

# Multilocus Analysis of Introgression Between Two Sympatric Sister Species of *Drosophila*: *Drosophila yakuba* and *D. santomea*

Ana Llopart,<sup>\*,1</sup> Daniel Lachaise<sup>†</sup> and Jerry A. Coyne<sup>\*</sup>

<sup>\*</sup>Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637 and <sup>†</sup>Laboratoire Populations, Génétique and Evolution, CNRS, 91198 Gif-sur-Yvette, France

Manuscript received July 16, 2004  
Accepted for publication June 9, 2005

## ABSTRACT

*Drosophila yakuba* is widely distributed in sub-Saharan Africa, while *D. santomea* is endemic to the volcanic island of São Tomé in the Atlantic Ocean, 280 km west of Gabon. On São Tomé, *D. yakuba* is found mainly in open lowland forests, and *D. santomea* is restricted to the wet misty forests at higher elevations. At intermediate elevations, the species form a hybrid zone where hybrids occur at a frequency of ~1%. To determine the extent of gene flow between these species we studied polymorphism and divergence patterns in 29 regions distributed throughout the genome, including mtDNA and three genes on the *Y* chromosome. This multi-locus approach, together with the comparison to the two allopatric species *D. mauritiana* and *D. sechellia*, allowed us to distinguish between forces that should affect *all* genes and forces that should act on *some* genes (*e.g.*, introgression). Our results show that *D. yakuba* mtDNA has replaced that of *D. santomea* and that there is also significant introgression for two nuclear genes, *yellow* and *salb*. The majority of genes, however, has remained distinct. These two species therefore do not form a “hybrid swarm” in which much of the genome shows substantial introgression while disruptive selection maintains distinctness for only a few traits (*e.g.*, pigmentation and male genitalia).

ACCORDING to the biological species concept (BSC), the coexistence of distinct entities in sympatry suggests a severe reduction in gene flow between them (DOBZHANSKY 1937; MAYR 1942; COYNE 1992; COYNE and ORR 1998). It is clear, however, that there is more introgression between species than early advocates of the BSC suspected (COYNE and ORR 2004). This observation of introgression has prompted the proposal of alternative species concepts. MALLETT (1995), for example, proposed the “genotypic cluster species concept” (GCSC), in which factors beyond reproductive isolation, such as stabilizing selection or historical inertia, are claimed to maintain discrete species in one habitat. As COYNE and ORR (2004) noted, many of these “other factors” are in fact forms of reproductive isolation. The main difference between the BSC and many proposed alternatives—including the GCSC and Wu’s (2001) “genic species concept”—is in how much gene flow between sympatric entities can occur while those entities still remain distinct. Sympatric, recognizably distinct clusters, for example, may differ at only a few loci while exchanging genes freely throughout the rest of the genome.

For several reasons it is important to determine the degree of gene exchange between sympatric or para-

patric clusters. First, the evolutionary independence of such taxa depends on the degree to which they can exchange generally adaptive alleles. Second, we want to know whether specific regions of the genome introgress more readily than others. For example, regions linked to genes causing species-specific adaptations or hybrid sterility may be limited in their ability to move between species (TUCKER *et al.* 1992; MACHADO *et al.* 2002). Chromosomal rearrangements that differ between species may also serve as “traps” for genes causing hybrid incompatibilities and thus also show limited introgression (RIESEBERG *et al.* 1995; NOOR *et al.* 2001; MACHADO *et al.* 2002; NAVARRO and BARTON 2003). Third, it has been claimed that introgression can be a source of genetic variation that allows species to adapt to new environments and hence can serve as an engine of adaptation (ARNOLD 1997). Finally, we want to determine on the genic level the applicability of various species concepts: for example, Are the genomes of sympatric, hybridizing species largely impermeable to new variation, as the BSC might predict?

Answers to these questions have generally come from hybrid zones, which show in general that introgression, while more frequent than previously suspected, is limited between sympatric taxa (COYNE and ORR 2004). Unfortunately, there is a dearth of hybrid zones in the most genetically well-studied group, *Drosophila*. LACHAISE *et al.* (2000), however, recently described a hybrid zone between the sister species *Drosophila yakuba* and *D. santomea*

<sup>1</sup>Corresponding author: Department of Biological Sciences, University of Iowa, 212 Biology Bldg. (BB), Iowa City, IA 52242.  
E-mail: ana-llopart@uiowa.edu

on the Island of São Tomé, a small volcanic island 280 km off the coast of Gabon. *D. yakuba* is widely distributed in sub-Saharan Africa and in the islands near the continent (including Madagascar), while *D. santomea* is endemic to São Tomé. The most striking morphological difference between *D. yakuba* and *D. santomea* is the pigmentation pattern: *D. yakuba* has the characteristic pattern of the *D. melanogaster* group (females' yellow abdomens are striped with black, while those of males have black tips), while *D. santomea* completely lacks this dark pigmentation (LACHAISE *et al.* 2000; LLOPART *et al.* 2002). Other traits, such as male genital morphology and the number of sex comb teeth, also distinguish these species (LACHAISE *et al.* 2000; COYNE *et al.* 2004). Moreover, *D. yakuba* is highly polymorphic for chromosome inversions, with the right arm of chromosome 2 showing the most polymorphism (LEMEUNIER and ASHBURNER 1976). In contrast, *D. santomea* shows no polymorphism for chromosomal arrangements, and no fixed inversions distinguish *D. yakuba* and *D. santomea* (LACHAISE *et al.* 2000). Molecular data suggest that *D. santomea* arose on the island after the colonization by its common ancestor with *D. yakuba* ~400,000 years ago (CARIOU *et al.* 2001; LLOPART *et al.* 2002). Tentative molecular evidence suggests that the current presence of *D. yakuba* on the island reflects a more recent colonization (CARIOU *et al.* 2001).

As expected from its status as an open forest/savannah species, *D. yakuba* usually occurs at low elevations on São Tomé, mainly in areas cleared by humans beginning in the 16th century. In contrast, *D. santomea* inhabits the wet forests found at higher elevations. On the largest mountain in the island, Pico de São Tomé, *D. yakuba* is found only below 1450 m, while *D. santomea* occurs between 1150 and 2024 m. Between 1150 and 1450 m—an area that coincides with the transition between agricultural land and virgin rainforest—the species coexist, with the abundance ratio of *D. yakuba*/*D. santomea* changing from 2 to 0.05 as one moves upward through the zone (LACHAISE *et al.* 2000). Hybrids have been reported in this zone at a frequency of ~1%; these were diagnosed as hybrids by their possession of intermediate genitalia and pigmentation. In the laboratory, these species show several forms of reproductive isolation, which include mate discrimination (LACHAISE *et al.* 2000; COYNE *et al.* 2002), conspecific sperm precedence (CHANG 2004), and sterility that conforms to Haldane's rule: F<sub>1</sub> hybrid males are completely sterile and F<sub>1</sub> hybrid females are partly or fully fertile (COYNE *et al.* 2004).

Here, we report the patterns of polymorphism and divergence between *D. yakuba* and *D. santomea* on the basis of the examination of a substantial number of genes spread throughout the genome. Our aim is to determine the extent of introgression for nuclear and mitochondrial genomes and to compare the amount and patterns of introgression with those predicted from laboratory data on the genetics of hybrid sterility.

Because our multilocus approach investigates genome-wide variation, it allows us to distinguish between forces that should affect *all* genes and forces that should act on some but not all genes (*e.g.*, introgression). This multilocus approach, pioneered by J. Hey and colleagues, has proved useful in revealing the pattern and extent of gene exchange in nature (HEY and KLIMAN 1993; HILTON *et al.* 1994; WANG *et al.* 1997; KLIMAN *et al.* 2000; MACHADO *et al.* 2002; BROUGHTON and HARRISON 2003; RAMOS-ONSINS *et al.* 2004).

## MATERIALS AND METHODS

**Loci and flies analyzed:** We used *D. yakuba* or *D. melanogaster* sequences to design primers to sequence 45 nuclear genes (Table 1) in *D. yakuba* Tai18 and *D. santomea* STO.4 strains (LLOPART *et al.* 2002). Polymorphism data in both species were collected for 28 out of 45 randomly selected nuclear genes and the mitochondrial region *ND5-ND4* (see Figure 1), with a sample size ranging from 21 to 32 chromosomes depending on the loci (Table 2). The "sympatric" specimens of *D. yakuba* and *D. santomea* were captured in the hybrid zone (elevation 1250 m) in the Obo Natural Reserve in March 2003 on São Tomé Island. At the same time, "allopatric" individuals of *D. yakuba* and *D. santomea* were collected, respectively, in a garden outside São Tomé City (elevation 5 m) and on Pico Calvario (a shoulder on the main mountain of Pico de São Tomé at 1566 m). In each of these locations only one species is found, although on Pico Calvario *D. yakuba* occurs within a few kilometers. Upon collection, all flies were immediately preserved in absolute ethanol until DNA extraction was performed in the laboratory. For mtDNA and Y chromosome genes, we analyzed a more geographically diverse sample of isofemale lines. The *D. yakuba* lines include strains from the Ivory Coast (Tai18 and TM34), Cameroon (CAM), Gabon (GAB), and Príncipe Island (ANTON-1) and from the zones of allopatry (SJI and BAR2) and sympatry (BOSU, COST2, SA1, and OBAT5) with *D. santomea* on São Tomé Island. For *D. santomea*, the sample includes one allopatric isofemale line (CAR1566.6) from Pico Calvario (elevation 1566 m) and nine additional sympatric lines. Eight of these were collected in the hybrid zone on Pico de São Tomé (STO.4, STO.10, STO.15, STO.18, COST 1235.1, OBAT 1200.13, OBAT 1200.14, and LAGO 1482.11), and one (QUIJA 650.1) came from Rio Quija, in southwest São Tomé, an area that also harbors *D. yakuba*. More detailed information on some of these strains can be found in COYNE *et al.* (2002). We also used the isofemale strains *D. mauritiana* B (from Mauritius Island) and *D. sechellia* SY 001 (from the Seychelles archipelago) to distinguish between selective constraints and introgression as explanations for sequence similarity in regions of low frequency of crossing over. To assess introgression between *D. yakuba* and *D. teissieri*, we also sequenced the *ND5-ND4* mitochondrial region and the Y chromosome genes in the Brazzaville 8 strain (collected in the Republic of the Congo).

