Patterns of DNA Sequence Variation Suggest the Recent Action of Positive Selection in the *janus-ocnus* Region of *Drosophila simulans*

John Parsch, Colin D. Meiklejohn and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 01238-2020

Manuscript received April 2, 2001
Accepted for publication July 16, 2001

ABSTRACT

Levels of nucleotide polymorphism in three paralogous *Drosophila simulans* genes, *janusA* (*janA*), *janusB* (*janB*), and *ocnus* (*ocn*), were surveyed by DNA sequencing. The three genes lie in tandem within a 2.5-kb region of chromosome arm 3R. In a sample of eight alleles from a worldwide distribution we found a significant departure from neutrality by several statistical tests. The most striking feature of this sample was that in a 1.7-kb region containing the *janA* and *janB* genes, 30 out of 31 segregating sites contained variants present only once in the sample, and 29 of these unique variants were found in the same allele. A restriction survey of an additional 28 lines of *D. simulans* revealed strong linkage disequilibrium over the *janA*-*janB* region and identified six more alleles matching the rare haplotype. Among the rare alleles, the level of DNA sequence variation was typical for *D. simulans* autosomal genes and showed no departure from neutrality. In addition, the rare haplotype was more similar to the *D. melanogaster* sequence, indicating that it was the ancestral form. These results suggest that the derived haplotype has risen to high worldwide frequency relatively recently, most likely as a result of natural selection.

THE comparison of DNA sequences both between and within species provides valuable information for understanding the evolutionary forces affecting genetic loci. Interspecific comparisons are useful for measuring rates of molecular evolution, inferring functional domains that are under strong selective constraint, and understanding processes such as gene duplication. Intraspecific comparisons reveal the standing genetic variation within a population and the microevolutionary forces that have shaped such variation. Previously we investigated the rates of molecular evolution of three paralogous male-reproductive genes, *janusA* (*janA*), *janusB* (*janB*), and *ocnus* (*ocn*), in the *Drosophila melanogaster* species subgroup (Parsch *et al.* 2001). The three genes are the result of two duplication events. The initial duplication of an ancestral sequence produced the *janA* and *janB* genes and clearly predates the divergence of the *D. melanogaster* and *D. obscura* species groups (Yanicostas *et al.* 1995), which is estimated to have occurred 25 million years ago (mya) (Russo *et al.* 1995). The subsequent duplication of *janB* to produce *ocn* appears to have occurred after the divergence of the *D. melanogaster* and *D. obscura* species groups but prior to divergence of the *D. melanogaster* species subgroup (Parsch *et al.* 2001), which occurred ~10 mya (Russo *et al.* 1995). Our results indicated significant heterogeneity in rates of evolution (as measured by the ratio of the nonsynonymous and synonymous substitution rates, d_N/d_S) among the three genes, suggesting that each gene has evolved under different selective constraints following duplication. In addition, all three genes showed a faster rate of evolution than genes encoding metabolic enzymes. This result was consistent with a general pattern of increased evolutionary rates in genes with reproductive function (Civetta and Singh 1998). Some reproductive genes, such as the Drosophila accessory protein gene *Acp26Aa*, show evidence for positive selection by a d_N/d_S ratio that is significantly greater than 1 (Tsaur and Wu 1997). This ratio was much less than 1 for *janA*, *janB*, and *ocn*, and thus there was no evidence for positive selection by this strict criterion. However, a number of powerful statistical techniques have been developed to detect patterns of selection from intraspecific DNA polymorphism data. For this reason, we chose to investigate intraspecific variation in the *janA*, *janB*, and *ocn* genes.

