Inferring the History of Speciation from Multilocus DNA Sequence Data: The Case of *Drosophila pseudoobscura* and Close Relatives

Carlos A. Machado, Richard M. Kliman,[1] *Jeffrey A. Markert,*[2] *and Jody Hey*

Department of Genetics, Rutgers University

The divergence of *Drosophila pseudoobscura* from its close relatives, *D. persimilis* and *D. pseudoobscura bogotana*, was examined using the pattern of DNA sequence variation in a common set of 50 inbred lines at 11 loci from diverse locations in the genome. *Drosophila pseudoobscura* and *D. persimilis* show a marked excess of low-frequency variation across loci, consistent with a model of recent population expansion in both species. The different loci vary considerably, both in polymorphism levels and in the levels of polymorphisms that are shared by different species pairs. A major question we address is whether these patterns of shared variation are best explained by gene flow or by persistence since common ancestry. A new test of gene flow, based on patterns of linkage disequilibrium, is developed. The results from these, and other tests, support a model in which *D. pseudoobscura* and *D. persimilis* have exchanged genes at some loci. However, the pattern of variation suggests that most gene flow, although occurring after speciation began, was not recent. There is less evidence of gene flow between *D. pseudoobscura* and *D. p. bogotana*. The results are compared with recent work on the genomic locations of genes that contribute to reproductive isolation between *D. pseudoobscura* and *D. persimilis*. We show that there is a good correspondence between the genomic regions associated with reproductive isolation and the regions that show little or no evidence of gene flow.

Introduction

Studies of gene flow via interspecific hybridization can be invaluable for understanding the role that natural selection plays during the formation of new species and for identifying genomic regions involved in reproductive isolation. When reproductive isolation is not complete (i.e., when F1 hybrids are not completely sterile), genes can pass between species. Therefore, incipient or hybridizing species may exchange genes and share genetic variation. This process of gene flow between species (also called introgressive hybridization) was first discussed by Anderson and Hubricht (1938) and later by Anderson (1949) with respect to its importance as a mechanism for generating new adaptations in plants.

Gene flow between incipient species is a component of the divergence-with-gene-flow models of speciation (i.e., sympatric or parapatric models) (Maynard Smith 1966; Endler 1977; Felsenstein 1981; Rice and Hostert 1993). Interestingly, these models have the consequence that incipient or hybridizing species can become divergent over some part of the genome although they may continue to share variation at others (Wang, Wakeley, and Hey 1997). This is so because some regions of the genome may introgress more readily than others (Clarke, Johnson, and Murray 1996; della Torre et al. 1997; Wang, Wakeley, and Hey 1997; Rieseberg, Whitton, and Gardner 1999; Jiang et al. 2000; Noor et al. 2001). Natural selection is expected to preclude gene flow at regions of the genome that are associated with (or linked to genes for) species-specific adaptations. Thus, natural selection can maintain species that are distinct from each other at some genes, in spite of persistent gene flow at other genes.

Under divergence-with-gene-flow models, natural selection has a direct role in generating and strengthening barriers to gene flow, and therefore a direct role in generating species. The role of natural selection in these models differs sharply from that in the classic and most accepted genetic model of speciation, the Dobzhansky-Muller model (Dobzhansky 1937; Muller 1940) (which was originally described by Bateson [Orr 1996]), in which natural selection plays an indirect role in speciation. In that model, reproductive isolation is simply the result of incompatibilities between gene variants that have arisen independently in each species and that are deleterious in a different genetic background.

Recently, speciation studies have taken advantage of several modern population genetic and phylogenetic techniques to analyze multilocus DNA sequence data (Bernardi, Sordino, and Powers 1993; Hey and Kliman 1993; Burton and Lee 1994; Hey 1994; Hilton and Hey 1997; Wang, Wakeley, and Hey 1997; Hare and Avise 1998; Kliman et al. 2000). The overall approach involves detailed population genetic analysis of species divergence for each of the several loci as well as an analysis of patterns that appear to be common among loci. This general methodology has been called divergence population genetics (DPG) (Kliman et al. 2000). By including multiple loci, the approach permits inferences regarding historical gene flow and natural selection that have acted on some, but not all genes. It is, therefore, possible to investigate whether different regions of the genome of incipient species have undergone more gene flow than others. This makes the DPG approach a powerful one to assess the importance of gene flow and natural selection during species divergence.

[1] Present address: Department of Biology, Kean University.

[2] Present address: Department of Biological Sciences, University of Cincinnati.

Abbreviations: DPG, divergence population genetics.

Key words: *Drosophila pseudoobscura*, speciation, polymorphism, natural selection, gene flow, reproductive isolation.

Address for correspondence and reprints: Jody Hey, Department of Genetics, Rutgers University, Nelson Biological Labs, 604 Allison Road, Piscataway, New Jersey 08854-8082.
E-mail: jhey@mbcl.rutgers.edu.

Mol. Biol. Evol. 19(4):472–488. 2002

Drosophila pseudoobscura and *D. persimilis* are a classic species pair for the study of speciation (Dobzhansky 1936; Dobzhansky and Epling 1944; Powell 1983; Orr 1987; Wang, Wakeley, and Hey 1997; Noor et al. 2001). It is estimated that the species started to diverge about 500,000 years ago (Aquadro et al. 1991; Wang, Wakeley, and Hey 1997), and reproductive isolation is not complete. F1 hybrid females are fertile, but F1 hybrid males are sterile; backcross hybrid males are fertile, but some of the hybrid backcross females are sterile (Dobzhansky 1936; Orr 1987, 1989); and there is geographic variation in *D. pseudoobscura* for the degree of premating isolation with *D. persimilis* (Noor 1995*b*). Hybridization does occur in nature, as a small number of backcross hybrid individuals have been collected in the field (Dobzhansky 1973; Powell 1983). Therefore, there is the potential for gene introgression across species via backcross of hybrid females to the parental species. Although there are fixed inversion differences on chromosome XL and chromosome 2, which should impede gene introgression at loci located in these chromosome regions (Tan 1935; Dobzhansky and Epling 1944; Anderson, Ayala, and Michod 1977; Moore and Taylor 1986), a study of hybrids using 14 codominant molecular markers (microsatellites and RFLPs) found no evidence of major barriers decreasing the potential for gene flow across most of the autosomal chromosomes (Noor et al. 2001). A DPG study of three loci found evidence of gene flow for one locus (*Adh*) located in the fourth chromosome (Wang, Wakeley, and Hey 1997).

In 1963 *D. pseudoobscura* was found to have a closer relative, *D. pseudoobscura bogotana*, which occurs in allopatry in Colombia (Dobzhansky et al. 1963). These two subspecies are estimated to have begun diverging about 200,000 years ago (Wang, Wakeley, and Hey 1997). Although there is very little premating isolation between these subspecies (Noor 1995*a*), hybrid *D. pseudoobscura-D. p. bogotana* males are fertile when *D. pseudoobscura* is the mother but sterile when *D. p. bogotana* is the mother.

Here we address the question of how much gene flow among these taxa has occurred historically and how has it varied for different regions of the genome by collecting and analyzing sequence data from 11 loci. We consider these data, together with previously collected data, using a broad population genetic approach which includes a new method for assessing gene flow.

Materials and Methods

Drosophila Stocks

Fifty-one inbred lines were established, one each from a set of isofemale lines that were collected from locations in the western United States by Noor (Noor et al. 1998; Noor, Schug, and Aquadro 2000) and from Colombia by Álvarez and Ruíz-García (Universidad Javeriana, Bogotá, Colombia). The locations include: Flagstaff, Arizona; Abajo Mountains, American Fort Canyon (AFC, AF), Utah; Mather, Mount St. Helena (MSH), California; Salem, Oregon; Sutatausa, Susa, Tabio (Toro), and La Calera (Potosí), in the vicinities of the Sabana de Bogotá (Cundinamarca, Colombia). Figure 1 shows the approximate locations and species for the North American collections. We used 20 lines of *D. pseudoobscura* (Abajo36, AF2, AFC3, AFC7, AFC12, Flagstaff5, Flagstaff6, Flagstaff14, Flagstaff16, Flagstaff18, Mather10, Mather17, Mather32, Mather48, Mather52, MSH9, MSH10, MSH21, MSH24, and MSH32), 14 lines of *D. persimilis* (Mather6, Mather27, Mather37, Mather39, Mather40, Mather41, MatherB, MatherG, MSH1, MSH3, MSH7, MSH25, MSH42, and Salem), 14 lines of *D. p. bogotana* (Susa1, Susa2, Susa3, Susa6, Sutatausa1, Sutatausa2, Sutatausa3, Sutatausa5, Toro1, Toro4, Toro6, Toro7, Potosí2, and Potosí3), and 3 lines of *D. miranda* (MSH22, MSH38, and Mather28). Half of the *D. pseudoobscura* lines are from locations where this species is sympatric with *D. persimilis* (MSH and Mather, California).

DNA Extractions

The original isofemale lines went through 12–17 generations of full sib–mating prior to DNA sequencing. Genomic DNA from each inbred line was extracted using protocols 47 and 48 of Ashburner (1989, pp. 106–109).

Loci

DNA sequences were collected for 11 loci (table 1). Nine of these are noncoding regions that flank or include microsatellite markers (or both) developed for *D. pseudoobscura* (Noor, Schug, and Aquadro 2000), and two are protein coding genes (*bcd* and *rh1*). The sequences of three loci (*X010, 4002* and *bcd*) contain the microsatellite, but the repeats were not included in the analyses. Previously reported sequences from *Adh/Adh-dup* (Schaeffer and Miller 1992*b;* Wang, Wakeley, and Hey 1997), *per* (*period*), and *Hsp82* (Schaeffer and Miller 1992*b;* Wang, Wakeley, and Hey 1997) were also included in the analyses. The *D. pseudoobscura Adh/Adh-dup* data set consists of a subset of 10 sequences from the Apple Hill population (Schaeffer and Miller 1992*a*, 1992*b*). These 14 loci are scattered across the genome of *D. pseudoobscura* (table 1, fig. 2). Chromosomal locations and recombinational distances among the microsatellite markers have been previously reported (Noor, Schug, and Aquadro 2000; Noor and Smith 2000). The cytological location of the markers was determined by in situ hybridization using the method of Lim (1993).