**DNA extraction and sequencing:** We extracted DNA from single flies using the Puregene DNA isolation kit for paraffin-embedded tissue (Gentra Systems, Minneapolis) and performed PCR amplifications using ~25 ng of genomic DNA. To ensure the specific amplification of the genes on the Y chromosome, we performed PCR reactions using genomic DNA extracted from females as a control. PCR products were purified using the Wizard MagneSil PCR clean-up system (Promega, Madison, WI) and sequenced directly with an ABI

TABLE 1

Synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) substitutions per site between *D. yakuba* and *D. santomea*

Gene	$K_s$	$K_a$	Size (bp) <sup>b</sup>
<i>bmb</i> <sup>a</sup>	0.047	0.0033	1248
<i>f</i>	0.063	0.0047	843
<i>Hex-A</i> <sup>a</sup>	0.018	0	1290
<i>per</i> <sup>a</sup>	0.097	0	801
<i>rux</i> <sup>a</sup>	0.066	0.027	945
<i>sn</i>	0.04	0	1134
<i>sog</i> <sup>a</sup>	0.033	0	1200
<i>su(f)</i> <sup>a</sup>	0.021	0	414
<i>y</i> <sup>a</sup>	0.0029	0	1041
<i>CG17629</i> <sup>a</sup>	0.023	0	936
<i>Dhc-Yh3</i> <sup>a</sup>	0.0104	0	924
<i>PpY1</i> <sup>a</sup>	0.017	0	486
<i>Adh</i>	0.0047	0.004	630
<i>barr</i>	0.033	0.0057	954
<i>Hex-C</i> <sup>a</sup>	0.14	0.0022	1272
<i>His3</i> <sup>a</sup>	0.0073	0	408
<i>Kr</i> <sup>a</sup>	0.013	0.0056	1350
<i>l(2)gl</i> <sup>a</sup>	0.033	0.0027	903
<i>Ngf</i> <sup>a</sup>	0.029	0.0037	741
<i>Pgi</i>	0.013	0	717
<i>Rad1</i> <sup>a</sup>	0.054	0.0016	741
<i>Rep4</i> <sup>a</sup>	0.063	0.0053	846
<i>RpL27A</i> <sup>a</sup>	0.015	0.0029	414
<i>salr</i> <sup>a</sup>	0.045	0.0012	1005
<i>Sara</i> <sup>a</sup>	0.036	0.0073	915
<i>trp1</i>	0.062	0.0027	1011
<i>wkg</i>	0.023	0.014	867
<i>AP-50</i> <sup>a</sup>	0.039	0	1128
<i>dib</i>	0.097	0.0027	1359
<i>dos</i>	0.032	0	867
<i>Est6</i> <sup>a</sup>	0.077	0.011	600
<i>hb</i>	0.036	0.0005	2199
<i>Hsc70-4</i>	0.018	0	1086
<i>Lsp1-γ</i> <sup>a</sup>	0.027	0	597
<i>Mlc1</i> <sup>a</sup>	0.091	0	78
<i>Pgm</i>	0.053	0.0012	1185
<i>RpL14</i>	0.019	0	411
<i>Rpn5</i>	0.039	0.0012	963
<i>sfl</i> <sup>a</sup>	0.09	0	864
<i>Sod</i>	0.13	0.0034	360
<i>SsI1</i> <sup>a</sup>	0.033	0.0037	654
<i>Xdh</i> <sup>a</sup>	0.051	0.0031	1269
<i>ymp</i>	0.066	0.0078	504
Total <sup>c</sup>	0.044	0.0029	38160

<sup>a</sup> Loci included in the polymorphism survey.

<sup>b</sup> Size of the coding region analyzed.

<sup>c</sup> Divergence based on the concatenated sequence including the regions *AnnX* and *v*, 12 and 18 bp long, respectively.

PRISM 3100 genetic analyzer (Applied Biosystems, Foster City, CA) after we performed cycle sequencing reactions using Big Dye 3.0 (Applied Biosystems). Both strands were sequenced. We edited sequences with the software Sequencher 3.0 (Gene Codes, Ann Arbor, MI) and aligned them using the ClustalX program (THOMPSON *et al.* 1997). Haplotypes were reconstructed (STEPHENS and DONNELLY 2003) using the Phase 2.0 program (STEPHENS *et al.* 2001), excluding indels. We deposited all newly obtained sequences in GenBank, EMBL,

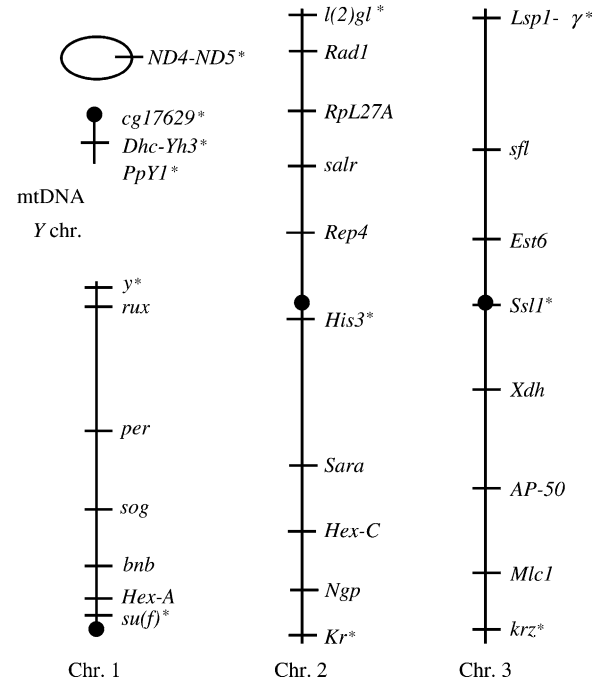


FIGURE 1.—Chromosomal locations of the loci included in the polymorphism survey. Cytological positions were inferred from *D. melanogaster* and the comparative salivary-gland banding maps of LEMEUNIER and ASHBURNER (1976). Solid circles symbolize the centromeric regions of the chromosomes. Loci in regions of reduced crossing over are indicated by asterisks.

and DDBJ database libraries under accession nos. AY804458–AY804777, AY804780–AY804885, and AY804888–AY805047.

**Data analysis:** Basic polymorphism analyses were performed using DnaSP 4.0 (ROZAS *et al.* 2003). To determine whether regions under study were heterogeneous in their polymorphism-to-divergence ratios, and to test whether the global frequency spectrum of polymorphisms conforms to the neutral expectation, we used the HKA program (HEY and KLIMAN 1993). This program obtains the significance of the observed statistics ( $\chi^2_{\text{HK}}$  and  $D$ ), using multilocus coalescent simulations.

To detect the presence of significant introgression between *D. yakuba* and *D. santomea*, we used two different strategies. In regions with nonreduced frequency of crossing over we fit the data to an isolation model (*i.e.*, a model of genic patterns based on the assumption that there is no gene exchange between species) (WAKELEY and HEY 1997), following WANG *et al.* (1997) and using the Wakeley-Hey (WH) program. This model assumes that all variation is neutral and that population size remains constant over time, but allows for differences in selective constraints among loci (WANG *et al.* 1997). The parameters of the model ( $\theta$  for the ancestral and descendant populations and  $\tau$ , the time since physical separation) are estimated by choosing values that most closely equate expectations with observations; hence tests based on this model are conservative because all shared variation is assumed to be ancestral polymorphism. A significant excess of shared variation is the unequivocal fingerprint of introgression in *D. yakuba* and *D. santomea*, for they most likely went through a period of allopatry and came in contact secondarily recently. In particular, we calculated a  $\chi^2$ -statistic on the basis of the difference between expected and observed numbers of shared polymorphisms:  $\chi^2_{\text{SS}}$ . We generated the null distribution of this statistic under the isolation model by multilocus coalescent simulations (WANG *et al.* 1997) with recombination, and we