In species of the *D. melanogaster* species subgroup, *janA*, *janB*, and *ocn* lie in tandem within a 2.5-kb region of the right arm of chromosome 3 (Figure 1). *janA* produces two alternatively spliced transcripts, one that is specific to testes and another that is found in various tissues and in both sexes (Yanicostas *et al.* 1989). The two *janA* transcripts differ in their 5′ untranslated regions (UTRs) and their translation begins at different AUG initiation codons, with initiation of the sperm-specific polypeptide occurring 48 bp downstream of the general initiation site (Yanicostas *et al.* 1989). The 3′ UTR of *janA* overlaps with the 5′ UTR of *janB* and the beginning of the *janB* protein-encoding region (Yani-

Corresponding author: John Parsch, Department of Organismic and Evolutionary Biology, Harvard University Biological Laboratories, 16 Divinity Ave., Cambridge, MA 02138-2020.
E-mail: jparsch@oeb.harvard.edu

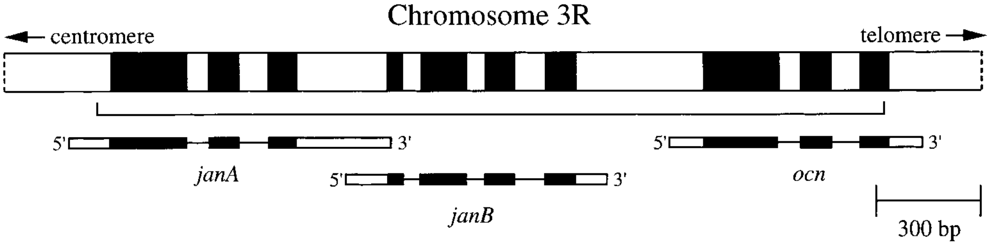

Figure 1.—Diagram of the *janus-ocnus* genomic region. The relative chromosomal location of the three genes is shown at the top, with coding regions represented as solid boxes. A bracket indicates the region that was PCR amplified and sequenced in this study. The *janA*, *janB*, and *ocn* transcriptional units are shown below. Solid boxes represent coding regions, open boxes represent untranslated regions, and lines represent introns.

costas *et al.* 1989; Figure 1). Despite this overlap, both *janA* and *janB* produce monocistronic transcripts that are controlled by independent promoters (Yanicostas and Lepesant 1990). *janB* and *ocn* produce only testis-specific transcripts (Yanicostas *et al.* 1989; Parsch *et al.* 2001). The *ocn* transcriptional unit lies ~250 bp downstream from the *janB* polyadenylation site, and there is no overlap between the *janB* and *ocn* transcripts (Figure 1). The *janB* 5′ UTR contains translational control elements that have been shown to restrict translation to the postmeiotic stages of sperm development (Yanicostas *et al.* 1995). The high degree of similarity between the *janB* and *ocn* 5′ UTR sequences suggests that *ocn* translation is under similar post-transcriptional regulation (Parsch *et al.* 2001).

In this article, we report DNA sequence variation in the *janA-ocn* region of *D. simulans*. We find reduced levels of polymorphism at these loci relative to other third chromosome loci surveyed previously. More strikingly, we find two highly divergent haplotypes in a 1.7-kb region spanning the *janA* and *janB* genes. The distribution of variants at segregating sites is shown to differ significantly from the neutral expectation by several statistical tests. These results are consistent with a model of genetic hitchhiking and suggest the recent action of positive selection in this region of the genome.

MATERIALS AND METHODS

Fly stocks: Each *D. simulans* line was derived from a single wild-caught female and maintained by brother/sister mating for >50 generations. The lines were collected from various geographic locations and at various times and were kindly provided to us by P. Capy and Y. Tao. The Canton-S strain of *D. melanogaster* was used as an outgroup. Genomic DNA was prepared from a single male of each line as described previously (Parsch *et al.* 2001).

PCR and DNA sequencing: The *janA-ocn* region was PCR amplified as a single 2.4-kb fragment from genomic DNA using primers and amplification conditions described in Parsch *et al.* (2001). The amplified region contained the complete coding sequences of *janA* and *janB* and a large portion of the *ocn* coding sequence extending into exon 3 (Figure 1). PCR products were cloned following the protocol of the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Plasmid DNA was then purified by the alkaline lysis procedure (Sambrook *et al.* 1989) and used as a template for DNA sequencing. Alternatively, PCR products were purified using the QIAquick PCR purification kit (QIAGEN, Valencia, CA) and used directly as sequencing templates. DNA sequencing was performed with the dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA), using the amplification primers and gene-specific internal primers (Parsch *et al.* 2001). In addition, universal M13 forward and reverse primers were used for sequencing plasmid templates. Sequencing gels were run on an ABI 373 automated sequencer. DNA was sequenced on both strands and a minimum of either two independently cloned plasmid templates or one plasmid template and one PCR template were sequenced for each line. Additional clones or PCR templates were sequenced when necessary to resolve ambiguities. We did not encounter any heterozygous positions within the sequenced regions. DNA sequences have been submitted to the GenBank database under accession nos. AF393330–AF393368.