DNA Sequencing of Microsatellite-Flanking Regions

Sequence information for the noncoding regions containing the microsatellites was obtained using an inverse PCR protocol (Offringa and van der Lee 1995). Briefly, genomic DNA from single individuals of the four species was digested with a four-cutter restriction enzyme (either *BfaI*, *HhaI*, *MspI*, *NlaIII*, or *TaqI*) that did not cut the sequence of the clones that included the microsatellites. DNA ligase was added to the restricted DNA and incubated overnight to generate circularized DNA. The sequences of the clones, including the mi-

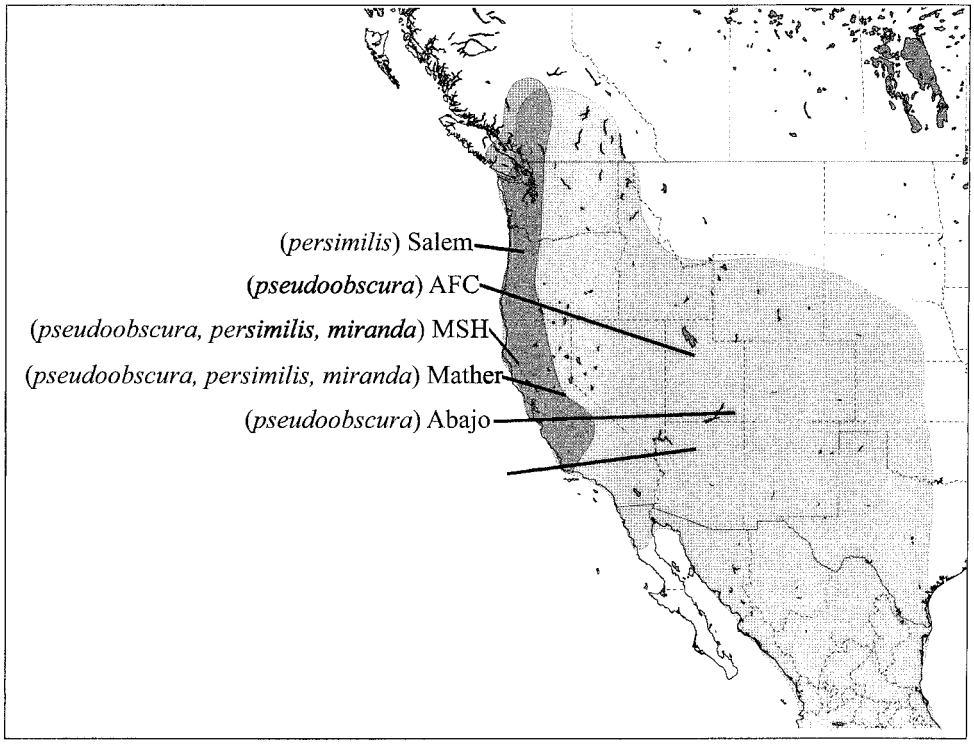

FIG. 1.—Map of the western part of the United States showing the location of the sites where *D. pseudoobscura*, *D. persimilis*, and *D. miranda* were collected. The light gray region corresponds approximately to the geographical range of *D. pseudoobscura* and the dark gray region corresponds to the region where both *D. persimilis* and *D. pseudoobscura* occur.

crosatellites, were used to design pairs of primers going toward the outside of the fragment (inverse PCR primers). PCR was performed using each circularized DNA as template (one solution per restriction enzyme) and the inverse PCR primers (i.e., inverse PCR). PCR products ranging from 0.5 to 2 kbp were generally observed. The PCR products included both 5′ and 3′ ends of the fragments that included the microsatellites plus the flanking region in each direction up to the sites recognized by the restriction enzyme used. These PCR fragments were then sequenced, and the sequence of the flanking regions was used to design new PCR primers that amplified fragments 0.8–1.5 kbp in nondigested genomic DNA from each of the four species.

For DNA sequencing, a PCR reaction was performed using one of the primers with an M13 forward

Table 1
Genomic Location of the Sequenced Loci in *D. pseudoobscura* and *D. melanogaster*

Locus	Chromosome	Cytological Location[a]	Chromosomal Location in *D. melanogaster*	*D. melanogaster* GenBank Accession Number[b]
X008	XL	15	X	AC004114
X009	XR	21	3L	AE003479
X010	XR	39	3R	AF315732
2001	2	58	3R	AE003758
2002	2	54	3R	AE003764
bcd	2	50	3R	AE003674
rh1[c]	2	45	3R	AE003728
2003	2	43	3R	AE003691
3002	3	74–76	2R	AE003466
4002	4	98–99	[d]	[d]
4003	4	86	2L	AE003613
Adh	4	88	2L	AE003644
per	XL	2	X	AE003425
Hsp82	XR	23	3L	AE003477

[a] Location determined by in situ hybridization with regard to the standard cytological maps (Dobzhansky and Tan 1936; Stocker and Kastritsis 1972). Cytological locations reported in the literature: *rh1* (Carulli and Hartl 1992); *Adh* (Schaeffer and Aquadro 1987); *bcd* (Segarra, Ribó and Aguadé 1996); and *Hsp82* (Segarra, Ribó and Aguadé 1996).

[b] Accession number for the sequence of the clone or genomic scaffold containing the homologous *D. melanogaster* sequence.

[c] The homologue of *rh1* in *D. melanogaster* is the *ninaE* gene (O'Tousa *et al.* 1985).

[d] No clear homologous sequence was found for *4002* (see text).

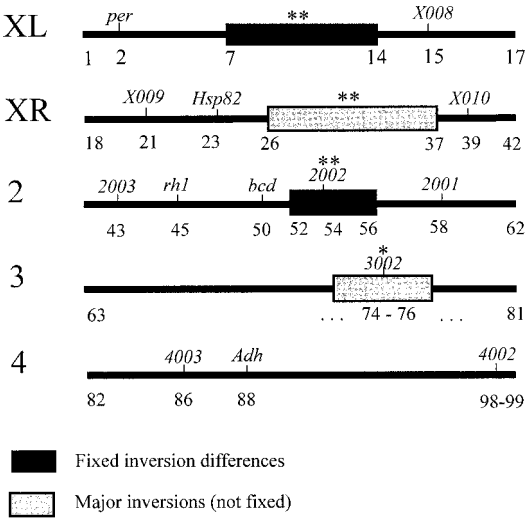

FIG. 2.—Cytological locations of the sequenced loci. Chromosome arm sizes are not drawn at physical or recombinational scales. Cytological bands are shown below each marker or each inversion. The dot chromosome is not shown because none of the markers are located in that linkage group. Information about the break points of the fixed and major inversions is from Moore and Taylor (1986). The inversion in the XR arm is fixed among *D. pseudoobscura* and non–Sex-Ratio (SR) XR *D. persimilis* strains. The two species also differ in the inversion polymorphisms for the third chromosome, but they share the standard arrangement. The break points of the third chromosome inversion are marked as (. . .) because they differ for each arrangement. Asterisks indicate which regions of the genome are strongly (**) or weakly (*) associated with isolation mechanisms between *D. pseudoobscura* and *D. persimilis* (Noor et al. 2001).

tail and the other with an M13 reverse tail. For fragments longer than 1 kbp, internal M13-tailed primers were designed to carry out secondary PCR amplifications of two smaller overlapping fragments. The PCR fragments were either gel purified (QIAGEN), column purified (MILLIPORE), or diluted 1/10 and sequenced bidirectionally using fluorescently labeled M13 primers in a LI-COR 4200 automated sequencer (Lincoln, Neb.). The sequence files were edited and assembled using the program ALIGN-IR (LI-COR, Lincoln, Neb.). PCR and sequencing primer information is available upon request.

bcd and *rh1* Sequencing

Primers were designed to amplify a 1.4-kbp PCR fragment, including introns 1–3 and exons 2 and 3 of the *bcd* (*bicoid*) gene. The sequence data used for the analyses encompasses positions 776–2147 of the complete *D. pseudoobscura* sequence (Seeger and Kaufman 1990). Primers were designed to amplify a 1.5-kbp PCR fragment, including introns 2–4 and exons 2–5 from the *rh1* (*Rhodopsin 1*) gene. The sequenced region corresponds to positions 611–2055 of the published *D. pseudoobscura* sequence (Carulli and Hartl 1992). Two smaller overlapping fragments were amplified using M13-tailed primers and sequenced as described previously.

Data Analyses

The sequences from each homologous data set were initially aligned with the program PileUp (Wisconsin Package v. 10, Genetics Computer Group, Madison, Wis.). Manual alignments were further performed in some data sets to improve the PileUp alignments. BLAST searches (Altschul et al. 1990) against the genome sequence of *D. melanogaster* were performed for each microsatellite-flanking region of *D. pseudoobscura* using the tool available at the Berkeley Drosophila genome project web site (http://www.fruitfly.org). Basic polymorphism analyses were performed with the program SITES (Hey and Wakeley 1997). Indels were not included in the analyses. The data from *D. miranda* were primarily used to root the variation found among the other species. Analyses of molecular variance (AMOVA) (Excoffier, Smouse, and Quattro 1992) were carried out with the Arlequin computer program (Schneider, Roessli, and Excoffier 2000). McDonald-Kreitman tests (McDonald and Kreitman 1991) were performed using the data from all four species, counting a given site as polymorphic if it was variable in any one of the species and performing the *G*-tests of independence using Williams' correction (Sokal and Rohlf 1981, p. 704). The polymorphism data were fitted to a model of speciation with no gene flow (Wakeley and Hey 1997) using the method described by Wang, Wakeley, and Hey (1997). A new method to assess gene flow using patterns of linkage disequilibrium (LD) is described subsequently (see *Results*). We discuss the results based on the traditional phylogeny of the species ([pseudoobscura, bogotana], persimilis). We focus primarily on the *pseudoobscura-bogotana* and *pseudoobscura-persimilis* comparisons because *D. pseudoobscura* and *D. p. bogotana* are the most closely related species and *D. pseudoobscura* and *D. persimilis* are partially sympatric.

Results

Physical Locations of the Loci

Chromosome and cytological locations are presented in table 1 and figure 2. No markers from chromosome five (dot chromosome) were sequenced. The locus *2002* is located in a region of a fixed inversion difference between *D. pseudoobscura* and *D. persimilis*. The locus *3002* is found in a region that is spanned by several polymorphic inversions found within and among these species (Powell 1992). The conservation of chromosome elements in the genus *Drosophila* (Muller 1940; Sturtevant and Novitski 1941) allows predictions of the chromosome location of *D. pseudoobscura* markers by determining the location of the homologous marker in the genome of *D. melanogaster*. The results of BLAST (Altschul et al. 1990) searches show that, with the exception of two loci (*X010* and *4002*), the location of the loci obtained with classic physical mapping techniques (Noor, Schug, and Aquadro 2000) and in situ hybridization (this study) is consistent with the location of homologous sequences (or sequences that generate the most significant alignments) in the *D. melanogaster* genome (table 1). For *X010*, BLAST search-

es to GenBank found similarities with only a 207-bp portion at the 5′ end, which corresponds to an uncharacterized region in *D. melanogaster* named *Jon99C* and to a repetitive sequence reported from *D. miranda* (Steinemann and Steinemann 1992). The sequence is probably a part of an old retrotransposon (Steinemann and Steinemann 1992). This interpretation is also consistent with the fact that *X010* was the only locus for which we could not amplify DNA from *D. miranda*. BLAST searches with *4002* produced only very short alignments with diverse sequences from several genomes.