**TABLE 2**  
**Polymorphism data summary**

Locus	Species	$n^a$	$S^b$	Syn <sup>c</sup>	Rep <sup>c</sup>	$\theta_t^d$	$\theta_s^d$	$D^e$	$L$ (bp) <sup>f</sup>
A. Regions of reduced frequency of crossing over									
<i>ND5-ND4</i>	<i>D. yakuba</i>	11	8	6	2	0.0016	0.0046	-1.27	1657
	<i>D. santomea</i>	10	4	4	0	0.00085	0.0032	-0.52	
<i>Dhc-Yh3</i>	<i>D. yakuba</i>	11	0	0	0	0	0	—	1219
	<i>D. santomea</i>	10	0	0	0	0	0	—	
<i>CG17629</i>	<i>D. yakuba</i>	11	1	1	0	0.0004	0.0016	0.67	936
	<i>D. santomea</i>	10	3	1	2	0.0011	0.001	-1.03	
<i>Pp1Y1</i>	<i>D. yakuba</i>	11	1	0	1	0.0007	0	-1.13	490
	<i>D. santomea</i>	10	0	0	0	0	0	—	
<i>y</i>	<i>D. yakuba</i>	16	6	6	0	0.0019	0.0078	-1.17	965
	<i>D. santomea</i>	16	0	0	0	0	0	—	
<i>su(f)</i>	<i>D. yakuba</i>	16	1	0	0	0.0005	0.0008	0.15	607
	<i>D. santomea</i>	16	0	0	0	0	0	—	
<i>l(2)gl</i>	<i>D. yakuba</i>	16	1	1	0	0.0005	0.0021	1.03	608
	<i>D. santomea</i>	16	3	2	1	0.0015	0.0042	0.17	
<i>His3</i>	<i>D. yakuba</i>	16	3	0	0	0.0018	0.0022	-0.65	496
	<i>D. santomea</i>	16	0	0	0	0	0	—	
<i>Kr</i>	<i>D. yakuba</i>	16	3	3	0	0.002	0.0086	0.01	456
	<i>D. santomea</i>	16	0	0	0	0	0	—	
<i>Lsp1-γ</i>	<i>D. yakuba</i>	16	3	2	0	0.0018	0.006	-0.41	508
	<i>D. santomea</i>	16	4	0	3	0.0024	0.002	-1.55*	
<i>Ssl1</i>	<i>D. yakuba</i>	16	2	2	0	0.0012	0.0051	-0.58	501
	<i>D. santomea</i>	16	1	1	0	0.0006	0.0025	0.15	
<i>krz<sup>s</sup></i>	<i>D. yakuba</i>	16	0	—	—	0	0	—	165
	<i>D. santomea</i>	16	2	—	—	0.0036	0.0036	-1.50*	
B. Regions of nonreduced frequency of crossing over									
<i>rux</i>	<i>D. yakuba</i>	16	28	20	7	0.0090	0.023	-1.62*	940
	<i>D. santomea</i>	16	21	17	3	0.0067	0.02	-1.42	
<i>per</i>	<i>D. yakuba</i>	16	3	0	0	0.001	0.0013	-1.35	905
	<i>D. santomea</i>	14	5	0	1	0.0017	0.0019	-1.36	
<i>sog</i>	<i>D. yakuba</i>	16	25	23	2	0.007	0.024	-0.63	1100
	<i>D. santomea</i>	16	18	17	1	0.0049	0.018	-0.59	
<i>bnb</i>	<i>D. yakuba</i>	14	33	24	9	0.012	0.037	-0.86	837
	<i>D. santomea</i>	16	25	18	7	0.009	0.026	-0.97	
<i>Hex-A</i>	<i>D. yakuba</i>	16	3	3	0	0.0014	0.058	-0.63	639
	<i>D. santomea</i>	16	9	8	1	0.0042	0.016	-0.67	
<i>Rad1</i>	<i>D. yakuba</i>	16	16	14	1	0.011	0.04	-1.45	428
	<i>D. santomea</i>	16	10	8	1	0.007	0.024	-1.48*	
<i>RpL27A</i>	<i>D. yakuba</i>	14	20	2	1	0.0076	0.01	0.13	822
	<i>D. santomea</i>	14	19	2	1	0.0073	0.0099	-1.17	
<i>salr</i>	<i>D. yakuba</i>	16	28	25	2	0.0086	0.038	-0.72	869
	<i>D. santomea</i>	16	26	22	3	0.0090	0.033	-0.58	
<i>Rep4</i>	<i>D. yakuba</i>	16	25	19	6	0.0093	0.032	-0.76	808
	<i>D. santomea</i>	16	14	11	3	0.0052	0.019	-1.61*	
<i>Sara</i>	<i>D. yakuba</i>	16	21	13	7	0.0068	0.02	-0.26	887
	<i>D. santomea</i>	16	12	8	4	0.0040	0.012	0.52	
<i>Hex-C</i>	<i>D. yakuba</i>	16	21	19	2	0.011	0.04	0.65	552
	<i>D. santomea</i>	16	13	12	1	0.0071	0.021	-0.73	
<i>Ngp</i>	<i>D. yakuba</i>	16	3	3	0	0.0012	0.005	0.47	767
	<i>D. santomea</i>	16	5	4	1	0.0020	0.0067	-0.32	
<i>sfl</i>	<i>D. yakuba</i>	16	40	40	0	0.014	0.058	-0.86	867
	<i>D. santomea</i>	16	37	37	0	0.013	0.053	-0.59	
<i>Est6</i>	<i>D. yakuba</i>	16	37	22	8	0.016	0.043	0.17	687
	<i>D. santomea</i>	16	27	15	9	0.012	0.027	0.12	
<i>Xdh</i>	<i>D. yakuba</i>	16	32	23	9	0.012	0.036	-0.14	782
	<i>D. santomea</i>	16	24	18	6	0.0092	0.028	-1.95**	

(continued)

TABLE 2  
(Continued)

Locus	Species	$n^a$	$S^b$	Syn <sup>c</sup>	Rep <sup>c</sup>	$\theta_t^d$	$\theta_s^d$	$D^e$	$L$ (bp) <sup>f</sup>
B. Regions of nonreduced frequency of crossing over									
AP-50	<i>D. yakuba</i>	16	35	25	2	0.01	0.035	-0.63	1025
	<i>D. santomea</i>	16	29	22	1	0.0085	0.029	-1.84*	
Mlc1	<i>D. yakuba</i>	16	7	0	0	0.0059	0.0071	-1.52	359
	<i>D. santomea</i>	16	10	0	0	0.0084	0.01	-1.54*	

\*  $P < 0.05$ , \*\*  $P < 0.01$ . —, values could not be estimated because of lack of informative sites.

<sup>a</sup> Sample size.

<sup>b</sup> Number of polymorphic sites.

<sup>c</sup> Number of synonymous (syn) and nonsynonymous (rep) polymorphic sites estimated using NEI and GOJOBORI (1986).

<sup>d</sup> WATTERSON'S (1975) estimate of total and silent (synonymous and noncoding) heterozygosity.

<sup>e</sup> TAJIMA'S (1989b)  $D$ -statistic. Significance was calculated with coalescent simulations without recombination.

<sup>f</sup> Size of the sequenced region.

<sup>g</sup> This locus does not include the coding region sequence.

obtained the statistical significance of the observed  $\chi_{SS}^2$  by comparing it to the null distribution generated. We estimated recombination between sites ( $\rho = 4N_e r$ ) within DNA fragments using the maximum-composite-likelihood method implemented in the Maxhap software (HUDSON 2001). The estimates obtained ranged between 0.0025 and 0.17 in *D. yakuba* and between 0.0025 and 0.5 in *D. santomea*.

Loci on the tips and near centromeres of chromosomes show a reduced frequency of crossing over (Figure 1) in species of the *D. melanogaster* subgroup (ASHBURNER 1989; TRUE *et al.* 1996). This reduction, usually associated with decreased levels of polymorphism, compromises the detection of shared variation. Thus in these regions and also in mtDNA and Y chromosome genes, our strategy to detect introgression was based on comparing divergence estimates between *D. yakuba* and *D. santomea* with those between *D. mauritiana* and *D. sechellia*. We derived these estimates for synonymous and nonsynonymous sites ( $K_s$  and  $K_a$ , respectively), as well as confidence intervals, using the program *K*-Estimator 5.5 (COMERON 1999). To estimate overall divergence between species pairs, we constructed a concatenated sequence for genes on each chromosome separately and also for all genes taken together as a single unit. To obtain maximum-likelihood estimates of the population migration rates we fit the polymorphism and divergence data on genes of reduced crossing over to an isolation with migration (IM) model (HEY and NIELSEN 2004). This Markov chain Monte Carlo method assumes no intragenic recombination and yields estimates of multiple population parameters. Convergence of the Markov chain simulations was assessed by comparing the results and using different starting points.

## RESULTS

**Intraspecific nucleotide variation:** We collected polymorphism data in *D. yakuba* and *D. santomea* from 29 regions: 28 nuclear and 1 mitochondrial (Table 2). *D. yakuba* shows significantly higher levels of within-species variation than does *D. santomea*: the weighted average values of WATTERSON'S (1975) estimator of  $4N_e\mu$  (where  $N_e$  is the effective population size and  $\mu$  is the neutral mutation rate) per site ( $\theta$ ) are 0.0055 for *D. yakuba* and 0.0044 for *D. santomea* (Wilcoxon signed rank test using individual genes;  $Z = -2.22$ ,  $P = 0.026$ ). This result is

consistent with *D. santomea* having a smaller  $N_e$ , although this reduction of only 20% contrasts with a large difference in *current* population size inferred from the geographical distribution of both species.

In both species, silent variation (synonymous and noncoding) for genes on the X chromosome tends to be smaller than that for genes on the autosomes. The difference is close to the neutral expectation (a 25% reduction for the X chromosome compared to autosomes) based on their difference in  $N_e$ :  $\theta_{\text{autosomal}} = 0.025$  vs.  $\theta_X = 0.015$  for *D. yakuba*, and  $\theta_{\text{autosomal}} = 0.018$  vs.  $\theta_X = 0.012$  for *D. santomea*. Polymorphism in the mitochondrial region *ND5-ND4* in *D. yakuba* and *D. santomea* is similar to that seen in *D. melanogaster* (RAND *et al.* 1994; RAND and KANN 1996). Consistent with previously published data from *D. melanogaster* (ZUROVCOVA and EANES 1999), in both *D. santomea* and *D. yakuba* we see extremely low levels of variation on the Y chromosome.

