.7 (see Recombination, gene tree estimates, and gene flow). If persimilis-40 is excluded from the calculations for this estimate, the value rises to 28. This is an estimate of the number of recombination events per $3N_r$ generations per 1512 bases of *per*. If we take the time of introgression as the midpoint between the expected time for five and six recombination events, then the time since introgression is estimated to be $0.6N_r$ generations ($5.5 * 3N_r/28$).

The two methods yield estimates of $1.2N_s$ generations

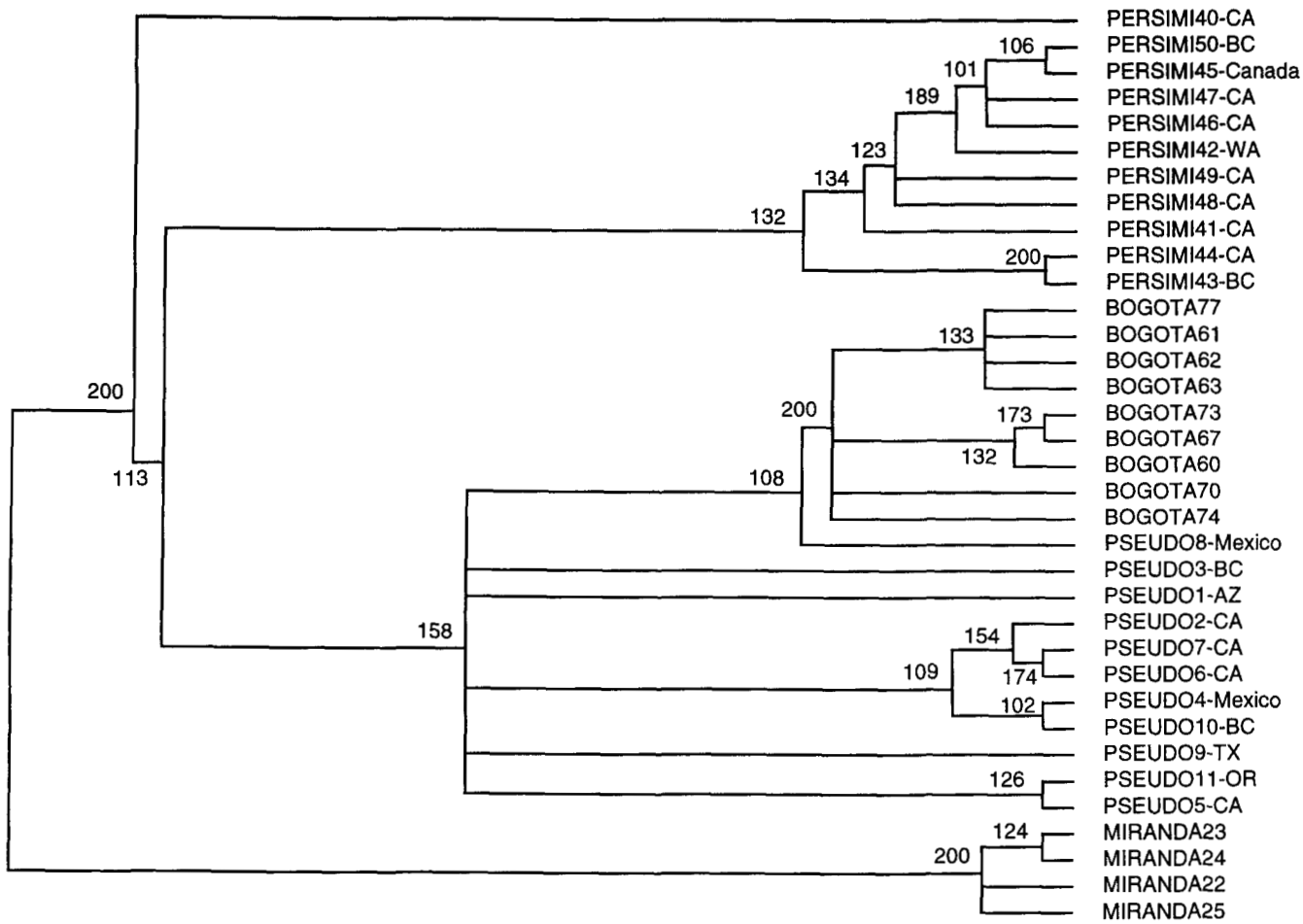

FIGURE 5.—A neighbor-joining consensus tree representing 200 bootstraps. The nodes with values <100 are not shown.