Description of Intraspecific Variation

Polymorphism analyses are summarized in table 2. Consistent with previous observations based on data from three genes (Wang, Wakeley, and Hey 1997), *D. pseudoobscura* and *D. p. bogotana* have the most and the least nucleotide variation, respectively. The only loci showing exceptions to that pattern are *X009* and *Adh*, where *D. persimilis* has more variation than *D. pseudoobscura*, and *4003*, where both taxa have similar levels of variation. These observations suggest a larger historic effective population size for *D. pseudoobscura*, which is consistent with its more extensive geographic distribution and agree with previous findings showing that this species is highly polymorphic (Riley, Kaplan, and Veuille 1992; Schaeffer and Miller 1992*b;* Veuille and King 1995; Wells 1996; Hamblin and Aquadro 1999). The most noteworthy observation in the protein-coding genes is the complete lack of replacement polymorphism and fixed replacement differences at *rh1*. This is not unexpected, however, given that *rh1* is a very conserved gene in *Drosophila* (Carulli and Hartl 1992).

The weighted average values of Watterson's estimator, $\hat{\theta}$ (Watterson 1975), of the population mutation rate parameter $4Nu$ (or θ, where N is the effective population size and u is the neutral mutation rate), per base pair for autosomal loci of *D. pseudoobscura*, *D. persimilis*, and *D. p. bogotana* are 0.0148, 0.0097, and 0.0059, respectively, whereas the values for X-linked loci are 0.0149, 0.0090, and 0.0014. Interestingly, with the exception of *D. p. bogotana* the expected reduction in the average value of $\hat{\theta}$ for X-linked loci is not observed in these samples. The locus *X008* in *D. pseudoobscura* and *X009* in *D. persimilis*, in particular, have high levels of polymorphism. When these loci are not included in the calculation of the weighted average values, X-linked variation becomes about 67% of the average autosomal variation in *D. pseudoobscura* (0.0099 without *X008*) and 63% in *D. persimilis* (0.0068 without *X009*).

AMOVA analyses (Excoffier, Smouse, and Quattro 1992) show that, with respect to sequence variation at these loci, these taxa are largely panmictic throughout their geographical range (table 3). The distribution of variation is similar across most loci, with almost all variation caused by within-population and between-species variation. In *X009* there is a significant covariance component attributable to differences among populations ($F_{SC} = 0.3546$ and $P = 0.007$) that explains about 15% of the total variation. No evidence of population structure is observed for *X009* when only *D. p. bogotana* and *D. persimilis* are compared ($F_{SC} = -0.1482$ and $P = 0.5$), but the covariance component is significant in the *D. pseudoobscura-D. persimilis* and *D. pseudoobscura-D. p. bogotana* comparisons ($F_{SC} = 0.3946$ and $P = 0.01$; $F_{SC} = 0.4578$ and $P = 0.009$). The evidence of population structure in *D. pseudoobscura* based on *X009* is caused by an interesting pattern of haplotype structure in this locus, where the first 246 bp of the aligned sequence of three out of four haplotypes from one sympatric (Mather) and one allopatric (AFC) locality are quite different from the rest of *D. pseudoobscura* and *D. p. bogotana* haplotypes but almost identical to those of *D. persimilis*. If the analyses are repeated without that region, the evidence of population structure disappears ($F_{SC} = -0.0175$ and $P = 0.16$; $F_{ST} = 0.4603$ and $P < 0.001$; $F_{CT} = 0.4695$ and $P = 0.009$). These results are generally consistent with earlier allozyme (Prakash, Lewontin, and Hubby 1969; Singh 1983; Keith et al. 1985), RFLP (Riley, Hallas, and Lewontin 1989), sequence (Schaeffer and Miller 1992*a;* Wang and Hey 1996), and microsatellite (Noor, Schug, and Aquadro 2000) data supporting the lack of geographic population structure in *D. pseudoobscura*.

Testing the Neutral Hypothesis

HKA tests (Hudson, Kreitman, and Aguade 1987) were performed to determine whether the amounts of polymorphism and divergence across loci are correlated, as expected under neutrality. Because these species are closely related, the HKA test statistic is not expected to follow the χ^2 distribution. Therefore, the test statistic was compared with a distribution generated from 10,000 coalescent simulations (see e.g., Hilton, Kliman, and Hey 1994). HKA tests were applied to each of the three taxa, in each case using a single sequence from *D. miranda* as an outgroup (*D. pseudoobscura*, $\chi^2 = 4.47$, $P = 0.821$; *D. persimilis*, $\chi^2 = 11.18$, $P = 0.309$; *D. p. bogotana*, $\chi^2 = 19.4$, $P = 0.048$) or using all sequences from two taxa (*D. pseudoobscura-D. persimilis*, $\chi^2 = 7.70$, $P = 0.944$; *D. pseudoobscura-D. p. bogotana*, $\chi^2 = 18.54$, $P = 0.331$; *D. persimilis-D. p. bogotana*, $\chi^2 = 22.55$, $P = 0.408$). Only the analysis of polymorphism within *D. p. bogotana*, using *D. miranda* as an outgroup, was statistically significant. Among the contributions to the χ^2 statistic in this case, the largest came from lower than expected polymorphism within *D. p. bogotana* at the *period* locus, a finding that was previously noted (Wang and Hey 1996).

The McDonald-Kreitman test (McDonald and Kreitman 1991) uses a contrast similar to that of the HKA test but examines different types of sites that are interspersed with each other over the sequence of a locus. This test examines whether the ratio of silent to replacement variation is the same for polymorphisms as it is for fixed differences between species. Under the assumption that these two kinds of variation are selectively neutral, the ratios are expected to be the same. The McDonald-Kreitman test revealed no departure

Table 2
Polymorphism Statistics

Locus	Species	n[a]	L[b]	S[c]	syn[d]	rep[d]	θ[e]	π[f]	D[g]	4Nc[h]	Div.[i]
X008	*pseudoobscura*	17	998.0	109	—	—	0.0323	0.0210	−1.4865	0.0629	0.0465
	bogotana	12	1,040.5	15	—	—	0.0047	0.0037	−0.9669	0.0028	0.0426
	persimilis	13	996.3	35	—	—	0.0113	0.0077	−1.3849	0.0459	0.0478
	miranda	3	1,067.3	22	—	—	0.0137	0.0137	—	—	
X009	*pseudoobscura*	18	693.9	40	—	—	0.0167	0.0139	−0.6977	0.0168	0.0343
	bogotana	14	701.0	1	—	—	0.0004	0.0002	−1.1552	—	0.0336
	persimilis	14	701.0	42	—	—	0.0188	0.0175	−0.3112	0.0657	0.0325
	miranda	1	704.0	—	—	—	—	—	—	—	
X010[j] ...	*pseudoobscura*	20	869.2	17	—	—	0.0055	0.0026	−1.9945*	0.0000	—
	bogotana	14	888.9	0	—	—	0.0000	0.0000	—	—	—
	persimilis	14	872.9	9	—	—	0.0032	0.0015	−2.0942**	—	—
	miranda	—	—	—	—	—	—	—	—	—	
2001	*pseudoobscura*	17	678.1	38	—	—	0.0166	0.0108	−1.4482	0.0450	0.0210
	bogotana	13	694.4	5	—	—	0.0023	0.0018	−0.8419	—	0.0222
	persimilis	14	677.2	20	—	—	0.0093	0.0073	−0.9079	0.0898	0.0175
	miranda	1	690.0	—	—	—	—	—	—	—-	
2002	*pseudoobscura*	19	924.8	66	—	—	0.0204	0.0152	−1.0396	0.0694	0.0193
	bogotana	13	922.8	17	—	—	0.0059	0.0071	0.8372	0.0000	0.0162
	persimilis	13	905.2	16	—	—	0.0057	0.0038	−1.3807	0.0206	0.0171
	miranda	3	918.3	21	—	—	0.0152	0.0152	—	—	
bcd.....	*pseudoobscura*	20	1,367.9	47	25	11	0.0097	0.0074	−0.9633	0.0344	0.0192
	bogotana	14	1,370.8	14	8	4	0.0032	0.0030	−0.2488	0.0000	0.0191
	persimilis	14	1,375.4	30	18	7	0.0068	0.0058	−0.6620	0.0306	0.0176
	miranda	2	1,369.0	0	0	0	0.0000	0.0000	—	—	
rh1.....	*pseudoobscura*	17	1,443.0	52	14	0	0.0106	0.0071	−1.3903	0.0251	0.0151
	bogotana	11	1,200.0	3	3	0	0.0008	0.0007	−0.7494	—	0.0148
	persimilis	14	1,443.1	41	12	0	0.0089	0.0069	−1.0102	0.0204	0.0155
	miranda	2	1,443.0	1	1	0	0.0007	0.0007	—	—	
2003	*pseudoobscura*	18	512.8	18	—	—	0.0102	0.0068	−1.2721	0.0114	0.0195
	bogotana	14	528.3	7	—	—	0.0042	0.0027	−1.2767	—	0.0143
	persimilis	14	532.3	11	—	—	0.0065	0.0038	−1.6374	—	0.0176
	miranda	3	507.0	1	—	—	0.0013	0.0013	—	—	
3002	*pseudoobscura*	11	607.3	57	—	—	0.0320	0.0305	−0.2318	0.0463	0.0812
	bogotana	13	615.7	24	—	—	0.0126	0.0125	−0.0305	0.0704	0.0817
	persimilis	13	615.0	26	—	—	0.0136	0.0106	−0.9557	0.0471	0.0735
	miranda	1	617.0	—	—	—	—	—	—	—	
4002	*pseudoobscura*	18	829.2	14	—	—	0.0049	0.0026	−1.7266	0.000	0.0034
	bogotana	14	819.0	5	—	—	0.0019	0.0019	−0.0502	0.000	0.0053
	persimilis	13	821.6	8	—	—	0.0031	0.0015	−2.0240*	—	0.0007
	miranda	1	831.0	—	—	—	—	—	—	—	
4003	*pseudoobscura*	15	623.2	50	—	—	0.0247	0.0194	−0.9216	0.1633	0.0434
	bogotana	14	627.1	29	—	—	0.0145	0.0113	−0.9591	0.0285	0.0490
	persimilis	13	615.7	45	—	—	0.0235	0.0196	−0.7382	0.1241	0.0439
	miranda	3	624.7	31	—	—	0.0330	0.0325	—	—	
Adh	*pseudoobscura*	10	3,449.4	110	31	4	0.0113	0.0100	−0.5578	0.0695	0.0322
	bogotana	8	3,447.0	61	16	3	0.0068	0.0066	−0.1454	0.0149	0.0318
	persimilis	6	3,447.3	94	23	9	0.0119	0.0118	−0.0786	0.0798	0.0312
	miranda	1	3,461.0	—	—	—	—	—	—	—	
per.....	*pseudoobscura*	11	1,459.2	48	22	6	0.0112	0.0084	−1.2002	0.0271	0.0292
	bogotana	9	1,479.2	3	1	1	0.0007	0.0009	0.6021	0.0000	0.0331
	persimilis	11	1,481.2	36	12	9	0.0083	0.0069	−0.7600	0.0226	0.0248
	miranda	4	1,480.7	9	5	3	0.0033	0.0031	−0.4915	—	
Hsp82 ...	*pseudoobscura*	11	1,976.1	34	6	1	0.0059	0.0042	−1.3608	0.0025	0.0255
	bogotana	9	1,917.1	6	0	1	0.0011	0.0012	0.1386	0.0000	0.0258
	persimilis	11	1,937.6	10	2	0	0.0017	0.0012	−1.2645	0.0000	0.0242
	miranda	4	1,980.7	4	1	0	0.0011	0.0012	0.6501	—	

NOTE. (—) values could not be obtained for small samples or for groups of sequences with few informative sites. * significant at $P < 0.05$; ** significant at $P < 0.01$.