We also compared nucleotide variation between allopatric and sympatric populations of *D. yakuba* and *D. santomea* for all genic regions under study. If introgression has occurred recently in sympatric populations, one expects a certain degree of genetic differentiation between sympatric and allopatric populations of the different species.  $K_{ST}^*$  values (HUDSON *et al.* 1992), a measure of population substructure, are consistently close to 0. Permutation tests (HUDSON *et al.* 1992) show that only 5 of the 29 regions have significantly nonzero values of  $K_{ST}^*$  (*salr* and *Kr* in *D. yakuba* and *bnb*, *Est6*, and *sara* in *D. santomea*), and none of these remain significant after Bonferroni correction for multiple tests (RICE 1989). Even for these 5 regions, differences between allopatric and sympatric populations account only for 2.3–5.8% of the total molecular variance (EXCOFFIER *et al.* 1992). Overall, our results show no genetic difference between sympatric and allopatric populations of *D. yakuba* and *D. santomea* on São Tomé Island. Consistently, the fraction of shared/total polymorphisms is similar

for allopatric and sympatric flies (7.7 and 7.9%, respectively).

**Tests for neutrality:** Under the neutral theory of molecular evolution, regions of the genome that evolve rapidly should also show high levels of intraspecific variation (KIMURA 1983), a prediction that can be evaluated with the HKA test (HUDSON *et al.* 1987). For data from several loci one can test the heterogeneity of the polymorphism-to-divergence ratio across all the regions under study (HEY and KLIMAN 1993). When applied to the 29 loci with polymorphism data in both *D. yakuba* and *D. santomea*, this test shows a significant departure from the neutral expectation ( $\chi^2_{\text{HKA}} = 73.32$ ,  $P < 0.001$ ; see MATERIALS AND METHODS for details). This departure is also observed when the five regions with significant population structure are excluded ( $\chi^2_{\text{HKA}} = 61.64$ ,  $P < 0.005$ ), and it is due primarily to regions located on tips and near centromeres of chromosomes. If these regions are excluded from the analysis, the data fit the neutral expectations ( $\chi^2_{\text{HKA}} = 19.02$ ,  $P = 0.71$ ).

We also used Tajima's *D* (TAJIMA 1989b) to determine whether the frequency of mutations in our sample is consistent with neutral expectations. Most genes in our data set show negative *D*-values in both species (implying an excess of low-frequency variants over that expected under the neutral theory), although none of these values are statistically significant after multiple test correction. However, a multilocus analysis (HEY and KLIMAN 1993) shows that in both species the mean *D*-value among loci departs from neutral expectations in a negative direction ( $D = -0.49$ ,  $P = 0.009$  for *D. yakuba* and  $D = -0.89$ ,  $P < 0.0001$  for *D. santomea*). We observe the same pattern after excluding regions that have a reduced frequency of crossing over ( $D = -0.59$ ,  $P = 0.009$  for *D. yakuba*, and  $D = -0.95$ ,  $P < 0.0001$  for *D. santomea*) or regions with significant population structure ( $D = -0.53$ ,  $P = 0.007$  for *D. yakuba*, and  $D = -1.03$ ,  $P < 0.0001$  for *D. santomea*). Significant negative values of *D* imply either purifying selection or population expansion (TAJIMA 1989a,b).

**Analysis of introgression:** Speciation can be thought of as the fragmentation of an ancestral population into two descendant populations that, in time, acquire reproductive isolation (*i.e.*, genetic fixed differences). If the two incipient species exchange genes, the accumulation of fixed differences is retarded and shared polymorphisms are introduced. Yet this same pattern is also expected in isolated incipient species, due to shared variants inherited from their common ancestor, particularly if speciation did not involve a strong bottleneck. There is also a third source of shared variation: independent parallel mutations, which may be a special problem in species that are very polymorphic.

To test for independent mutations as a factor that may mimic introgression, we calculated the expected number of recurrent mutations using a hypergeometric distribution conditioning on the number of polymorphisms in each species and the number of sites under

analysis (CLARK 1997; KLIMAN *et al.* 2000). In *D. yakuba* and *D. santomea*, 56 of the 719 polymorphisms are shared (Table 3), while the expectations are just 5.85 under a model in which variation is evenly distributed along all sites, and slightly more—13.01—when only silent sites are considered. The number of observed shared polymorphisms substantially exceeds both of these expectations ( $P < 1 \times 10^{-6}$ ). Therefore a significant fraction of shared variation cannot be explained by parallel mutation and must be accounted for by either introgression or common ancestry.

To claim shared variation as evidence of introgression we ought to test it against a model that takes into account possible ancestral polymorphism. This approach based on detecting an excess of shared variation, however, is less informative in regions where polymorphism is severely reduced. In *Drosophila*, as in many other eukaryotes, regions of reduced frequency of crossing over show a strong reduction of intraspecific variation because of recurrent selection (AGUADÉ *et al.* 1989; STEPHAN and LANGLEY 1989; BERRY *et al.* 1991; BEGUN and AQUADRO 1992; LANGLEY *et al.* 1993; DVORAK *et al.* 1998; KRAFT *et al.* 1998; NACHMAN *et al.* 1998; PRZEWORSKI *et al.* 2000). Thus to investigate introgression between *D. yakuba* and *D. santomea* we analyzed separately genes of nonreduced and reduced crossing over.

**Regions of nonreduced crossing over:** To determine whether the number of shared polymorphisms exceeds the expectations estimated using the isolation model proposed by WAKELEY and HEY (1997) we used the  $\chi^2_{\text{SS}}$ -statistic (see MATERIALS AND METHODS). We performed two sets of tests. In the first, we contrast  $\chi^2_{\text{SS}}$  for each locus against the expected distribution of this statistic under the isolation model, and in the second we add the  $\chi^2_{\text{SS}}$ -values of all loci to assess whether the observed variance among loci differs from that expected under the isolation model. Tests of individual loci show a highly significant excess of shared polymorphisms in the *sabr* region ( $\chi^2_{\text{SS}} = 12.82$ ,  $P = 0.0002$ ). The same tendency is observed for the *sfl* gene ( $\chi^2_{\text{SS}} = 5.55$ ,  $P = 0.042$ ) on the third chromosome, but this is not significant after the Bonferroni correction. *sabr* is on the second chromosome and, as a consequence, the isolation model is also rejected when we test the whole chromosome ( $\chi^2_{\text{SS}} = 29.56$ ,  $P = 0.0075$ ). This result should not be interpreted as evidence for whole-chromosome exchange between *D. yakuba* and *D. santomea*, but as an indication that gene flow involving this chromosome is heterogeneous, with at least one locus having a number of shared polymorphisms that is not consonant with the background level. Despite the significant departure from the isolation model for *sabr* and the second chromosome, the overall patterns of variation among all regions of nonreduced frequency of crossing over are compatible with the isolation model (Table 5), which suggests that introgression is not pervasive.

**Regions of reduced crossing over:** Loci on the tips and near centromeres of chromosomes show a severe

**TABLE 3**  
**Shared and fixed variation between *D. yakuba* and *D. santomea***

Locus	SS <sup>a</sup>	F <sup>b</sup>	Sx1 <sup>c</sup>	Sx2 <sup>d</sup>	Locus	SS <sup>a</sup>	F <sup>b</sup>	Sx1 <sup>c</sup>	Sx2 <sup>d</sup>
A. Regions of reduced frequency of crossing over									
<i>ND5/ND4</i>	2 (0.01)	0	6	2	<i>His3</i>	0 (0)	6	3	0
<i>Y chr.</i> <sup>e</sup>	0 (0)	8	2	3	<i>Kr</i>	0 (0)	6	3	0
<i>y</i>	0 (0)	0	6	0	<i>Lsp1-γ</i>	0 (0.02)	4	3	4
<i>su(f)</i>	0 (0)	7	1	0	<i>Ssl1</i>	0 (0)	6	2	1
<i>l(2)gl</i>	0 (0)	1	1	3	<i>krz</i>	0 (0)	3	0	2
B. Regions of nonreduced frequency of crossing over									
<i>ruv</i>	3 (0.51)	22	25	19	<i>Sara</i>	0 (0.27)	2	20	12
<i>per</i>	0 (0.02)	13	3	5	<i>Hex-C</i>	2 (0.38)	0	19	11
<i>sog</i>	2 (0.33)	1	23	16	<i>Ngp</i>	0 (0.02)	3	3	5
<i>bnb</i>	5 (0.67)	2	28	20	<i>sft</i>	12 (0.81)	2	28	25
<i>Hex-A</i>	0 (0.04)	3	3	9	<i>Est6</i>	5 (0.99)	5	31	22
<i>Rad1</i>	2 (0.26)	1	14	8	<i>Xdh</i>	3 (0.78)	1	29	21
<i>RpL27A</i>	0 (0.46)	2	20	19	<i>AP-50</i>	7 (0.60)	4	28	22
<i>salr</i>	10 (0.23)	0	13	14	<i>Mlc1</i>	2 (0.11)	3	5	8
<i>Rep4</i>	1 (0.39)	4	24	13					

The number of polymorphisms expected by recurrent independent mutations is indicated in parentheses (CLARK 1997).

<sup>a</sup> Shared polymorphisms.

<sup>b</sup> Fixed differences.

<sup>c</sup> Exclusive polymorphisms in *D. yakuba*.

<sup>d</sup> Exclusive polymorphisms in *D. santomea*.