and $0.6N_r$ generations. To place these values in context, it is necessary to estimate the time since speciation between *D. pseudoobscura* and *D. persimilis*. Consider a simple speciation model in which the net divergence between species is proportional to the time since speciation (HUDSON *et al.* 1987; HEY 1994). Under this simple model, net divergence is equal to $2Tu$, where T is the time, in generations, since speciation. Between *D. pseudoobscura* and *D. persimilis* the net divergence (excluding persimilis-40) is 15.6 (Table 4) so that the estimate of the product of speciation time and mutation rate, Tu, is 7.8. Within *D. persimilis*, the average pairwise divergence (excluding persimilis-40) is 8.58. Since the average pairwise divergence can be taken as an estimate of $3N_ru$, the time since speciation is estimated to be $\sim 2.7N_r$ generations (*i.e.*, $3N_r * 7.8/8.58$). Within *D. pseudoobscura* there are estimated to be 12.22 mutations per $3N_s$ generations so that the time since speciation is estimated to be $1.91N_s$ generations ($3N_s * 7.8/12.22$). The *D. persimilis* recombination estimate of the time of introgression ($0.6N_r$ generations) is ~22% of the time since speciation (0.6/2.7). The *D. pseudoobscura* divergence estimate of the time of introgression ($1.2N_s$ generations) is ~63% of the time since speciation (1.2/1.91). This analysis has included several simplifying assumptions, and has dealt only in expectations. Consequently, the estimated times of introgression do not have confidence intervals. However, the analyses show that the data is most consistent with a historical model in which an introgression from *D. pseudoobscura* to *D. persimilis* occurred well after the two species began to diverge. Both methods also suggest that the introgression was not a recent event.

Time of divergence: If a molecular clock is assumed along with a yearly substitution rate based on fossil evidence, it is possible to estimate the approximate divergence times among species. The divergence between *D. melanogaster* and *D. obscura* was estimated to be 30–40 mya based on amber fossil evidence (THROCKMORTON 1975). Using these estimates, and a measure of divergence between these species for synonymous sites at several genes, SHARP and LI (1989) proposed a rate of 1.6×10^{-8} substitutions per synonymous site per year for genes with low codon bias and one half that rate for high-bias genes. We take an average of 1.2×10^{-8}. When applied to the net synonymous site divergence estimates for the species in this study, we estimate the divergence between *D. miranda* and the rest of the

group to be ~2 mya; the date separating *D. pseudoobscura* and *D. persimilis* to be ~1 mya (1.1 mya excluding persimilis-40); and that between *D. pseudoobscura* and *D. p. bogotana* ~0.5 mya (Table 7). These estimates are based on net divergence at synonymous sites which is the absolute divergence minus the average of the amount of variation within each of the species compared (NEI 1987, p. 276). For the case of contrasts involving *D. p. bogotana,* because of the near absence of variation within this species, net divergence was calculated as absolute divergence minus the amount of intraspecific variation within just one species (the one *D. p. bogotana* was compared with).

DISCUSSION

Diversity: The estimates of nucleotide diversity at per for *D. pseudoobscura* are similar to, though a bit lower than, estimates from other loci (Table 3). For *D. persimilis* this is the first study of DNA sequence variation in a nuclear gene and the amount is substantial. Under a neutral model, both π and θ have an expected value of *4Nu* (*3Nu* for a sex-linked locus), so these levels of variation probably reflect histories of large effective population sizes. In neither of these species was evidence found for a departure from neutrality at *per,* and the data suggest a history of large long-term effective population sizes in both species. Data from *Adh* are consistent with this interpretation. Although only a single haplotype was studied from *D. persimilis* at *Adh,* it fell within the gene tree estimate for a larger sample of highly variable *D. pseudoobscura* haplotypes (SCHAEFFER *et al.* 1987).