[a] Number of lines sequenced.

[b] Average length (bp) of the sequences from each species.

[c] Number of polymorphic sites.

[d] Number of synonymous (syn) and replacement (rep) polymorphisms in the coding regions.

[e] Estimate of 4Nu (3Nu for X-linked loci) per base pair using the number of polymorphic sites (Watterson 1975).

[f] Estimate of 4Nu (3Nu for X-linked loci) using the average number of nucleotide differences per site (Nei 1987).

[g] Tajima's statistic (1989*b*). Significance was determined using table 2 of Tajima (1989*b*).

[h] Estimate of the population recombination rate (4Nc) per base pair (Hey and Wakeley 1997).

[i] Average divergence per base pair between alleles from each taxon and the alleles of *D. miranda*.

[j] No amplification of *X010* could be obtained for *D. miranda*.

Table 3
ANOVA and Hierarchical Analyses

Source of Variation (%)	*X008*	*X009*	*X010*	*2001*	*2002*	*2003*	*3002*	*4002*	*4003*	*bcd*	*rh1*[a]
Among speices	44.32	55.36	69.38	32.64	59.67	42.39	49.03	59.23	29.76	31.65	33.73
Among populations within species[b]	−0.84	15.83	0.81	−2.59	−0.36	−4.33	−5.12	−0.90	−1.02	1.84	−3.68
Within populations	56.52	28.81	29.81	69.95	40.69	61.94	56.09	41.67	71.26	66.51	69.95
Fixation indices[c]											
F_{CT} (species/total)	0.443	0.554	0.694	0.326	0.597	0.424	0.490	0.592	0.298	0.316	0.337
F_{SC} (population/species)[b]	−0.015	0.355	0.026	−0.038	−0.009	−0.075	−0.100	−0.022	−0.014	0.027	−0.055
F_{ST} (population/total)	0.435	0.712	0.702	0.300	0.593	0.381	0.439	0.583	0.287	0.335	0.300

NOTE.—Variance partitioning and associated fixation indices were only calculated for the 11 loci for which geographical sampling was the same.

[a] A 242-bp deletion present in the second intron of the *D. p. bogotana* sequences was not included in the analyses (see text).

[b] Negative variance values and fixation indices occur because they are neither covariances nor correlation coefficients, respectively, and indicate that there is no genetic structure within species.

[c] All F_{CT} and F_{ST} values are highly significant ($P < 0.001$). The only significant F_{SC} value is that estimated for *X009* ($P = 0.007$) (see text).

from the neutral model at *bcd* ($G = 1.434$, $P = 0.231$), *Hsp82* ($G = 1.824$, $P = 0.177$), or *per* ($G = 0.535$, $P = 0.464$). The test for the *Adh* region, comprised of *Adh* ($G = 1.726$, $P = 0.189$) and *Adh-dup* ($G = 1.374$, $P = 0.241$), is significant ($G = 3.882$, $P = 0049$) but only before correcting for multiple tests. Although it is impossible to assign with confidence the cause of the departure from neutrality, the observed pattern suggests that there may be an excess of replacement differences between species at this locus.

We also examined whether the pattern of variation at each locus within each species was consistent with the neutral model. Table 2 shows the value of Tajima's *D* (Tajima 1989*b*), which is proportional to the difference between two estimates of the population mutation parameter θ, the mean pairwise differences between the sampled sequences (π), and Watterson's estimator $\hat{\theta}$. Under a neutral model with constant population size, both estimators have the same expected value. In our sample, Tajima's *D* was negative in almost all the cases, but its value was significantly different from zero only in the *X010* sample from *D. pseudoobscura* and *D. persimilis* and in the *4002* sample from *D. persimilis* (table 2). Negative values of *D* are expected in the presence of purifying selection, or following a selective sweep, or in samples from populations that are expanding in size (Tajima 1989*a*, 1989*b*). To test whether the average value of Tajima's *D*, within each species, departs significantly from zero, we conducted a test using the same simulations used in the HKA test. For *D. pseudoobscura* and *D. persimilis* the mean values of *D* (D) were less than all of the means found in 10,000 simulations ($D = -1.100$, $P < 0.0001$; $D = -1.009$, $P < 0.0001$, respectively), whereas the *D. p. bogotana* value was not significantly different from zero ($D = -0.372$, $P = 0.124$). The consistency of the negative value of *D* across loci suggests a demographic explanation because demographic forces affect all loci simultaneously. A recent population expansion in *D. pseudoobscura* and *D. persimilis* is the likely explanation for this general pattern.

A relative rate test for multiple sequences (Li and Bousquet 1992) was used to examine whether there is evidence of differences in the rate of substitution among taxa. After correcting for multiple comparisons, the tests show evidence of rate heterogeneity across taxa in the sequences from *4002* and *per*. The sequences of *D. p. bogotana* have evolved faster than the sequences of *D. persimilis* (*4002:* $Z = 6.222$ and $P < 0.0001$; *per:* $Z = 4.805959$ and $P < 0.0001$); and the sequences of *D. pseudoobscura* have evolved faster than those of *D. persimilis* (*4002* $Z = 4.883$ and $P < 0.0001$; *per:* $Z = 2.835$ and $P = 0.0046$). However, the fact that in some loci the outgroup *D. miranda* shares variation with the ingroup species (variation that probably predates the divergence of these taxa) reduces the utility of this test. For instance, the significant result for the *D. pseudoobscura-D. persimilis* comparison of *4002* can be explained by the fact that the sequence of *D. miranda* (Mather28) is identical to the sequences of eight *D. persimilis* lines.

Table 4
The Number of Shared Polymorphisms and Fixed Differences Between Species

	pseudoobscura-bogotana		*pseudoobscura-persimilis*	
Locus	Shared	Fixed	Shared	Fixed
X008	*5 (1.60)*	*2*	*2 (3.82)*	*10*
X009	*0 (0.06)*	*1*	*7 (2.40)*	*1*
X010	*0 (0)*	*0*	*0 (0.17)*	*6*
2001	*1 (0.27)*	*3*	*6 (1.12)*	*0*
2002	*8 (1.21)*	*0*	*0 (1.15)*	*6*
bcd	*4 (0.48)*	*0*	*5 (1.02)*	*0*
rh1	*1 (0.11)*	*2*	*8 (1.47)*	*0*
2003	1 (0.24)	0	2 (0.37)	0
3002	13 (2.23)	0	7 (2.42)	0
4002	0 (0.08)	0	0 (0.13)	1
4003	13 (2.31)	0	20 (3.63)	0
Adh	37 (1.94)	0	47 (2.99)	0
per	1 (0.09)	6	6 (1.17)	2
Hsp82	0 (0.10)	0	1 (0.17)	8

NOTE.—The expected number of shared polymorphisms on the basis of recurrent mutation (Clark 1997) are shown in parentheses.

In conclusion, neutral model assumptions, including selective neutrality and constant rate of mutation accumulation, are not generally violated by the data. However, the consistent negative values of Tajima's *D* suggest that the assumption of constant population size might not be correct for these data.

Shared Variation and Sequence Divergence

Under the null speciation model (see subsequently), two very recently diverged species are expected to share some polymorphisms that were present in the ancestral population. As the species diverge from each other, genetic drift within each species leads to an accumulation of fixed differences and a loss of shared polymorphisms. With observations from a number of loci, one expects to find a negative correlation between fixed differences and shared polymorphisms across loci. In particular, for a locus that has no history of recombination and no recurrent mutation, shared polymorphisms and fixed differences are mutually exclusive (Wakeley and Hey 1997). The loss of shared polymorphisms and the accumulation of fixed differences is expected to occur more rapidly at loci involved in adaptive divergence or at loci linked to such regions. On the other hand, if the strict isolation model is not correct and gene flow has occurred, then the divergence of a given locus will be retarded. That happens because gene flow removes, and prevents the accumulation of, fixed differences at the same time as it introduces shared polymorphisms.

Table 4 shows the number of shared and fixed differences between species. The expected negative relationship between the two quantities is observed, and in several genes the species pairs share large numbers of polymorphisms. Markers from chromosomes 2 and 4 show the largest counts of shared polymorphisms. Of the three markers showing no shared polymorphisms between *D. pseudoobscura* and *D. persimilis*, one is located in a region spanned by a fixed inversion difference (*2002*), and the other two (*X010* and *4002*) have few low-frequency polymorphisms. Shared polymorphism can also be caused by recurrent mutation (homoplasy). However, homoplasy can only explain a fairly small fraction of the observed shared polymorphisms in most of the genes (table 4).