<sup>e</sup> *Y* chromosome genes *DhcYh3*, *CG17629*, and *Pp1Y1* were pooled.

reduction of polymorphism in *D. yakuba* and *D. santomea* (Mann-Whitney test;  $Z = -3.76$ ,  $P = 0.0002$  for *D. yakuba* and  $Z = -4.31$ ,  $P < 0.0001$  for *D. santomea*). This is consistent with these regions having a reduced frequency of crossing over, but makes difficult the detection of introgression through shared variation. *Divergence*, however, can be used to examine gene flow, for one of the consequences of introgression is the reduction of interspecific differences in the regions being exchanged. Such a reduction is also expected under strong selective constraints. To distinguish between these two possibilities, we compared estimates of genetic divergence between *D. yakuba* and *D. santomea* with estimates from a pair of *allopatric* species in the same group: species in which gene flow is very unlikely. These species are *D. mauritiana* and *D. sechellia*. Each is endemic to an Indian Ocean island very far from the other (Mauritius and the Seychelles, respectively), and these well-studied species almost certainly arose after the independent colonization of the islands by their common ancestor with the mainland African species *D. simulans* (Figure 2).

Ideally, one would compare species pairs separated by similar genetic distances. (Using this criterion assumes that introgression is not so pervasive as to distort the genetic differences at *all* loci used to calculate divergence times.) To check that this was so for the species pairs used in our test, we therefore estimated divergence between *D. yakuba* and *D. santomea* using newly obtained sequences of 45 nuclear genes (Table 1). We also estimated divergence between *D. mauritiana* and *D. sechellia*,

using published sequences of 25 genes (supplementary Table S1 at <http://www.genetics.org/supplemental/>). For *D. yakuba* and *D. santomea*, the estimated number of nonsynonymous substitutions per site,  $K_a$ , is 0.0029 [95% confidence intervals (C.I.) 0.0022–0.0035], while divergence at synonymous sites,  $K_s$ , is 0.044 (C.I. 0.039–0.048). For *D. mauritiana* and *D. sechellia*, estimates for  $K_a$  and  $K_s$  are 0.0083 (C.I. 0.0069–0.0097) and 0.047 (C.I. 0.04–0.053), respectively. Clearly the  $K_s$  values, which are critical in determining species age, are very similar for these species pairs. Thus, comparative analysis of interspecific divergence between these two pairs of species allows us to determine what genes—if any—are able to cross the species boundary between *D. yakuba*

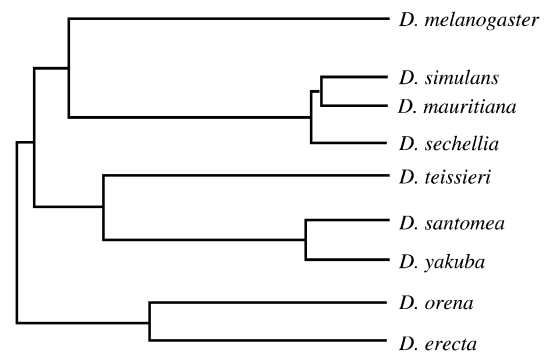


FIGURE 2.—Schematic of the phylogenetic relationships among the nine species of the *D. melanogaster* subgroup (KLIMAN *et al.* 2000; LACHAISE *et al.* 2000; PARSCH 2003).

**TABLE 4**  
**Substitutions between *D. yakuba*-*D. santomea* and**  
***D. mauritiana*-*D. sechellia* in regions with reduced**  
**frequency of crossing over**

Locus	Location <sup>a</sup>	<i>D. yakuba</i> - <i>D. santomea</i> <sup>b</sup>		Size (bp)
		<i>D. yakuba</i>	<i>D. santomea</i>	
<i>ND4-ND5</i> <sup>c</sup>	mtDNA	1	62	1658
<i>Y chr.</i> <sup>d</sup>	<i>Y</i>	7	6	2155
<i>y</i> <sup>e</sup>	<i>X</i>	1	6	1041
<i>su(f)</i>	<i>X</i>	7	0	717
<i>l(2)gl</i>	2	9	9	903
<i>His3</i>	2	5	11	811
<i>Kr</i>	2	10	9	1329
<i>LspI-γ</i>	3	7	4	659
<i>SsII</i>	3	15	7	754
<i>krz</i>	3	2	4	207
<i>ey<sup>f</sup></i>	4	6	5	918

<sup>a</sup> In *D. melanogaster*.

<sup>b</sup> Number of substitutions based on single sequence comparison (*D. yakuba* Tai18 and *D. santomea* STO.4).

<sup>c</sup> *D. mauritiana* and *D. sechellia* sequences were obtained from GenBank (accession nos. AF200830 and AF200832).

<sup>d</sup> Includes *Dhc-Yh3* and *CG17629*, but not *Pp1Y1* because we could not amplify it in *D. mauritiana* and *D. sechellia*. *D. mauritiana* and *D. sechellia* sequences for *Dhc-Yh3* were obtained from GenBank (accession nos. AF136265 and AF136264).

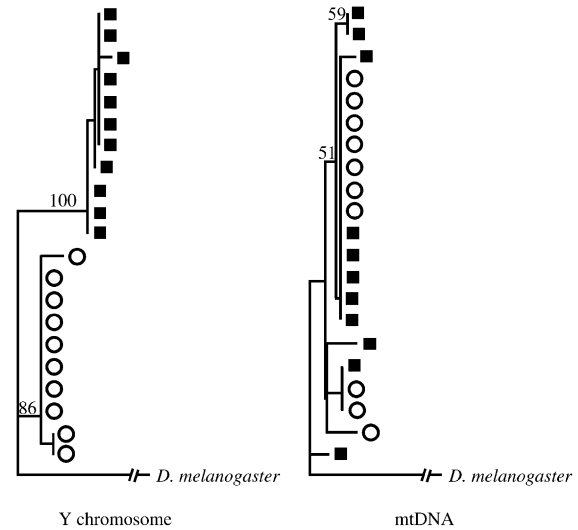
<sup>e</sup> The *D. mauritiana* sequence was obtained from GenBank (accession no. AJ300669).

<sup>f</sup> *D. mauritiana* and *D. sechellia* sequences for the intron 2 of the *eyeless* gene were obtained from GenBank (accession nos. AF491794 and AF491797).

and *D. santomea*. Such introgression is revealed by a reduction of divergence compared to that seen between truly allopatric species.

We compared estimates of divergence in the 12 regions of reduced crossing over for *D. yakuba* and *D. santomea* with estimates in the same regions between *D. mauritiana* and *D. sechellia* (Table 4). MtDNA and genes on the *Y* chromosome have, *a priori*, similar abilities to reveal the genetic history of species, as they both experience an equivalent reduction in  $N_e$  compared to that in autosomal loci. However, mtDNA and the *Y* chromosome show strikingly different patterns of genetic divergence between *D. yakuba* and *D. santomea*. The number of substitutions in the *Y* chromosome is similar for the *D. yakuba*/*D. santomea* and *D. mauritiana*/*D. sechellia* comparisons (Table 4). In contrast, *D. yakuba*/*D. santomea* show a >60-fold reduction in mitochondrial divergence compared to that in *D. mauritiana*/*D. sechellia* ( $K = 0.0006$  vs.  $K = 0.039$ ,  $P < 0.0001$  and Table 4). The reduction is >50-fold if only synonymous sites are considered. These results suggest that introgression of mtDNA has occurred between *D. yakuba* and *D. santomea*.

Analysis of population differentiation (HUDSON *et al.* 1992) indicates that the mtDNA of *D. yakuba* is not genetically different from that of *D. santomea* ( $K_{ST}^* = -0.01$ ,  $P = 0.54$ ). Indeed, 9 out of 10 lines analyzed in



**FIGURE 3.**—Gene genealogies from the *Y* chromosome and mtDNA. The reconstruction was done using the neighbor-joining algorithm (SAITOU and NEI 1987). The numbers above the branches are bootstrap values based on 1000 replicates (>50%). Solid squares, *D. yakuba*; open circles, *D. santomea*.

*D. santomea* have the same mtDNA haplotype as do 5 out of 6 *D. yakuba* lines collected in São Tomé. In addition, mtDNA haplotypes of *D. yakuba* collected in mainland Africa are significantly different from those of *D. yakuba* and *D. santomea* lines collected on the island ( $K_{ST}^* = 0.072$ ,  $P = 0.02$ ), another hint of mtDNA introgression. As this observation implies, mainland and island strains of *D. yakuba* are also genetically different ( $K_{ST}^* = 0.88$ ,  $P = 0.044$ ). Consequently, gene genealogies based on mtDNA show alleles of *D. santomea* clustering with alleles of *D. yakuba*, a pattern that stands in strong contrast to that seen for the *Y* chromosome (Figure 3).

LACHAISE *et al.* (2000) also reported a high level of similarity among sequences of the mitochondrial gene *cytochrome b* in *D. yakuba*, *D. santomea*, and *D. teissieri* (MONNEROT *et al.* 1990). We investigated whether this similarity could reflect an introgression of mtDNA between more than two species by examining synonymous divergence at nuclear and mtDNA genes among these species (supplementary Table S2 at <http://www.genetics.org/supplemental/>). Synonymous divergence at nuclear genes gives us an estimate of the times of these species splits, which we can then compare to divergence times estimated from mtDNA and *Y* chromosome genes. If there is introgression of mtDNA, one expects reduced divergence times for mtDNA compared to those for nuclear genes.