This is also the first study of nuclear DNA sequence variation in *D. miranda,* which revealed considerable variation though not as much as *D. pseudoobscura* and *D. persimilis. D. p. bogotana,* however, revealed little variation, and *per* locus diversity estimates in this species are about one order of magnitude lower than those of other loci from the same species, and from the same locus of *D. pseudoobscura.* At *per, D. p. bogotana* is similar to *D. sechellia,* an island species with small effective population size (HEY and KLIMAN 1993).

The divergence between *D. pseudoobscura* and *D. p. bogotana*: *D. pseudoobscura* and *D. p. bogotana* have also been extensively studied for 3700 bp around the *Adh* locus (SCHAEFFER and MILLER 1991, 1992). Considerable variation and recombination are apparent at both loci within *D. pseudoobscura.* However the two loci differ markedly in *D. p. bogotana,* with *per* exhibiting much less variation than *Adh* (Table 3). The two loci also differ in the pattern of divergence between *D. pseudoobscura* and *D. p. bogotana.* At *Adh* the two species exhibited no fixed differences and 42 of the 66 polymorphisms that were observed in *D. p. bogotana* were also observed in *D. pseudoobscura* (SCHAEFFER and MILLER 1992). At *per,* there were six fixed differences between the species, and one of the three polymorphisms that were observed in *D. p. bogotana* also occurred in *D. pseudoobscura.* However, there was no evidence of recent gene flow at *per* between *D. pseudoobscura* and *D. p. bogotana* in agreement with what has been found for the mitochondrial genome (POWELL 1983). Also consistent is the *Nm* estimate of only 0.13 between the two (sub)-species at *per.* In contrast, we estimate from SCHAEFFER and MILLER's *Adh* data (1992) that the corresponding *Nm* for this locus is 0.55.

Based on HKA tests (Tables 5 and 6), variation at *per* in *D. p. bogotana* appears to have been reduced by recent natural selection at or near the *per* gene. The low variation at *per* in this species is not consistent with either a very small effective population size in *D. p. bogotana* or a slower mutation rate at *per.* A small population size should reduce variation across an entire genome, and then all the loci sampled from this species should have reduced genetic variation, and this is not the case (Table 3). If the low diversity at *per* for *D. p. bogotana* had been caused by a slower mutation rate at this locus, we would also expect lower variation at *per* in other closely related species, but this is not observed. We are left with natural selection as the simplest explanation for low *per* variation in *D. p. bogotana.* The pattern is consistent with a selective sweep, the selective fixation of a beneficial mutation that has caused a loss of DNA sequence variation in that portion of the genome that is tightly linked to the beneficial mutation. Natural selection could have acted within the sequenced region or it could have acted on a nearby location and have affected the per locus by genetic hitchhiking (MAYNARD-SMITH and HAIGH 1974). It is unlikely that natural selection in *D. p. bogotana* at *Adh* (which was also included in the HKA tests) has caused the significant departure from the neutral model. Under this alternative explanation, natural selection may have acted to elevate variation at *Adh* within *D. p. bogotana.* However, at *Adh* both *D. p. bogotana* and *D. pseudoobscura* have experienced considerable recombination (SCHAEFFER and MILLER 1991), which means that natural selection acting on any one base position can have only limited effect on flanking positions (HUDSON and KAPLAN 1988). The observations of considerable recombination at nuclear genes in *D. pseudoobscura* and at *Adh* in *D. p. bogotana* also mitigates the likelihood that other sources of variation have caused the significant findings. The HKA test assumes a particular model for the variation among independent loci and employs the assumption that recombination does not occur within loci. If the true history of a species has been such that variation among independent loci is augmented somehow, such as by changing population structure or just by chance, then an HKA test may be significant even in the absence of selective sweeps or balancing selection. However, the assumption of no recombination is conservative (HUDSON *et al.* 1987). If recombination is actually common within the