Regarding patterns of sequence divergence across loci, under the strict isolation model it is expected that net divergence for each locus (Nei 1987, p. 276) should be proportional to the time since speciation. If gene flow has occurred at some loci, but not at others, we expect a large variance in levels of net divergence across loci, and we expect to find that loci with low values of net divergence should have more shared polymorphisms and higher population migration estimates. In addition, if these three taxa have diverged without gene flow, we expect to find that net divergence values should be similarly ranked when comparing the two species pairs *D. pseudoobscura-D. persimilis* and *D. pseudoobscura-D. p. bogotana*. Estimates of net divergence and population migration rates between the three taxa are shown in table 5. Interestingly, 6 out of 14 *D. pseudoobscura-D. persimilis* net divergence values (at *2001, bcd, rh1, 3002, 4003,* and *Adh*) are lower than the *D. pseudoobscura-D. p. bogotana* values. This pattern is also reflected in the population migration rate estimates, which are larger in the same six loci of the *D. pseudoobscura-D. persimilis* comparison and vary within one to two orders of magnitude across loci for the two species comparisons. These same six loci, in addition to *2003,* are the only ones showing no fixed differences between *D. pseudoobscura* and *D. persimilis*, and they have some of the largest counts of shared polymorphisms between these two taxa (table 4). These qualitative analyses of the data, therefore, reveal variation across loci in levels of net divergence and gene flow which suggests that the isolation model is not an accurate one for these speciation events.

Testing the Null Model of Speciation

The simplest neutral model, in the context of the population genetics of species divergence, is an isolation model (Hey 1994; Wakeley and Hey 1998). In its most basic form, the isolation model includes an ancestral population that has split into two populations at a point in time, after which genetic divergence and polymorphism have accumulated independently in the two new populations (Wakeley and Hey 1997). The model also employs the standard assumptions of selective neutrality of mutations and constant population sizes. Thus, it is straightforward to do coalescent simulations that correspond to the assumptions of the model and that use parameters estimated from actual data. In this way, it is possible to assess, in various ways, how well the data fit the assumptions of the model.

The assumptions of the isolation model are violated if genes have been exchanged between species. Gene flow will elevate the numbers of shared polymorphisms and reduce both the number of exclusive polymorphisms and the number of fixed differences between taxa. Furthermore, if gene flow occurs at some loci and not at

Table 5
Estimates of Population Migration Rates and Net Divergence

Locus	Net divergence per base pair[a]		Population migration rate (Nm)[b]	
	pseudoobscura-bogotana	*pseudoobscura-persimilis*	*pseudoobscura-bogotana*	*pseudoobscurat-persimilis*
X008	*0.00844*	*0.01850*	*0.362*	*0.198*
X009	*0.00289*	*0.01013*	*0.607*	*0.391*
X010	*0.00015*	*0.00785*	*2.125*	*0.065*
2001	*0.00374*	*0.00282*	*0.417*	*0.799*
2002	*0.00286*	*0.00822*	*0.978*	*0.295*
bcd	*0.00310*	*0.00250*	*0.418*	*0.657*
rh1	*0.00148*	*0.00140*	*0.780*	*1.247*
2003	0.00116	0.00158	1.037	0.842
3002	0.01347	0.00892	0.400	0.575
4002	0.00160	0.00211	0.357	0.248
4003	0.00605	0.00372	0.637	1.314
Adh	0.00197	0.00121	1.057	2.258
per	0.00879	0.00967	0.131	0.198
Hsp82	0.00176	0.00413	0.386	0.165

[a] Calculated using equation 10.21 of Nei (1987).
[b] Estimated using the method of Hudson *et al.* (1992).

others, it will elevate the variance among loci in numbers of shared polymorphisms and fixed differences. This last mentioned idea has been used as the basis for a test of gene flow by Wang, Wakeley, and Hey (1997) (hereafter referred as WWH). They used a simple measure (the difference between the highest and lowest counts of shared polymorphisms among a set of loci plus the difference between the highest and lowest counts of fixed differences observed over the same group of loci) and compared the observed value to a simulated distribution. Alternatively, one can use a χ^2 statistic to measure the overall fit of the data to the isolation model (Kliman et al. 2000). This χ^2 statistic compares the observed and expected counts of each type of polymorphic site (exclusive polymorphisms for each species, shared polymorphisms, and fixed differences between the species) over all the loci. The expected counts are obtained using the methods described by Wakeley and Hey (1997) and Wang, Wakeley, and Hey (1997).

The results of simulations assessing the significance of the observed values of the test statistics for the *D. pseudoobscura-D. persimilis* and *D. pseudoobscura-D. p. bogotana* comparisons are shown in table 6. The isolation model is not rejected for any of the two comparisons when the χ^2 statistic is used, as the observed values of the statistic do not depart exceptionally from those of the simulated distribution. However, use of the WWH test statistic leads to a clear rejection of the isolation model for the *D. pseudoobscura-D. persimilis* comparison ($P = 0.015$) but not for the *D. pseudoobscura-D. p. bogotana* comparison ($P = 0.06$). The WWH values and test results obtained here closely resemble those found with just three loci (Wang, Wakeley, and Hey 1997), and the simulation results are very similar.

The appropriateness of the isolation model can also be assessed by consideration of the parameter value estimates. Note, in particular, that the estimated population size of the ancestor ($\hat{\theta}_A$) of *D. pseudoobscura* and *D. persimilis* is larger than the population size of either descendant species (table 6), suggesting that the population size of these species might have contracted since their time of divergence. However, the consistently negative value of Tajima's *D* across loci (table 2) suggests a process of population expansion rather than contraction. These conflicting signals can be partly reconciled if we consider that the isolation model may not be accurate because of gene flow. If that has been the case, the elevated variance in shared and fixed differences, caused by gene flow, leads to an elevated estimate of the ancestral population size (Wang, Wakeley, and Hey 1997). On the other hand, in the *pseudoobscura-bogotana* comparison, where the isolation model was not re-

Table 6
Isolation Model Fitting

Species 1	Species 2	θ_1	θ_2	θ_A	T	χ^2	Pχ^2	WWH	P_{WWH}
Pseudoobscura . .	*persimilis*	228.8 160.4–338.1	115.7 84.8–158.8	259.8 182.8–346.7	0.229 0.159–0.296	198.3	0.088	57	0.015
Pseudoobscura . .	*bogotana*	315.9 143.5–1308.0	34.6 22.2–51.6	218.0 121.9–312.6	0.075 0.018–0.0125	176.7	0.308	43	0.060

Note.—For each contrast, the data were fit to the isolation model as described (Wang, Wakeley and Hey 1997). The estimated value for the primary parameters are shown, along with the 95% confidence intervals determined by simulation. θ_A is the estimate of the population mutation parameter for the ancestor of species 1 and 2. T is the estimated time of divergence between the two species in $2N_i$ generation units (where N_i is the estimate of the effective population size of species 1). The *P* values, for both the χ^2 and the Wang, Wakeley and Hey (WWH) test statistics, are the proportion of simulated values greater than or equal to the observed. The test is one-tailed because the focus is on detecting a departure from the model in the direction expected if historical gene flow had occurred.

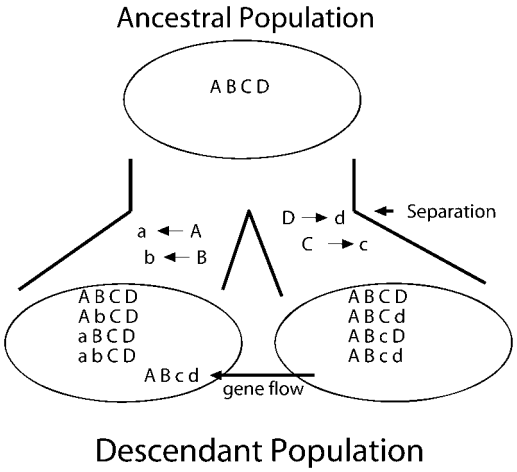

FIG. 3.—Each row of letters (e.g., A, B, C, and D) represents a haplotype, with upper case representing the ancestral state and lower case representing the derived state. After separation, each population experiences mutation and recombination so that multiple haplotypes are generated in each population. After gene flow, the shared polymorphisms in the recipient population (i.e., C/c and D/d) are in positive LD with each other and in negative LD with the other exclusive polymorphisms in that population (e.g., A/a and B/b).

jected, $\hat{\theta}_A$ is lower than the current estimate of *D. pseudoobscura* but much larger than the estimate of *D. p. bogotana*. This qualitatively matches our knowledge about the inferred population expansion in *D. pseudoobscura* and a founder effect in *D. p. bogotana*.

LD Tests of Gene Flow

In the two explanations of shared polymorphisms: (1) persistence since species coancestry, and (2) gene flow, we have differing expectations regarding patterns of LD within loci. According to the persistence model, shared polymorphisms are relatively old, at least as old as the time of population splitting, and they will have had more time, relative to nonshared polymorphisms, to recombine with other polymorphisms within each species. The general expectation is that LD among shared polymorphisms within species may be closer to zero than LD among other nonshared polymorphisms or between shared polymorphisms and nonshared polymorphisms. However, if polymorphisms have been introduced by gene flow at some time after the species began to diverge, then there will have been less time for recombination, and thus more LD is expected among these polymorphisms and between these polymorphisms and nonshared polymorphisms. Furthermore, we can also generate predictions regarding the sign of LD. If polymorphisms are rooted by an outgroup, they can be sorted into ancestral and derived character states. Among rooted polymorphisms, positive LD occurs when both ancestral bases or both derived bases of two polymorphic sites appear together more often than expected on the basis of their frequencies. Negative LD occurs when there is an excess of haplotypes carrying an ancestral base at one site and a derived base at the other site.

Figure 3 depicts two populations, each with two exclusive polymorphisms that have arisen since the onset of isolation. In general, gene flow between two populations need not change the haplotype distribution as it may involve haplotypes that are already shared. However, if gene flow moves polymorphisms that are exclusive to one population into the other, then it creates shared polymorphisms. The LD among these new shared polymorphisms, within the recipient species, will tend to be positive as the derived bases that immigrated together are preferentially associated with one another (fig. 3). Consider too the LD between these shared polymorphisms (C and D in fig. 3) and the other nonshared, exclusive polymorphisms (A and B in fig. 3). The introgressed haplotype will tend to carry ancestral bases at those sites where exclusive polymorphisms have arisen in the recipient species. Thus, the derived bases that come in via gene flow and cause shared polymorphisms will tend to be linked to ancestral bases at sites that support exclusive polymorphisms in the recipient species. A negative LD between shared polymorphisms and exclusive polymorphisms is expected.

Let DSS be the average of an estimate of LD that is found among all pairs of shared polymorphisms, and let DSX be the average among all pairs of sites for which one member is a shared polymorphism and the other is an exclusive polymorphism. Then for species *i*, which shares some polymorphisms with species *j*, and for locus *k*, we can consider the quantity $x_{(i,j)k} = \mathrm{DSS}_{(i,j)k} - \mathrm{DSX}_{(i,j)k}$. On the basis of the argument framed previously, in the case of gene flow, DSS should tend to be positive, whereas DSX should tend to be negative, and thus *x* should also tend to be positive and might be fairly large. However, if polymorphisms are not shared because of gene flow, then there should be relatively little LD among shared polymorphisms and similarly between shared and exclusive polymorphisms. Note that $x_{(i,j)k}$ can vary somewhat independently of $x_{(j,i)k}$, unless there is considerable gene flow, and thus the measure may be useful for assessing the directionality of gene flow.