We sequenced the *ND5-ND4* region and two genes in the *Y* chromosome (*Dhc-Yh3* and *CG17629*; we could not amplify *Pp1Y1* in *D. teissieri*) in *D. teissieri*. In addition, we combined our data with sequences for 26 nuclear genes in *D. teissieri* and *D. yakuba* taken from GenBank. The overall estimate for synonymous and nonsynonymous divergence between these 26 nuclear genes for *D. yakuba*

and *D. teissieri* is  $K_s = 0.087$  (C.I. 0.079–0.094) and  $K_a = 0.017$  (C.I. 0.015–0.019), respectively.  $K_s$  for the mitochondrial *ND5–ND4* region between *D. yakuba* and *D. teissieri* is 0.011, the lowest value among all genes analyzed. Therefore, our results are consistent with a reduced mtDNA divergence between *D. yakuba* and *D. teissieri* compared to that in nuclear genes, but we also find fixed differences between the pair *D. yakuba/D. santomea* on one hand and *D. teissieri* on the other in the regions studied (4 for the *ND5–ND4* region of mtDNA, 32 for the *Y* chromosome). These results suggest that while there was ancient introgression of mtDNA between the ancestor of *D. teissieri* and that of *D. santomea/D. yakuba*, there has also been a more recent introgression of mtDNA between *D. yakuba* and *D. santomea*, presumably after *D. yakuba* experienced secondary contact with *D. santomea* on São Tomé.

For the remaining genes in regions of reduced crossing over, exon 2 of *yellow* is the only one showing significantly less divergence between *D. yakuba* and *D. santomea* than between *D. mauritiana* and *D. sechellia* after correcting for multiple tests (RICE 1989) ( $K = 0.001$  vs.  $K = 0.0058$ ,  $P < 0.0001$ ). In fact, *D. santomea* shows no variation in this region: all sequences are identical to that of the most common haplotype in *D. yakuba*. Moreover, the multilocus HKA test (HEY and KLIMAN 1993) shows a significant heterogeneity of the polymorphism-to-divergence ratio across these eight regions when the *Y* chromosome and mtDNA loci are excluded ( $\chi^2_{\text{HK}} = 29.39$ ,  $P < 0.005$ ). This heterogeneity is not observed when only the *yellow* region is excluded from the analysis ( $\chi^2_{\text{HK}} = 14.26$ ,  $P = 0.19$ ). We therefore suggest that the lack of fixed differences in the *yellow* gene between *D. yakuba* and *D. santomea* is the main cause of the heterogeneity among these regions having reduced crossing over. The similarity between these species in the *yellow* region, however, is not observed in a fragment located ~5 kb upstream from the start codon (data not shown). Indeed, we obtained the sequence for this additional 0.7-kb-long fragment in the same flies for which the second exon of *yellow* was studied, and found eight fixed differences between *D. yakuba* and *D. santomea* and no shared polymorphism (see DISCUSSION). The introgression in the *yellow* region may reflect its lack of effect on hybrid sterility (see DISCUSSION).

Finally, we obtained maximum-likelihood estimates (MLEs) of the population migration rates between *D. yakuba* and *D. santomea* by fitting polymorphism and divergence data on genes with reduced frequency of crossing over to the recently proposed IM model (HEY and NIELSEN 2004). MLEs of population migration rates ( $M$ ;  $M = 2 N_e m$ , where  $m$  is the migration rate) are  $m_1 = 0.0008$  and  $m_2 = 0.062$  for *D. santomea* to *D. yakuba* and for *D. yakuba* to *D. santomea*, respectively. Ninety-five percent confidence intervals of  $m_2$  rule out a scenario with no introgression of genes from *D. yakuba* to *D. santomea*.

**Autosomal vs. sex-linked genes:** The study of hybrid male sterility in *D. yakuba* and *D. santomea* revealed at least three male “sterility genes” that map to the *X* chromosome, as well as a significant effect of foreign *Y* chromosomes on sterility (COYNE *et al.* 2004). Thus, all else being equal, the movement of sex chromosomes across the hybrid zone should be limited. We therefore examined whether the *Y* and *X* chromosomes show less introgression between these species than do the autosomes. For the *Y* chromosome, none of the three genes studied shows shared variation and all show fixed interspecific differences (Table 3). Thus there is no evidence of recent gene flow between *D. yakuba* and *D. santomea* involving this chromosome. This result is, of course, consistent with the expectation that the *Y* chromosome is the least likely to introgress because it not only causes sterility by itself, but also is restricted to males, which are completely sterile as  $F_1$  hybrids and largely sterile in backcrosses (COYNE *et al.* 2004).

To compare the amount of introgression between the *X* chromosome and autosomes, we calculated the number of shared polymorphisms and fixed differences between *D. yakuba* and *D. santomea* for these chromosomes. We used only variation in regions with “normal” frequencies of crossing over to avoid artificially distorting the polymorphism-to-divergence ratio differentially for *X* chromosomes and autosomes. The ratio of shared to fixed differences is significantly lower for genes on the *X* chromosome than for those on the autosomes (10/41 for the *X* chromosome and 44/27 for the autosomes,  $G$ -test,  $G = 22.72$ ,  $P < 0.0001$ ). While this result is consistent with reduced introgression of the *X* chromosome, there is an alternative explanation: the difference in  $N_e$  between *X* chromosomes and autosomes is also expected to reduce both  $\theta$  and the number of shared polymorphisms and to increase the number of fixed differences for *X*-linked compared to autosomal regions. To properly judge the likelihood of introgression, we must correct for this difference in  $N_e$ . To do so, we estimated the number of shared and fixed differences for the *X* chromosome and autosomes under the isolation model (9.47 and 28.79, respectively, for the *X* chromosome, and 44.52 and 39.20, respectively, for the autosomes). The deviation of the observed values from the expectations is estimated with a  $\chi^2$ -statistic, and the statistical significance is obtained by comparing the observed  $\chi^2$ , 8.98, with a null distribution of this statistic obtained by multilocus coalescent simulations. Note that in very closely related species, the variance of the outcome of the coalescent process has a substantial effect on the number of fixed differences. Although we observe a higher than expected number of fixed differences for the *X* chromosome (and a deficit of fixed differences on the autosomes), the deviation from expectation is not statistically significant ( $P = 0.18$ ). Therefore, although the difference between the *X* chromosome and the autosomes is in the direction

**TABLE 5**  
**Testing the isolation model in *D. yakuba* and *D. santomea***

	$\theta_1$	$\theta_2$	$\theta_A$	$\chi_{SS}^2$	$P_{SS}^a$	$\chi_T^2$	$P_T^a$	WWH	$P_{WWH}^a$
<i>X</i>	18.93 0.12–38.01	15.48 0.12–32.41	109.59 54.98–165.82	3.77	0.70	54.24	0.24	26	0.16
<i>2</i>	34.37 21.00–53.07	23.69 15.59–34.62	46.98 24.54–69.47	29.56	0.0075	50.97	0.034	14	0.063
<i>3</i>	28.26 9.52–52.80	22.75 7.92–42.60	75.63 40.56–120.3	4.91	0.80	14.40	0.97	14	0.73
Total <sup>b</sup>	76.91 50.29–106.76	61.15 40.76–83.68	232.37 169.42–308.15	33.98	0.46	134.12	0.4	34	0.12
Total <sup>c</sup>	78.72 45.65–113.24	62.54 37.22–87.60	321.30 221.64–411.42	4135.15	<0.0001	4382.35	<0.0001	34	0.21

Estimated values of parameters of the isolation model are shown (WAKELEY and HEY 1997) as performed in WANG *et al.* (1997) with 95% confidence intervals based on coalescent simulations.

<sup>a</sup> Fractions of simulations with values of  $\chi_{SS}^2$  (see text),  $\chi_T^2$  (see KLIMAN *et al.* 2000 for details), and WWH (see WANG *et al.* 1997 for details) statistics equal to or more extreme than the values observed.

<sup>b</sup> Only genes in regions of nonreduced crossing over (17 genes).

<sup>c</sup> Regions of nonreduced and reduced crossing over (mtDNA and *Y* chromosome included, 29 genes).

indicating less introgression of the former between *D. yakuba* and *D. santomea*, this difference is not significant.

## DISCUSSION

We examined intra- and interspecific variation for 29 randomly selected loci regions of *D. yakuba* and *D. santomea*. Overall, the data are not compatible with an isolation model that assumes no gene flow between these species after they began to diverge ( $P = 0.0001$ , Table 5). Therefore, the genomes of these entities have not remained completely distinct, almost certainly because of past hybridization. Nevertheless, this gene flow has not been extensive: only 3 of the 29 regions (mtDNA and the nuclear regions containing the *yellow* and *sabr* loci) show statistically significant evidence for introgression. Thus these species do not form a “hybrid swarm” on São Tomé in which there is pervasive introgression in much of the genome, while disruptive selection maintains distinctness for only a few traits (*e.g.*, pigmentation and male genitalia). However, we must bear in mind that the introgression we are able to detect today is unlikely to reflect very recent gene flow, for neutral mutations require substantial time to reach detectable frequencies in a population (KIMURA and OHTA 1972; NEI and FELDMAN 1972).

**Introgression in regions of nonreduced crossing over:** The *sabr* gene, located in 2L, shows evidence of introgression between *D. yakuba* and *D. santomea*. Its level of shared polymorphism and the absence of fixed differences between species stand out against a chromosomal background practically devoid of shared variation. There is a strong reduction of the ratio of shared-to-exclusive polymorphisms on the right arm of

the second chromosome, which has the most inversion polymorphism in *D. yakuba*, compared to that on chromosome 3 (2/70 *vs.* 29/219,  $G = 6.4$ ,  $P = 0.011$ ). These results suggest that gene flow is particularly restricted in 2R, an observation consistent with the hypothesis that genes contributing to reproductive isolation accumulate faster in chromosomal inversions than in colinear regions of the genome (RIESEBERG *et al.* 1995; NOOR *et al.* 2001; NAVARRO and BARTON 2003). In agreement with this proposal, MACHADO *et al.* (2002) reported that regions that are distinct between *D. pseudoobscura* and *D. persimilis* are also located in chromosomal inversions. Clearly the effect of inversions on limiting gene flow is greatest if they are fixed between species, but polymorphic inversions may also contribute, to a lesser degree, to restricting gene flow. The combination of shared variation for *sabr* on 2L and the *absence* of shared variation on 2R leads to the result that patterns of variation in the entire second chromosome are incompatible with the isolation model—a result not seen in the *X* or third chromosomes.