TABLE 7

Net average pairwise divergence at synonymous sites and estimated divergence time between species

Species	1	2	3	4
1. *D. p. bogotana*		3.18 (0.012)	16.27 (0.062)	10.33 (0.039)
2. *D. pseudoobscura*	0.5		11.35 (0.043)	6.29/6.83[c] (0.024/0.026[a])
3. *D. miranda*	2.6	1.8 ± 1.22		12.54 (0.048)
4. *D. persimilis*	1.6	1.0 ± 1.0 (1.1 ± 0.93[a])	2.0 ± 1.0	

Pairwise divergences are stated as divergence per synonymous site in the upper (right) triangle; divergence times are means ± SE in million years ago, lower (left) triangle. Divergence time is estimated based on a rate of 1.2×10^{-8} substitutions per synonymous site per year (SHARP and LI 1989) and the presence of 265 synonymous positions within the sequenced region. The standard error was calculated following TAKAHATA and NEI (1985) for net average pairwise divergence at synonymous sites. The divergence between *D. p. bogotana* and the other species was corrected for the low diversity within *D. p. bogotana* and the standard error was not calculated for these values.

[a] Sequence persimilis-40 excluded.

regions that have been sequenced and that are subjected to an HKA test, then the variance assumed for each locus will tend to be too high. In effect, recombination causes loci to act like multiple loci, so that a measure of variation that sums across a region with a history of recombination will have a variance that is reduced by the presence of recombination (HUDSON 1983).

Given what is known of the function of the *per* locus in *D. melanogaster*, it seems possible that the gene could be a candidate for natural selection at times of speciation. The *per* locus is known to be involved in the control of interspecific differences in male courtship songs (KYRIACOU and HALL 1980, 1989; KYRIACOU 1990). The region we have sequenced is adjacent to the Thr-Gly-like repeats, and this and the downstream flanking region have been explicitly implicated in such control (YU *et al.* 1987; WHEELER *et al.* 1991). However, *D. p. bogotana* does exhibit polymorphisms in the region between the Thr-Gly-like repeats and the region that is most divergent (between positions 618 and 1504) from *D. pseudoobscura*. It is possible that divergence at *per* may be associated with behavioral differences and contribute to reproductive isolation between *D. pseudoobscura* and *D. p. bogotana*. To our knowledge, however, there has been no experimental demonstration of interspecific differentiation in male courtship songs between them.

The divergence between *D. pseudoobscura* and *D. persimilis*: One sequence, persimilis-40, appears to be a hybrid between *D. persimilis* and *D. pseudoobscura* sequences. Two methods of estimating the time of the introgression, by divergence data and by the estimated number of associated recombination events, suggest that the introgression occurred well after the initial divergence of *D. persimilis* and *D. pseudoobscura*. This finding is consistent with evidence of limited gene flow from chromosomal (DOBZHANSKY 1973) and mitochondrial data (POWELL 1983; HALE and BECKENBACH 1985), and it seems possible that there has been a small degree of ongoing gene flow between the two siblings since their speciation.