In principle, *x* can be calculated for any measure of LD. In selecting a measure, it was necessary to consider the way in which different measures of LD vary as functions of allele frequencies. The simple isolation model assumes a constant population size, and the simulations under this model generate a particular distribution of allele frequencies. However, the broadly negative values of Tajima's *D* suggest that the species have undergone a recent population expansion. Regardless of the cause, the allele frequency distributions in these data sets are markedly shifted toward low-frequency polymorphisms. Thus, for statistical tests of the observed values of *x*, we selected a measure of LD that will be less sensitive to allele frequencies. We have chosen D′, which is equal to the conventional measure of LD divided by the maximum possible value given the allele frequencies (Lewontin 1964).

The actual expected sign and value of *x*, both with and without gene flow, is difficult to assess as it will depend on the relative ages of shared and nonshared polymorphisms and their relative allele frequencies, which will change depending on the time since popu-

Table 7
Linkage Disequilibrium Tests

	Species pair							
	PSEUDOOBSCURA-PERSIMILIS				*PSEUDOOBSCURA-BOGOTANA*			
	PSEUDOOBSCURA		*PERSIMILIS*		*PSEUDOOBSCURA*		*BOGOTANA*	
Locus	*Obs.*	*Sim.*	*Obs.*	*Sim.*	*Obs.*	*Sim.*	*Obs.*	*Sim.*
X008.....	NA	0.047	NA	0.011	−0.062	0.126	0.169	0.024
						0.191		0.181
X009.....	0.194	0.05	0.454	−0.007	NA	0.225	NA	0.111
		0.222		−0.036*				
X010.....	NA	0.288	NA	0.268	NA	0.318	NA	0.175
2001.....	0.578	0.075	0.321	0.035	NA	0.123	NA	0.043
		0.03*		0.113				
2002.....	NA	0.064	NA	0.044	0.838	0.126	0.565	0.003
						0.005*		0.016*
bcd.....	0.121	0.04	−0.128	0.01	−0.014	0.142	0.122	0.055
		0.247		0.779		0.776		0.303
rh1.....	0.372	0.091	0.165	0.054	NA	0.142	NA	0.046
		0.073		0.176				
2003.....	NA	0.107	NA	0.048	NA	0.201	NA	0.115
3002.....	0.077	0.071	0.366	0.06	0.035	0.054	0.149	0.012
		0.407		0.093		0.532		0.155
4002.....	NA	0.27	NA	0.262	NA	0.283	NA	0.086
4003.....	0.085	0.073	0.174	0.049	0.059	0.106	0.026	0.056
		0.391		0.196		0.535		0.404
Adh.....	0.194	0.059	0.203	0.038	0.178	0.07	0.167	0.02
		0.109		0.092		0.141		0.13
per.....	—	0.048	0.312	0.052	NA	0.137	NA	0.079
	0.074	0.744		0.171				
	NA	0.161	NA	0.142	NA	0.303	NA	0.184
hsp82.....	0.193	0.084	0.233	0.051	0.172	0.135	0.200	0.044
Mean		0.104		0.028*		0.323		0.096
Standard...	0.201	0.063	0.177	0.071	0.336	0.066	0.187	0.091
Deviation		0.045*		0.095		0.017*		0.141

Note.—Shown are the observed (Obs.) and simulated (Sim.) values of *x* (see text). Simulated values of *x* were obtained in the same simulations used for the isolation model (table 6). The estimated probability of observing a simulated value higher than the observed value of *x* is presented below the simulated value of *x*. Mean and standard deviations are based on all loci in each simulation for which *x* could be calculated. NA-Observed values are shown only if both DSS and DSX were calculated from at least four pairs of sites. Similarly, only those simulations that also had at least four pairs of sites for these quantities for a locus were used. * less than 5% of simulated values where higher than the observed value.

lation splitting. The argument behind *x* is not quantitative, and we do not have an expression for its expected value under a null isolation model. To test whether observed values of *x* are consistent with the null model, we used the same computer simulations of the isolation model that were used to test the WWH and χ^2 statistics (table 6). These simulations used the estimates of the population recombination rate listed in table 2. Table 7 shows the observed values of *x* for each locus, calculated using D′, for the species pairs *D. pseudoobscura-D. persimilis* and *D. pseudoobscura-D. p. bogotana* as well as the results for the overall mean and standard deviation (SD) of *x*. If there are only a small number of shared or exclusive polymorphisms, then these quantities cannot be calculated, and this was the case for several loci. The observed values are consistently positive across loci and across species comparisons, suggesting gene flow according to the argument presented previously. All species in both contrasts have observed values in the upper tail of the simulated distribution, but the overall mean value of *x* is significantly high only in *D. persimilis*, suggesting that this species has been the recipient of gene flow. Interestingly, although *D. pseudoobscura* did not have an overall significantly elevated mean of *x*, it did have an elevated standard deviation of *x* in both species contrasts.

For individual loci, only *2001* and *X009* have noticeably elevated values of *x* in the *D. pseudoobscura-D. persimilis* comparison: *2001* for *D. pseudoobscura* and *X009* for *D. persimilis*. In the *D. pseudoobscura-D. p. bogotana* comparison, *2002* has high values of *x* in both species. Interestingly, the test is not significant for *Adh*, the locus with the greatest impact on the WWH statistic, possibly reflecting high levels of recombination in this locus (table 2).

Cladistic and Qualitative Assessments of Gene Flow

Recent gene flow can also be inferred using cladistic and qualitative approaches. The cladistic approach (Slatkin and Maddison 1989) uses gene genealogies to estimate levels of gene flow. Unfortunately, the high levels of recombination in these data make it impossible to build accurate gene genealogies for each locus, thus reducing the utility of this approach. The qualitative approach looks for regions of sequence or full haplotypes that are atypical for the species on hand but that are identical or very similar to sequences that are typical for

the other species. No haplotypes were shared between any of the species, although several loci (*4002, X009, 3002, bcd, per*) show partial regions of sequence that resemble the typical sequence from the other species. The lack of full shared haplotypes indicate that the putative gene flow events occurred sufficiently long ago that there has been enough time since then for recombination to occur.

Discussion

The DPG approach to study the divergence of closely related species entails a full population genetic analysis of interspecific and intraspecific multilocus sequence data (Hey and Kliman 1993; Wang, Wakeley, and Hey 1997; Kliman et al. 2000). In principle, the approach can lead to inferences regarding the long-term effects of natural selection, gene flow, demography, and recombination on genetic variation at genomic regions that are, or are not, associated with the adaptive divergence of closely related species.

DPG analyses of the large multilocus sequence data set reported here have allowed us to generate an initial genome-wide portrait of the history of divergence of three closely related species: *D. pseudoobscura*, *D. persimilis*, and *D. p. bogotana*. The large variation across loci in patterns of fixed differences and shared polymorphisms leads us to reject the null model of speciation for *D. pseudoobscura* and *D. persimilis* but not for *D. pseudoobscura* and *D. p. bogotana*. We argue for gene flow as the main cause for the rejection of the isolation model in *D. pseudoobscura* and *D. persimilis*. However, factors other than gene flow could also increase the variance in fixed differences and shared polymorphisms across loci and in principle could lead us to reject that null model. Two models, in particular, natural selection at a subset of loci and population structure in the ancestor, could generate data patterns not consistent with the isolation model.

With regard to natural selection, HKA and McDonald-Kreitman tests found no evidence of selection in the data. HKA tests do not reject neutrality in any of the relevant ingroup comparisons, and, based on the McDonald-Kreitman test, the evidence for selection in *Adh* is weak. Nonneutral patterns were observed only in *4002* and *X010* (significant Tajima's *D*). The significant negative value of Tajima's *D* in *4002* and *X010* could be caused not only by selection but also by population expansion, a more plausible explanation supported by the consistently negative value of Tajima's *D* across loci. Therefore, we have no evidence that natural selection could have generated the observed variance in shared and fixed differences across loci in these data.

An informative comparison regarding the effect of selection is the recent DPG study of the *D. simulans* group (*D. simulans*, *D. sechellia*, *D. mauritiana*) using data from 14 genes (Kliman et al. 2000). In that study, the McDonald-Kreitman test was significant in three genes and the HKA test was significant for all three species. Despite the evidence of directional selection at about half of the loci the isolation model was not rejected, in contrast to the present case of *D. pseudoobscura* and *D. persimilis* where there is little evidence of directional selection, and yet the isolation model is rejected.

Population structure in the ancestral species could also increase the variance among genes. However, this explanation is, in effect, our conclusion, for we argue for a model in which an ancestral population diverged into two populations and engaged in gene flow during the process to lead to their becoming separate species. Thus, at some point, the distinction between that scenario and our explanation is a semantic one concerning when, during the history, it was appropriate to consider separate populations as separate species. It also bears noting, in this context, that populations that first experienced divergence were probably separated by considerable distance, or else selection against gene flow must have been quite strong. The reason is that these flies are highly mobile, and today we find no evidence of population structure at any of these loci over a range of 600 miles.

On balance, the simplest model of divergence, consistent with the data, is one that includes gene flow between *D. pseudoobscura* and *D. persimilis*. Additional evidence also supports this model (for additional discussion see Noor, Johnson, and Hey [2000]). First, *D. pseudoobscura* and *D. persimilis* are partially sympatric, they can hybridize in the lab, and F1 hybrids have been collected in the wild (Dobzhansky 1973; Powell 1983). Second, new data from regions of no recombination (mitochondrial and dot chromosome loci) provide clear evidence of gene flow between the two species (full haplotype sharing) (C. A. Machado and J. Hey, unpublished data). Third, the contrasting situation provided by the comparison between *D. p. bogotana* and *D. pseudoobscura*, provides indirect evidence to support our explanation. There, the isolation model is not rejected, providing a case that is quite consistent with the known history of geographical isolation between the two subspecies.

Regarding the timing of gene flow, the data do not suggest the occurrence of recent and pervasive gene flow between *D. pseudoobscura* and *D. persimilis*. Although they suggest that gene flow has occurred at a number of the surveyed loci, the lack of more evident cases of recent introgression (e.g., the sharing of complete haplotypes) suggests that what is observed mostly reflects older gene flow events. This may be surprising, given the potential for introgression via backcross of hybrid females, and the fact that most of the genome of these taxa can introgress between species (Noor et al. 2001). However, previous observations suggest low levels of hybridization in nature among these taxa (Dobzhansky 1951, 1973; Powell 1983), which are probably because of sexual isolation caused by strong female species discrimination (Merrell 1954; Noor 1996), a trait that probably evolved to reinforce isolation mechanisms between the two taxa (Noor 1995*b*).