The isolation model is not rejected when all genes in regions of nonreduced crossing over are considered. The estimated values of the parameter  $\theta_A$  of the isolation model (Table 5), however, suggest that the ancestral species of *D. yakuba* and *D. santomea* had an effective population size larger than that of either of the derived species, a result that is inconsistent with the observed negative values of  $D$  (*i.e.*, population expansion). Thus, it is conceivable that the degree of introgression between *D. yakuba* and *D. santomea* may have been underestimated.

**Introgression in regions with reduced crossing over:** Divergence between *D. yakuba* and *D. santomea* for the

mitochondrial region *ND5-ND4* is strongly reduced compared to both the rest of the genes in these species' genomes and the divergence between *D. mauritiana* and *D. sechellia*. This reduction is even more conspicuous when we consider that mitochondrial genes evolve 4.5–9 times *faster* than nuclear genes (MORIYAMA and POWELL 1997) at synonymous sites and is a sound indication that *D. yakuba* and *D. santomea* have exchanged mtDNA. In addition, the *D. yakuba* and *D. santomea* mtDNA haplotypes found on São Tomé are more similar to each other than to mtDNA haplotypes found on the African mainland. We therefore conclude that the lack of fixed differences in the *ND5-ND4* region between *D. yakuba* and *D. santomea*, together with the observation of shared haplotypes on the island, strongly suggests that these species have exchanged mitochondrial genomes.

There is ample evidence from many groups that mtDNA introgresses between species more frequently than does nuclear DNA (SMITH 1992; FERRIS *et al.* 1993; BERNATCHEZ *et al.* 1995; TAYLOR and MCPHAIL 2000; SHAW 2002; BALLARD and WHITLOCK 2004). The reason for this pattern is not yet understood, but HUDSON and COYNE (2002) and COYNE and ORR (2004) suggest two reasons. First, mitochondrial loci appear to have primarily “housekeeping” functions, such as production of tRNAs and enzymes used in respiration. Selection on such loci may be largely divorced from the external (but not the internal) environment, and therefore mitochondria from one species may function relatively well on the genetic background of a closely related species. In addition, even “neutral” nuclear DNA may often fail to introgress during hybridization because it is linked to other genes that are divergently selected between species.

Because mainland African populations of *D. yakuba* can be thought of as an outgroup, it is very likely that the direction of introgression of mtDNA has been from *D. yakuba* into *D. santomea*. (The alternative direction of introgression would require the unlikely scenario that *D. santomea* mtDNA introgressed into *D. yakuba* on São Tomé, and this mtDNA subsequently spread throughout *D. yakuba* on the African mainland.) This result is consonant with laboratory studies showing that, while *D. yakuba* and *D. santomea* produce hybrids in both directions, interspecific matings occur much more frequently between *D. yakuba* females and *D. santomea* males than vice versa (COYNE *et al.* 2002).

The mtDNA haplotype found in a single strain in *D. santomea* collected on the southwest part of São Tomé is identical to one *D. santomea* haplotype (though not the most frequent) detected on Pico de São Tomé, suggesting that *D. yakuba* haplotypes may be fixed on the entire island. The similarity of mtDNA between *D. yakuba* and *D. santomea* may have resulted from either genetic drift or selection following introgression. The presence of some polymorphisms in the mtDNA of the *D. yakuba* and *D. santomea* lines from São Tomé Island, as well as the

detection of shared variation on this organelle, points to a neutral explanation. On the other hand, given the average time to fixation for neutral mutations (KIMURA 1983), and the relatively short divergence time between these species, it remains plausible that selection may have played some role in the invasion of the *D. yakuba* mtDNA through a “transspecies-selective sweep” (HILTON *et al.* 1994; STEPHAN *et al.* 1998; MACHADO and HEY 2003). Under this scenario, an ancient selective sweep of *D. yakuba* mtDNA through both species was followed by the appearance of new mutations on mtDNA that later introgressed between the species. It is unlikely that this mitochondrial sweep could have been associated to the spread of a *Wolbachia* infection, as proposed in *D. simulans* (TURELLI and HOFFMANN 1995; BALLARD 2004), because *D. yakuba* and *D. santomea* do not show cytoplasmic incompatibility when infected by natural strains of *Wolbachia* (ZABALOU *et al.* 2004).

As with mtDNA, divergence in the *yellow* region between *D. yakuba* and *D. santomea* is reduced compared to that seen between *D. mauritiana* and *D. sechellia*. That this disparity is also due to introgression rather than to selection on amino acids is supported by the observed fivefold difference in  $K_s$  ( $K_s = 0.0029$  for *D. yakuba-D. santomea* and  $K_s = 0.015$  for *D. mauritiana-D. sechellia*). In addition, all *D. santomea yellow* haplotypes are identical to the most frequent haplotype seen in the *D. yakuba* sample.

Apart from introgression, there is one other hypothesis that could explain the pattern observed in *yellow*: stronger selective constraints at synonymous sites at *yellow* in *D. yakuba* and *D. santomea* than in the *D. simulans* clade. Indeed, TAKANO-SHIMIZU (1999) has suggested stronger selective constraints at *yellow* synonymous sites in the *D. yakuba* lineage compared to the *D. melanogaster* lineage (using *D. orena* as outgroup; Figure 2) on the basis of patterns of synonymous codon usage. However, there is no reduction in the number of synonymous substitutions in the *D. yakuba* lineage compared to the *D. melanogaster* lineage (TAKANO-SHIMIZU 1999). Moreover, synonymous polymorphism at *yellow* in *D. yakuba* is not drastically reduced compared to that in the other nuclear regions studied, with 14 of the 29 regions analyzed showing lower polymorphism than this locus (10/12 genes in regions of reduced crossing over and 4/12 in regions of nonreduced crossing over). Therefore, we conclude that the lack of divergence in the *yellow* locus between *D. yakuba* and *D. santomea* largely reflects introgression. This is also consistent with evidence that this region apparently does not contain genes affecting male sterility in *D. yakuba/D. santomea* hybrids (COYNE *et al.* 2004). However, we do not detect introgression in a region only 5 kb upstream from the start codon of *yellow*. It has been suggested that the telomere end of the *D. yakuba X* chromosome shows a substantial increase in recombination (~14-fold higher) compared to that of the *D. melanogaster X* chromosome (TAKANO-SHIMIZU

1999). Thus, we tentatively propose that introgression is restricted to the coding region of the *yellow* gene, while a 5' flanking region remains distinct. This in turn suggests that introgression can be a quite localized phenomenon on the chromosome (WANG *et al.* 1999; TING *et al.* 2000; TAKAHASHI *et al.* 2001).

**Autosomal vs. sex-linked genes:** We also tested the hypothesis that there is less interspecific introgression of genes on the X chromosome than on the autosomes because of the large effect of the X chromosome on hybrid male sterility in *D. yakuba* and *D. santomea* (COYNE *et al.* 2004). If the X chromosome contributes a disproportionate effect on hybrid male sterility and inviability (CHARLESWORTH *et al.* 1987; COYNE and ORR 1989), all else being equal, there should be less introgression for the X chromosome than for autosomes. The numbers of fixed differences between these species for the X chromosome and autosomes are higher and lower, respectively, than the expectations under the isolation model, but these differences are not statistically significant when tested by coalescent simulation. These differences, however, are in the direction expected if the X chromosome shows less introgression, and it is possible that in such closely related species the large variance of the coalescent process denies us the statistical power to demonstrate this.

Alternatively, there are two other explanations for a lack of difference in introgression of X-linked vs. autosomal genes. First, the large effect of the X chromosome may reflect not a difference in the number of genes causing sterility, but merely its expression of recessively acting sterility genes in males—the so-called “dominance theory,” for which there is substantial evidence (COYNE and ORR 2004). If the density of sterility genes is similar on all chromosomes, it is thus possible that “X effects” in males caused by recessivity alone (rather than a higher density of sterility genes) would not lead to a reduced rate of introgression. Theoretical work is needed to explore this possibility. However, recent experiments in *D. simulans* and *D. mauritiana* show that X-chromosomal sterility genes are not only partially recessive, but also more numerous than those on autosomes (TAO *et al.* 2003). If this were also the case in *D. yakuba* and *D. santomea*, one would indeed expect less introgression of X-linked vs. autosomal genes.

Alternatively, reduced introgression of the X chromosome caused by its larger effect on hybrid sterility may be balanced by the accumulation of autosomal genes contributing to other reproductive barriers, such as sexual or habitat isolation. Theoretical work by ORR and BETANCOURT (2001) (ORR and BETANCOURT 2001), for example, shows that when adaptation to a sudden change in environmental conditions is achieved through selection of existing genetic variation at initial mutation/selection equilibrium, genes in the X chromosome are expected to evolve *less* rapidly than those in autosomes.

We thank D. Charlesworth and two anonymous reviewers for comments and suggestions. We are also grateful to J. M. Comeron for helping us with different programs and to S. Elwyn and S. Powers for their assistance in the laboratory. We thank A. Chang for collecting flies, S. Sousa (Conservation et utilisation rationnelle des ÉCOsystèmes Forestiers D’Afrique Centrale, ECOFAC) for assistance in the field, and the Government of São Tomé and Príncipe for allowing us to work in the Obo National Park. This work was supported by National Institutes of Health grant GM58260 to J. A. Coyne.