Speciation via founder event probably did not occur in the divergence of *D. pseudoobscura* and *D. persimilis*. Both species reveal evidence of large long-term effective population sizes, and the pattern of introgression suggests that they may have never been completely isolated from each other, geographically or genetically. It remains possible that these species arose via sympatric speciation, a history that is supported by the geographic distribution and the pattern of chromosome inversion polymorphisms. The range of *D. persimilis* is entirely included within that of *D. pseudoobscura*, and both species segregate multiple inversion polymorphisms on the third chromosome (reviewed in POWELL 1992). However, only one arrangement, "Standard", is shared by both species. The Standard arrangement is relatively recent among the inversions in *D. pseudoobscura* but it appears to be the oldest arrangement in *D. persimilis*. The implication is that *D. persimilis* third chromosomes are derived from the Standard arrangement, an argument strongly favored by both allozyme and molecular data (PRAKASH and LEWONTIN 1968; AQUADRO *et al.* 1991). In *D. pseudoobscura*, the historical development of inversion polymorphisms is closely correlated with their geographic distributions (POWELL 1992), with the most ancestral type found dominating in southern Mexico and Guatemala and the more recent ones gradually extend northward in the western third of the U.S., forming a fairly good temporal gradient along the north-south axis. This pattern may not have been disrupted, since no glaciation occurred in this region during the Pleistocene era (MIELKE 1989), and it is possible that the geographic range of *D. pseudoobscura* has changed little in the past two million years. Today, the Standard arrangement of *D. pseudoobscura* is found in the highest frequency along the Pacific Coast, from southern Cali-

fornia to British Columbia, exactly matching the distribution of *D. persimilis,* the inversions of which are either the Standard or its direct derivatives (POWELL 1992).

Evidence that introgression occurred at *per* after the separation of *D. pseudoobscura* and *D. persimilis* argues against a direct contribution by *per* to their reproductive isolation. If reduced fitness of hybrids were due in part to *per* locus differences, then selection would prevent the introgression of *per* locus sequences. However, this argument does not preclude a mechanism where other sex-linked loci play a dominant role in the initial development of reproductive isolation.

Time of divergence: Estimates of the age of divergence between *D. pseudoobscura* and *D. p. bogotana* range from millions of years (AYALA and DOBZHANSKY 1974) to decades (PRAKASH 1972). Our estimate is 0.5 mya, assuming a substitution rate of 1.2×10^{-8} per synonymous site per year. However, this estimate is problematic because of the suspected directional selection or selective sweep at the *per* locus of *D. p. bogotana,* which will significantly alter the number of segregating sites and the divergence between *D. p. bogotana* and the others. SCHAEFFER and MILLER (1991) estimated a divergence time of 0.15 mya from the pattern of divergence at *Adh.* Our estimates of divergence times among *D. miranda, D. pseudoobscura,* and *D. persimilis* are not subject to this difficulty and may therefore be more accurate. These estimates are also similar to independent time estimates based on inversion polymorphisms. Of the 36 types of inversions on the third chromosome in *D. pseudoobscura* and 17 in *D. persimilis,* Standard is the only one shared by both species, therefore predating their speciation (POWELL 1992). Molecular data suggest this inversion arises ~1.7 mya (AQUADRO *et al.* 1991). Our divergence figures, 2 mya between *D. miranda* and the others and 1 mya between *D. pseudoobscura* and *D. persimilis,* are therefore compatible with the age of inversion polymorphisms.

We thank DANIEL WEINREICH for providing us several strains of *D. p. bogotana.* We are grateful to JOHN WAKELEY and HOLLY HILTON for helpful discussions. This study was funded in part by a grant from the National Science Foundation to J.H. (DEB-9306625) and a postdoctorate fellowship to R.L.W. by the Bureau of Biological Research, Rutgers University.

LITERATURE CITED

AQUADRO, C. F., A. L. WEAVER, S. W. SCHAEFFER and W. W. ANDERSON, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura*: the *amylase* gene region. Proc. Natl. Acad. Sci. USA **88:** 305–309.

ASHBURNER, M., 1989 Protocol 48:preparation of DNA from single flies, pp. 108–109 in *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

AVISE, J. C., and R. M. BALL, JR., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 45–67.

AYALA, F. J., and TH. DOBZHANSKY, 1974 A new subspecies of *Drosophila D. pseudoobscura.* Pan Pacific Entomol. **50:** 211–220.

BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature **309:** 425–430.

COHN, V. H., M. A. THOMPSON and G. P. MOORE, 1984 Nucleotide sequence comparison of the Adh gene in three Drosophilids. J. Mol. Evol. **20:** 31–37.

COLOT, H. V., J C. HALL and M. ROSBASH, 1988 Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of nonconserved coding DNA. EMBO J. **7:** 3929–3737.