Another piece of evidence showing that most of the gene flow is not recent is the observation that the proportions of shared polymorphism over the total number

of polymorphisms are almost identical in sympatric and allopatric populations of *D. pseudoobscura* (0.1616 vs. 0.1636). This is not surprising, given the high level of gene flow found among *D. pseudoobscura* populations, but if interspecific gene flow were currently ongoing at high rates, then we might see more evidence of it in sympatric populations.

Comparing Divergence and Isolation Mapping Studies of *D. pseudoobscura* and *D. persimilis*

Recently, Noor and co-workers (2001) used 14 codominant markers to map genomic regions associated with reproductive isolation (isolation map) between *D. pseudoobscura* and *D. persimilis*. All the markers linked to or located in the chromosomal inversions in the left and right arms of the X-chromosome (XL, XR) and the center of the second chromosome were strongly associated with barriers to gene exchange (fig. 2). A weak effect was observed in the center of the third chromosome, and the fourth and fifth chromosomes showed no detectable effects (fig. 2). These general results demonstrate that in laboratory conditions most of the genome of these two species can introgress.

Our results can also be interpreted as a kind of map—a divergence map showing which parts of the genome have diverged between species and which parts show evidence of gene flow. We can then ask: how does this divergence map compare with the isolation map developed by Noor et al. (2001)? If divergence is less for some genes because of gene flow, then we expect a correspondence between the two types of maps. Several of the markers used by Noor et al., correspond to the same microsatellite or RFLP loci for which we have collected flanking sequence data (*X009, 2002, 2003, 3002, 4002, 4003,* and *Adh*). We did not sequence any markers located within the XL fixed inversion, but two loci (*X008* and *per*) are located on that same chromosome arm, with one of them (*X008*) being physically close to the XL inversion breakpoint (fig. 2). Interestingly, the *X008* data show the largest number of fixed differences between *D. pseudoobscura* and *D. persimilis* and two shared polymorphisms that can be explained on the basis of recurrent mutation (table 4). The *period* locus did show evidence of one instance of gene flow some time ago, with a portion of one haplotype explaining all of the shared polymorphism (Wang and Hey 1996).

Three of the sequenced loci are located in the right arm of the X chromosome (*X009, Hsp82* and *X010*) (fig. 2), but none of these maps within the XR inversion (which is fixed among *D. pseudoobscura* and non–Sex-Ratio (SR) XR *D. persimilis* strains). As expected, the *Hsp82* and *X010* data suggest a fairly old cessation of gene flow between *D. pseudoobscura* and *D. persimilis* (these loci have the lowest values of population migration rates and the largest numbers of fixed differences after *X008*), whereas shared partial haplotypes suggest some recent introgression at *X009* (not shown).

The locus located in the fixed inversion of the second chromosome (*2002*) revealed no shared polymorphisms and a large number of fixed differences, consistent with complete isolation or an old termination of gene flow between *D. pseudoobscura* and *D. persimilis* (table 4). The other loci from the second chromosome (*2001, rh1, bcd,* and *2003*) show several shared polymorphisms and no fixed differences between *D. pseudoobscura* and *D. persimilis* (table 4). Two of the loci (*rh1* and *2001*) show high values of *x,* the measure of LD associated with shared polymorphisms (table 7). Interestingly, the same two loci show just one shared polymorphism but several fixed differences between *D. pseudoobscura* and *D. p. bogotana*. This observation suggests an older time for the cessation of gene flow at these loci between *D. pseudoobscura* and *D. p. bogotana* than between *D. pseudoobscura* and *D. persimilis*, which is supported by both estimates of net sequence divergence and population migration rates (table 5).

The one other locus that is associated with an inversion is *3002,* located in a region of the third chromosome where several inversions are known to occur. Interestingly, data from that locus show no fixed differences and a large number of shared polymorphisms between *D. pseudoobscura* and *D. persimilis* (table 4). The fact that *D. pseudoobscura* and *D. p. bogotana* also have a larger number of shared polymorphisms (13) may suggest that some of the shared variation between *D. pseudoobscura* and *D. persimilis* is ancestral. However, the data also show regions of the *3002* sequence from several *D. pseudoobscura* strains that resemble *D. persimilis* sequences (not shown), and in those regions all exclusive polymorphisms of *D. pseudoobscura* correspond to fixed derived bases in *D. persimilis*, suggesting recent introgression. The pattern observed in *3002* is intriguing, given its genomic location and the isolation mapping results which found a weak effect for reproductive isolation in that region of the genome (Noor et al. 2001). However, there are, in principle, no barriers for gene flow to occur across all the third chromosome of these species because both share the standard inversion arrangement, which is the most common third chromosome inversion arrangement of *D. pseudoobscura* in regions of sympatry with *D. persimilis* (Anderson et al. 1991; Powell 1992).

Apart from *4002,* the other markers located on the fourth chromosome (*4003* and *Adh*) have the largest numbers of shared polymorphisms and the highest estimates of population migration rate between *D. pseudoobscura* and *D. persimilis* (tables 4 and 5). This observation is consistent with the findings of Noor et al. (2001) and with the fact that this chromosome is colinear between the two species. The locus *4002* revealed a shared microsatellite allele with 15 dinucleotide repeats, a repeat number typical for *D. pseudoobscura* but quite different from that of *D. persimilis*, where the longest allele has only 10 repeats.

Thus, divergence and isolation maps are fairly consistent with each other. Genes that are located in genomic regions not associated with isolation phenotypes (Noor et al. 2001) show more evidence of introgression or more recent cessation of gene flow than those that are located in (or that are closely linked to) genomic regions associated with isolation phenotypes. This pat-

tern strongly suggests the action of natural selection preventing introgression at these regions. There are, however, two potential incongruences between the maps. First, the sequence data suggest the occurrence of gene flow and possibly recent introgression at *X009*, a locus near the XR inversion and which is significantly associated with several isolation phenotypes. The data also suggest some gene flow and recent introgression at *3002*, a locus located in the third chromosome inversion which is weakly associated with one isolation phenotype. One explanation for the apparent incongruence is that high levels of historical recombination and possibly not very large selection effects allowed *X009* and *3002* to introgress, despite their linkage to isolation factors. In addition, it is important to note that a similar comparison between the maps of *D. pseudoobscura* and *D. p. bogotana* is expected to show less congruence because of the history of old geographic isolation between the subspecies. If the current state of allopatry also existed during earlier stages of divergence, then gene flow should not have occurred at any loci.

Limitations of the Current Methods and Future Developments

The tools of our DPG approach have some limitations, particularly regarding the causes of shared polymorphisms. In this study, we have used tests of the isolation model of species divergence (WWH), patterns of LD, and qualitative assessments of shared haplotypes, to try to assess the impact of gene flow. However, none of these tests are ideal. The qualitative assessments are subjective, and the WWH and LD methods are strongly affected by the amount of recombination that is occurring (Wang, Wakeley, and Hey 1997). In order to carry out these tests, the simulations employed the γ estimates of 4Nc, the population recombination rate (Hey and Wakeley 1997) from table 2. These estimates are expected to underestimate the true value, on average (Hey and Wakeley 1997), which makes the statistical tests conservative with regard to rejection of the null model (which has no gene flow). If recombination is increased, then the variance of the WWH statistic and the LD measures under the null model goes down, and the apparent significance of the observations increases (results not shown, but available upon request). Nevertheless, the strong dependence of the tests on ad hoc estimates of recombination (i.e., recombination is not estimated simultaneously with other parameters) is a limitation.

The LD test described here is a useful addition to the basic DPG methodology. One of its main advantages is that it permits inferences on the direction of introgression for each locus, unlike the WWH test which addresses the pattern of variation for all loci simultaneously. However, although the overall LD test was significant across loci, the independent tests for each locus were significant for only two loci in the *D. pseudoobscura-D. persimilis* comparison (*X009, 2001*). These observations suggest that this test might not be powerful for studying species like *D. pseudoobscura* and *D. persimilis* that show large levels of recombination and for which gene flow at many loci seems to have ceased some time ago. Further, because the LD test is not based on explicit quantitative arguments, we have no expressions for the expected value of x under the null isolation model, and we do not know much about its power. Recent simulation results suggest that using patterns of LD among shared and exclusive polymorphic sites may not be a statistically powerful approach to test for gene flow, particularly when recombination is high (F. Depaulis, personal communication).

Another limitation is that we do not at present have ways to fit models with isolation and gene flow and to estimate the timing and magnitude of gene flow given such models. Developing such models is crucial, given the apparent inappropriateness of speciation models without gene flow.

In the future, divergence population genetics can be expected to rely more on maximum likelihood (ML) methods for testing the fit of the data to strict isolation model and constant–gene flow model of speciation. In principle, for the case of two diverged populations one could construct multilocus coalescent models that take into account mutation, recombination, population size changes, time since divergence, plus no migration (isolation model), constant levels of migration across loci (constant–gene flow model), or different levels of migration across loci (differential–gene flow model) (Wakeley and Hey 1998). Having the likelihood of the data, given each model and the estimated ML parameters, one could then compare the adequacy of the different models using likelihood ratio tests. Coalescent ML models that include migration and that can be adapted to multilocus cases have been implemented for the case of two populations (Beerli and Felsenstein 1999) or multiple populations (Beerli and Felsenstein 2001) with symmetric and nonsymmetric levels of migration. Recent developments using the Markov Chain Monte Carlo approach to better explore the genealogy space (Beerli and Felsenstein 1999; Nielsen 2000; Beerli and Felsenstein 2001) also provide hope for the implementation of better and more complex models in the near future.

Sequence Availability

Sequences have been deposited in GenBank with accession numbers AF450504–AF451008.

Acknowledgments

Special thanks to R. Favis for advice on inverse PCR, M. Noor, D. Álvarez, and M. Ruíz-García for providing the isofemale lines, and K. Shallop for help in the lab. S. Palumbi and two anonymous reviewers provided constructive comments on the manuscript. Research supported by NIH grant GM58060 to J.H.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

ANDERSON, E. 1949. Introgressive hybridization. Wiley, New York.

ANDERSON, E., and L. HUBRICHT. 1938. The evidence for introgressive hybridization. Am. J. Bot. **25**:396–402.

ANDERSON, W. W., J. ARNOLD, D. G. BALDWIN et al. (21 coauthors). 1991. Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **88**: 10367–10371.

ANDERSON, W. W., F. J. AYALA, and R. E. MICHOD. 1977. Chromosomal and allozymic diagnosis of three species of *Drosophila*. J. Hered. **68**:71–74.