#### LITERATURE CITED

- AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. *Genetics* **122**: 607–615.
- ARNOLD, M. L., 1997 *Natural Hybridization and Evolution*. Oxford University Press, Oxford.
- ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- BALLARD, J. W., 2004 Sequential evolution of a symbiont inferred from the host: Wolbachia and *Drosophila simulans*. *Mol. Biol. Evol.* **21**: 428–442.
- BALLARD, J. W. O., and M. C. WHITLOCK, 2004 The incomplete natural history of mitochondria. *Mol. Ecol.* **13**: 729–744.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- BERNATCHEZ, L., L. GLEMETE, C. C. WILSON and R. G. DANZMANN, 1995 Introgression and fixation of Arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Can. J. Fish. Aquat. Sci.* **52**: 179–185.
- BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**: 1111–1117.
- BROUGHTON, R. E., and R. G. HARRISON, 2003 Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* **163**: 1389–1401.
- CARIOU, M. L., J. F. SILVAIN, V. DAUBIN, J. L. D. LAGE and D. LACHAISE, 2001 Divergence between *Drosophila santomea* and allopatric or sympatric populations of *D. yakuba* using paralogous amylase genes and migration scenarios along the Cameroon volcanic line. *Mol. Ecol.* **10**: 649–660.
- CHANG, A. S., 2004 Conspecific sperm precedence in sister species of *Drosophila* with overlapping ranges. *Evolution* **58**: 781–789.
- CHARLESWORTH, B., J. A. COYNE and N. BARTON, 1987 The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**: 113–146.
- CLARK, A. G., 1997 Neutral behavior of shared polymorphism. *Proc. Natl. Acad. Sci. USA* **94**: 7730–7734.
- COMERON, J. M., 1999 *K-estimator*: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15**: 763–764.
- COYNE, J. A., 1992 Genetics and speciation. *Nature* **355**: 511–515.
- COYNE, J. A., and H. A. ORR, 1989 Patterns of speciation in *Drosophila*. *Evolution* **43**: 362–381.
- COYNE, J. A., and H. A. ORR, 1998 The evolutionary genetics of speciation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353**: 287–305.
- COYNE, J. A., and H. A. ORR, 2004 *Speciation*. Sinauer Associates, Sunderland, MA.
- COYNE, J. A., S. Y. KIM, A. S. CHANG, D. LACHAISE and S. ELWYN, 2002 Sexual isolation between sibling species with overlapping ranges: *Drosophila santomea* and *Drosophila yakuba*. *Evolution* **56**: 2424–2434.
- COYNE, J. A., S. ELWYN, S. Y. KIM and A. LLOPART, 2004 Genetic studies of two sister species in the *Drosophila melanogaster* subgroup, *D. yakuba* and *D. santomea*. *Genet. Res.* **84**: 11–26.
- DOBZHANSKY, T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York.
- DVORAK, J., M. C. LUO and Z. L. YANG, 1998 Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *agelops* species. *Genetics* **148**: 423–434.

- EXCOFFIER, L., P. E. SMOUSE and J. M. QUATTRO, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- FERRIS, C., J. M. RUBIO, L. SERRANO, J. GOSALVEZ and G. M. HEWITT, 1993 One way introgression of a subspecific sex chromosome marker in a hybrid zone. *Heredity* **71**: 119–129.
- 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 R. NIELSEN, 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747–760.
- 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., 2001 Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- HUDSON, R. R., and J. A. COYNE, 2002 Mathematical consequences of the genealogical species concept. *Evolution* **56**: 1557–1565.
- 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., D. D. BOOS and N. L. KAPLAN, 1992 A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- KIMURA, M., and T. OHTA, 1972 The age of a neutral mutant persisting in a finite population. *Genetics* **75**: 199–212.
- KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**: 1913–1931.
- KRAFT, T., T. SALL, I. MAGNUSON-RADING, N. O. NILSSON and C. HALLDEN, 1998 Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**: 1239–1244.
- LACHAISE, D., M. HARRY, M. SOLIGNAC, F. LEMEUNIER, V. BÉNASSI *et al.*, 2000 Evolutionary novelties in islands: *Drosophila santomea* a new *melanogaster* sister species from São Tomé. *Proc. R. Soc. Lond. Ser. B* **267**: 1487–1495.
- LANGLEY, C. H., J. MACDONALD, N. MIYASHITA and M. AGUADE, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **90**: 1800–1803.
- LEMEUNIER, F., and M. ASHBURNER, 1976 Relationship within the *melanogaster* subgroup of the genus *Drosophila* (Sophophora). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. *Proc. R. Soc. Lond. Ser. B* **193**: 275–294.
- LLOPART, A., S. ELWYN, D. LACHAISE and J. A. COYNE, 2002 Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* **56**: 2262–2277.
- MACHADO, C. A., and J. HEY, 2003 The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **270**: 1193–1202.
- MACHADO, C. A., R. M. KLIMAN, J. A. MARKERT and J. HEY, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19**: 472–488.
- MALLET, J., 1995 A species definition for the modern synthesis. *Trends Ecol. Evol.* **10**: 294–299.
- MAYR, E., 1942 *Systematics and the Origin of Species*. Columbia University Press, New York.
- MONNEROT, M., M. SOLIGNAC and D. R. WOLSTENHOLME, 1990 Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. *J. Mol. Evol.* **30**: 500–508.
- MORIYAMA, E. N., and J. R. POWELL, 1997 Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* **45**: 378–391.
- NACHMAN, M. W., V. L. BAUER, S. L. CROWELL and C. F. AQUADRO, 1998 DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**: 1133–1141.
- NAVARRO, A., and N. H. BARTON, 2003 Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**: 447–459.
- NEI, M., and M. W. FELDMAN, 1972 Identity of genes by descent within and between populations under mutation and migration pressures. *Theor. Popul. Biol.* **3**: 460–465.
- NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- NOOR, M. A. F., K. L. GRAMS, L. A. BERTUCCI and J. REILAND, 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98**: 12084–12088.
- ORR, H. A., and A. J. BETANCOURT, 2001 Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**: 875–884.
- PARSCH, J., 2003 Selective constraints on intron evolution in *Drosophila*. *Genetics* **165**: 1843–1851.
- PRZEWORSKI, M., R. R. HUDSON and A. DI RIENZO, 2000 Adjusting the focus on human variation. *Trends Genet.* **16**: 296–302.
- RAMOS-ONSINS, S. E., B. E. STRANGER, T. MITCHELL-OLDS and M. AGUADE, 2004 Multilocus analysis of variation and speciation in closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**: 373–388.
- RAND, D. M., and L. M. KANN, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**: 735–748.
- RAND, D. M., M. DORFSMAN and L. M. KANN, 1994 Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics* **138**: 741–756.
- RICE, W., 1989 Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- RIESEBERG, L. H., C. V. FOSSEN and A. M. DESROCHERS, 1995 Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375**: 313–316.
- ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 404–425.
- SHAW, K. L., 2002 Conflict between nuclear and mitochondrial DNA phylogenies of recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc. Natl. Acad. Sci. USA* **99**: 16122–16127.
- SMITH, G. R., 1992 Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. *Syst. Biol.* **41**: 41–57.
- STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermillion* and *forked* loci. *Genetics* **121**: 89–99.
- STEPHAN, W., L. XING, D. A. KIRBY and J. M. BRAVERMAN, 1998 A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl. Acad. Sci. USA* **95**: 5649–5654.
- STEPHENS, M., and P. DONNELLY, 2003 A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**: 1162–1169.
- STEPHENS, M., N. J. SMITH and P. DONNELLY, 2001 A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**: 978–989.
- TAJIMA, F., 1989a DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* **123**: 229–240.
- TAJIMA, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAKAHASHI, A., S. C. TSAUR, J. A. COYNE and C.-I. WU, 2001 The nucleotide changes governing phenomonal variation in *Drosophila* and their evolution. *Proc. Natl. Acad. Sci. USA* **98**: 3920–3925.
- TAKANO-SHIMIZU, T., 1999 Local recombination and mutation effects on molecular evolution in *Drosophila*. *Genetics* **153**: 1285–1296.
- TAO, Y., S. CHEN, D. L. HARTL and C. C. LAURIE, 2003 Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and

- D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* **164**: 1383–1397.
- TAYLOR, E. B., and J. D. MCPHAIL, 2000 Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proc. R. Soc. Lond. Ser. B* **267**: 2375.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- TING, C. T., S. C. TSAUR and C.-I. WU, 2000 The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97**: 5313–5316.
- TRUE, J. R., J. M. MERCER and C. C. LAURIE, 1996 Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**: 507–523.
- TUCKER, P. K., R. D. SAGE, J. WARNER, A. C. WILSON and E. M. EICHER, 1992 Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* **46**: 1146–1163.
- TURELLI, M., and A. A. HOFFMANN, 1995 Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics* **140**: 1319–1338.
- WAKELEY, J., and J. HEY, 1997 Estimating ancestral population parameters. *Genetics* **145**: 847–855.
- 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.
- WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. *Nature* **398**: 236–239.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WU, C.-I., 2001 The genic view of the process of speciation. *J. Evol. Biol.* **14**: 851–865.
- ZABALOU, S., S. CHARLAT, A. NIRGIANAKI, D. LACHAISE, H. MERCOT *et al.*, 2004 Natural Wolbachia infections in the *Drosophila yakuba* species complex do not induce cytoplasmic incompatibility but fully rescue the wRi modification. *Genetics* **167**: 827–834.
- ZUROVCOVA, M., and W. F. EANES, 1999 Lack of nucleotide polymorphism in the Ylinked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. *Genetics* **153**: 1709–1715.

Communicating editor: D. CHARLESWORTH