DOBZHANSKY, TH., 1935 *Drosophila miranda,* a new species. Genetics **20:** 377–391.

DOBZHANSKY, TH., 1973 Is there gene exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in their natural habitats? Am. Nat. **107:** 312–314.

DOBZHANSKY, TH., and C. EPLING, 1944 Contributions to the genetics, taxonomy, and ecology of *Drosophila pseudoobscura* and its relatives. Carnegie Inst. Washington Publ. **554:** 1–183.

DOBZHANSKY, TH., A. S. HUNTER, O. PAVLOVSKY, B. SPASSKY and B. WALLACE, 1963 Genetics of an isolated marginal population of *Drosophila pseudoobscura.* Genetics **48:** 91–103.

EDERY, I., J. E. RUTILA and M. ROSBASH, 1994 Phase shifting of the circadian clock by induction of the *Drosophila period* protein. Science **263:** 237–240.

FELLER, W. 1971 *An Introduction to Probability Theory and its Applications,* Vol II, Ed. 2. John Wiley and Sons, New York.

FELSENSTEIN, J., 1993 Phylip. Phylogeny Inference Package, version 3.5.

GODDARD, K., A. CACCONE and J. R. POWELL, 1990 Evolutionary implications of DNA divergence in the *Drosophila obscura* group. Evolution **44:** 1656–1670.

HALE, L. R., and A. T. BECKENBACH, 1985 Mitochondrial DNA variation in *Drosophila pseudoobscura* and related species in pacific Northwest populations. Can. J. Genet. Cytol. **27:** 357–364.

HEY, J. 1994 Bridging phylogenetics and population genetics with gene tree models, pp. 435–449 in *Molecular Approaches to Ecology and Evolution,* edited by B. SCHIERWATER, B. STREIT, G. WAGNER and R. DESALLE. Birkhäuser-Verlag, Basel.

HEY, J., and R. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

HILTON, H., R. KLIMAN and J. HEY, 1994 Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48:** 1900–1913.

HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

HUDSON, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 819–829.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

KIMURA, M. 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

KLIMAN, R. M., and J. HEY, 1993 DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133:** 375–387.

KONOPKA, R. J., and S. BENZER, 1971 Clock mutants of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **68:** 2112–2116.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster.* Nature **304:** 412–417

KYRIACOU, C. P., 1990 The molecular ethology of the period gene in *Drosophila.* Behav. Genet. **20:** 191–211.

KYRIACOU, C. P., and J. C. HALL, 1980 Circadian rhythm mutations in *Drosophila melanogaster* affect short term fluctuations in the male's courtship songs. Proc. Natl. Acad. Sci. USA **77:** 6729–2733.

KYRIACOU, C. P., and J. C. HALL, 1982 The function of courtship song rhythms in *Drosophila.* Anim. Behav. **30:** 794–801.

KYRIACOU, C. P., and J. C. HALL, 1984 Learning and memory mutations impair acoustic priming of mating behavior in *Drosophila.* Nature **308:** 62–65.

KYRIACOU, C. P., and J. C. HALL, 1989 Spectral analysis of Drosophila courtship songs. Anim. Behav. **37:** 850–859.

KYRIACOU, C. P., M. L. GREENACRE, M. G. RITCHIE and B. C. BYRNE, 1992 Genetic and molecular analysis of the love song preferences of *Drosophila* females. Am. Zool. **32:** 31–39.

LEWONTIN, R. C., J. A. MOORE, W. B. PROVINE and B. WALLACE, 1981 *Dobzhansky's Genetics of Natural Populations.* Columbia University Press, New York.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–35.

MAYR, E., 1946 Experiments on sexual isolation in Drosophila. VII. The nature of the isolating mechanisms between *Drosophila pseudoobscura* and *Drosophila persimilis.* Proc. Natl. Acad. Sci. USA **32:** 128–137.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the Adh locus in *Drosophila.* Nature **351:** 652–654.