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 **99**:305–309.

ASHBURNER, M. 1989. Drosophila, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

BEERLI, P., and J. FELSENSTEIN. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics **152**:763–773.

———. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. USA **98**:4563–4568.

BERNARDI, G., P. SORDINO, and D. A. POWERS. 1993. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. Proc. Natl. Acad. Sci. USA **90**:9271–9274.

BURTON, R. S., and B. N. LEE. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. Proc. Natl. Acad. Sci. USA **91**:5197–5201.

CARULLI, J. P., and D. L. HARTL. 1992. Variable rates of evolution among Drosophila opsin genes. Genetics **132**:193–204.

CLARK, A. G. 1997. Neutral behavior of shared polymorphism. Proc. Natl. Acad. Sci. USA **94**:7730–7734.

CLARKE, B. C., M. S. JOHNSON, and J. MURRAY. 1996. Clines in the genetic distance between two species of island land snails: how 'molecular leakage' can mislead us about speciation. Philos. Trans. R. Soc. Lond. Ser. B **351**:773–784.

DELLA TORRE, A., L. MERZAGORA, J. R. POWELL, and M. COLUZZI. 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics **146**:239–244.

DOBZHANSKY, T. 1936. Studies of hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21**:113–135.

———. 1937. Genetics and the origin of species. Columbia University Press, New York.

———. 1951. Experiments on sexual isolation in *Drosophila* X. Reproductive isolation between *Drosophila pseudoobscura* and *Drosophila persimilis* under natural and under laboratory conditions. Proc. Natl. Acad. Sci. USA **37**:792–796.

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

DOBZHANSKY, T., and T. EPLING. 1944. Taxonomy, geographic distribution and ecology of *Drosophila pseudoobscura* and its relatives. Pp. 1–46 *in* T. DOBZHANSKY and T. EPLING, eds. Contributions to the genetics, taxonomy, and ecology of *Drosophila pseudoobscura* and its relatives. Carnegie Institute of Washington, Washington, D.C.

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

DOBZHANSKY, T., and C. C. TAN. 1936. Studies on hybrid sterility III. A comparison of the gene arrangement in two species. Z. Indukt. Abstammungs.-Vererbungsl. **72**:88–114.

ENDLER, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, NJ.

EXCOFFIER, L., P. E. SMOUSE, and J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondiral DNA restriction data. Genetics **131**:479–491.

FELSENSTEIN, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution **35**:124–138.

HAMBLIN, M. T., and C. F. AQUADRO. 1999. DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. Genetics **153**:859–869.

HARE, M. P., and J. C. AVISE. 1998. Population structure in the american oyster as inferred by nuclear gene genealogies. Mol. Phylogenet. Evol. **15**:119–128.

HEY, J. 1994. Bridging phylogenetics and population genetics with gene tree models. Pp. 435–447 *in* B. SCHIERWATER, B. STREIT, G. P. WAGNER, and R. DESALLE, eds. Molecular ecology and evolution: approaches and applications. Birkhauser Verlag, Basel, Switzerland.

HEY, J., and R. M. 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.

HEY, J., and J. WAKELEY. 1997. A coalescent estimator of the population recombination rate. Genetics **145**:833–846.

HILTON, H., and J. HEY. 1997. A multilocus view of speciation in the *Drosophila virilis* group reveals complex histories and taxonomic conflicts. Genet. Res. **70**:185–194.

HILTON, H., R. M. 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., M. KREITMAN, and M. AGUADE. 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.

JIANG, C. X., P. W. CHEE, X. DRAYE, P. L. MORRELL, C. W. SMITH, and A. H. PATERSON. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). Evolution **54**:798–814.

KEITH, T. P., L. D. BROOKS, R. C. LEWONTIN, J. C. MARTINEZ-CRUZADO, and D. L. RIGBY. 1985. Nearly identical allelic distributions of xanthine dehydrogenase in two populations of *Drosophila pseudoobscura*. **2**:206–216.

KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN, A. J. BERRY, J. MCCARTER, J. WAKELEY, and J. HEY. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156**:1913–1931.

LEWONTIN, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49**:49–67.

LI, P., and J. BOUSQUET. 1992. Relative-rate test for nucleotide substitutions between two lineages. Mol. Biol. Evol. **9**: 1185–1189.

LIM, J. K. 1993. In situ hybridization with biotinylated DNA. Dros. Inf. Serv. **72**:73–77.

MAYNARD SMITH, J. 1966. Sympatric speciation. Am. Nat. **100**:637–650.

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

MERRELL, D. J. 1954. Sexual isolation between *Drosophila persimilis* and *Drosophila pseudoobscura*. Am. Nat. **88**:93–99.

MOORE, B. C., and C. E. TAYLOR. 1986. *Drosophila* of southern California III Gene arrangements of *Drosophila persimilis*. J. Hered. **77**:313–323.

MULLER, H. J. 1940. Bearings of the *Drosophila* work on systematics. Pp. 185–268 *in* J. HUXLEY, ed. The new systematics. Clarendon Press, Oxford, U.K.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NIELSEN, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics **154**:931–942.

NOOR, M. A. 1995*a*. Incipient sexual isolation in *Drosophila pseudoobscura bogotana* Ayala & Dobzhansky (Diptera: Drosophilidae). Pan-Pac. Entomol. **71**:125–129.

———. 1995*b*. Speciation driven by natural selection in Drosophila. Nature **375**:674–675.

———. 1996. Absence of species discrimination in *Drosophila pseudoobscura* and *D. persimilis* males. Anim. Behav. **52**:1205–1210.

NOOR, M. A., N. A. JOHNSON, and J. HEY. 2000. Gene flow between *Drosophila pseudoobscura* and *D. persimilis*. Evolution **54**:2174–2175.

NOOR, M. A., M. D. SCHUG, and C. F. AQUADRO. 2000. Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. Genet. Res. **75**:25–35.

NOOR, M. A., and K. R. SMITH. 2000. Recombination, statistical power, and genetic studies of sexual isolation in *Drosophila*. J. Hered. **91**:99–103.

NOOR, M. A. F., K. L. GRAMS, L. A. BERTUCCI, Y. ALMENDAREZ, J. REILAND, and K. R. SMITH. 2001. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution **55**:512–521.

NOOR, M. A. F., J. R. WHEATLEY, K. A. WETTERSTRAND, and H. AKASHI. 1998. Western North America obscura-group *Drosophila* collection data, summer 1997. Dros. Inf. Serv. **81**:136–137.

O'TOUSA, J. E., W. BAEHR, R. L. MARTIN, J. GIRSH, W. L. PAK, and M. L. ABBLEBURY. 1985. The *Drosophila* ninaE gene encodes an opsin. Cell **40**:839–850.

OFFRINGA, R., and F. VAN DER LEE. 1995. Isolation and characterization of plant genomic DNA sequences via (inverse) PCR amplification. Methods Mol. Biol. **49**:181–195.

ORR, H. A. 1987. Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. Genetics **116**:555–563.

———. 1989. Genetics of sterility in hybrids between two subspecies of Drosophila. Evolution **43**:180–189.

———. 1996. Dobzhansky, Bateson, and the genetics of speciation. Genetics **144**:1331–1335.

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.

———. 1992. Inversion polymorphisms in *Drosophila pseudoobscura* and *Drosophila persimilis*. Pp. 73–126 *in* C. B. KRIMBAS and J. R. POWELL, eds. *Drosophila* inversion polymorphism. CRC Press, Boca Raton, Fla.

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

RICE, W. R., and E. E. HOSTERT. 1993. Laboratory experiments on speciation: what have we learned in forty years? Evolution **47**:1637–1653.

RIESEBERG, L. H., J. WHITTON, and K. GARDNER. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152**:713–727.

RILEY, M. A., M. E. HALLAS, and R. C. LEWONTIN. 1989. Distinguishing the forces controlling genetic variation at the Xdh locus in *Drosophila pseudoobscura*. Genetics **123**:359–369.

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.

SCHAEFFER, S. W., and C. F. AQUADRO. 1987. Nucleotide sequence of the Adh region of *Drosophila pseudoobscura*: evolutionary change and evidence for an ancient gene duplication. Genetics **117**:61–73.

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

———. 1992*b*. Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics **132**:163–178.

SCHNEIDER, S., D. ROESSLI, and L. EXCOFFIER. 2000. Arlequin: a software for population genetics data analysis. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

SEEGER, M. A., and T. C. KAUFMAN. 1990. Molecular analysis of the bicoid gene from *Drosophila pseudoobscura*: identification of conserved domains within coding and noncoding regions of the bicoid mRNA. EMBO J. **9**:2977–2987.

SEGARRA, C., G. RIBÓ, and M. AGUADÉ. 1996. Differentiation of Muller's chromosomal elements D and E in the Obscura group of Drosophila. Genetics **144**:139–146.

SINGH, R. S. 1983. Genetic differentiation for allozyme and fitness characters between mainland and Bogota populations of *Drosophila pseudoobscura*. Can. J. Genet. Cytol. **25**:590–604.

SLATKIN, M., and W. P. MADDISON. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123**:603–613.

SOKAL, R. R., and F. J. ROHLF. 1981. Biometry: the principles and practice of statistics in biological research. W. H. Freeman, San Francisco.

STEINEMANN, M., and S. STEINEMANN. 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. Proc. Natl. Acad. Sci. USA **89**:7591–7595.

STOCKER, A. J., and C. D. KASTRITSIS. 1972. Developmental studies in Drosophila III. The puffing patterns of the salivary gland chromosomes of *D. pseudoobscura*. Chromosoma **37**:139–176.

STURTEVANT, A. H., and E. NOVITSKI. 1941. The homologies of the chromosome elements in the genus Drosophila. Genetics **26**:517–541.

TAJIMA, F. 1989*a*. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.

———. 1989*b*. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

TAN, C. C. 1935. Salivary gland chromosomes in the two races of *Drosophila pseudoobscura*. Genetics **20**:392–402.

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

WAKELEY, J., and J. HEY. 1997. Estimating ancestral population parameters. Genetics **145**:847–855.

———. 1998. Testing speciation models with DNA sequence data. Pp. 157–175 *in* R. DESALLE and B. SCHIERWATER, eds. Molecular approaches to ecology and evolution. Birkhäuser Verlag, Basel.

WANG, R. L., and J. HEY. 1996. The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. Genetics **144**:1113–1126.

WANG, R. L., J. WAKELEY, and J. HEY. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics **147**:1091–1106.

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

WELLS, R. S. 1996. Nucleotide variation at the Gpdh locus in the genus *Drosophila*. Genetics **143**:375–384.

STEPHEN PALUMBI, reviewing editor

Accepted November 27, 2001