MIELKE, H. W., 1989 *Patterns of Life—Biogeography of a Changing World,* Unwin Hyman, Boston.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

POWELL, J. R., 1983 Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila.* Proc. Natl. Acad. Sci. USA **80:** 492–495.

POWELL, J. R., 1992 Inversion polymorphism in *Drosophila pseudoobscura* and *Drosophila persimilis,* pp. 73–126 in *Drosophila Inversion Polymorphism,* edited by C. B. KRIMBAS and J. R. POWELL. CRC Press, Boca Raton, FL.

PRAKASH, S., 1972 Origin of reproductive isolation in the absence of apparent genic differentiation in a geographic isolate of *Drosophila pseudoobscura.* Genetics **72:** 143–155.

PRAKASH, S., and R. C. LEWONTIN, 1968 A molecular approach to the study of genic heterozygosity in natural populations. III. Direct evidence of coadaptation in gene arrangements of *Drosophila.* Proc. Natl. Acad. Sci. USA **59:** 398–405.

RILEY, M. A., S. R. KAPLAN and M. VEUILLE, 1992 Nucleotide polymorphism at the *Xanthine Dehydrogenase* locus in *Drosophila pseudoobscura.* Mol. Biol. Evol. **9:** 56–69.

RITCHIE, M. G., and C. P. KYRIACOU, 1994 Reproductive isolation and the *period* gene of *Drosophila.* Mol. Ecol. **3:** 595–599.

ROSBASH, M., and J. C. HALL, 1989 The molecular biology of circadian rhythms. Neuron **3:** 387–398.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Bio. Evol. **4:** 406–425.

SCHAEFFER, S. W., and E. L. MILLER, 1991 Nucleotide sequence analysis of Adh genes estimates the time of geographic isolation of the Bogota population of *Drosophila D. pseudoobscura.* Proc. Natl. Acad. Sci. USA **88:** 6097–6101.

SCHAEFFER, S. W., and E. L. MILLER, 1992 Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. Genetics **132:** 471–480.

SCHAEFFER, S. W., C. F. AQUADRO and W. W. ANDERSON, 1987 Restriction-map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura.* Mol. Biol. Evol. **4:** 254–265.

SHARP, P. M., and W. H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28:** 398–402.

SLATKIN, M., 1991 Inbreeding coefficients and coalescence time. Genet. Res. **58:** 167–175.

TAJIMA, F., 1989 Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

TAKAHATA, N., and M. NEI, 1985 Gene genealogy and variance of interpopulational nucleotide differences. Genetics **100:** 325–344.

TEMPLETON, A. R. 1994 The role of molecular genetics in speciation studies, pp. 455–477 in *Molecular Approaches to Ecology and Evolution,* edited by B. SCHIERWATER, B. STREIT, G. WAGNER and R. DESALLE. Birkhäuser-Verlag, Basel.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila,* pp. 421–469 in *Handbook of Genetics,* Vol. 3, edited by R. C. KING. Plenum Press, New York.

VEUILLE, M., and L. M. KING, 1995 Molecular basis of polymorphism at the esterase-5B locus in *Drosophila pseudoobscura.* Genetics **141:** 255–262.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

WHEELER, D. A., C. P. KYRIACOU, M. GREENACRE, Q. YU, J. RUTILA *et al.,* 1991 Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster.* Science **251:** 1082–1085.

WRIGHT, S., 1940 Breeding structure of populations in relation to speciation. Am. Nat. **74:** 232–248.

WRIGHT, S., 1951 The genetical structure of populations. Ann. Eugenet. **15:** 323–354.

WU, C.-I., and M. F. PALOPOLI, 1994 Genetics of postmating reproductive isolation in animals. Annu. Rev. Genet. **28:** 283–308.

YU, Q., H. V. COLOT, C. P. KYRIACOU, J. C. HALL and M. ROSBASH, 1987 Behavior modification by in vitro mutagenesis of a variable region within the *period* gene of *Drosophila.* Nature **326:** 765–769.

Communicating editor: W.-H. LI