Restriction analysis: Restriction enzymes and buffers were supplied by New England Biolabs (Beverly, MA). The following enzymes were used: *Bst*YI, *Rsa*I, *Mlu*I, and *Fok*I. For restriction analysis, the *janA*, *janB*, and *ocn* genes were amplified separately from each *D. simulans* line and an aliquot of the undigested product was run on a 1% agarose gel to ensure correct amplification. Five microliters of PCR product was then digested using the manufacturer's buffer and 2–6 units of enzyme. Digests were carried out at 37° for 2 hr. Digestion products were separated on 2% agarose gels, which allowed the unambiguous scoring of the presence or absence of a particular restriction site. Separate digests were performed for each restriction enzyme.

Sequence analysis: Standard DNA polymorphism analyses and coalescent simulations to determine the probabilities of the observed number of haplotypes, haplotype diversity, and Tajima's (1989) *D* statistic were performed using DnaSP 3.50 (Rozas and Rozas 1999). Coalescent simulations to determine the statistical significance of Fay and Wu's (2000) *H* statistic were performed using a program provided by J. Fay. The haplotype test of Hudson *et al.* (1994), which determines the probability of observing a subset of *i* alleles with *j* or fewer segregating sites given a total sample of *n* alleles with *S* segregating sites, was performed using a program provided by J. Braverman. Values of *i* and *j* were chosen to produce the most extreme subset possible from the observed data. The probability was corrected for the *a posteriori* choice of *i* and *j* by including the probability of all more extreme configurations theoretically possible. For all of the above, 10,000 random coalescent simulations (Hudson 1990) were performed under the conservative assumption of no recombination. The number of segregating sites was fixed at the observed value.

```
                                           1 1 1 1 1 1 1 1 1 1 1 1  1 1 2 2 2 2 2 2 2 2 2 2 2 2 2
          2 2 2 3 5 5 5 6 6 6 6 6 6 6 7  8 1 1 2 3 3 3 4 6 6 6 6 6  8 9 0 0 0 0 0 1 1 1 1 1 1 1 1
    3 8 9 4 9 9 6 3 4 8 4 5 5 5 5 6 6 0  4 3 3 9 0 2 8 5 1 7 8 8 9  3 6 0 0 1 2 2 5 6 7 7 7 8 9 9
    5 3 3 6 1 4 2 8 9 5 4 1 2 7 8 5 7 7  3 2 9 1 0 2 8 2 4 3 1 2 4  6 8 1 6 4 0 8 7 4 1 6 9 8 4 5
    -----------------------------------  -------------------------  -----------------------------
s1  c g a t c c a a t a t a a a g c t c  g a t a a g c c g a t t c  a c g t c t g a t a a g c g -
s2  . . c . . . . . . . . . . . . . . .  . . . . . . . g . . . . .  . . . . . . . . . . . . . . .
s3  a c . c a t g c c c g g g g a t c t  a t c c t c t g t t g c a  . . . . . . . . . . . . . . .
s4  . . . . . . . . . . . . . . . . . .  . . . . . . . . . . . . .  . . . . . . . . . . . . . . .
s5  . . . . . . . . . . . . . . . . . .  . . . . . . . g . . . . .  g . c . t g a g g g c c - t t
s6  . . . . . . . . . . . . . . . . . .  . . . . . . . . . . . . .  . . . . . . . . . . . . . . .
s7  . . . . . . . . . . . . . . . . . .  . . . . . . . g . . . . .  g . c . t g a g g g c c - t t
s8  . . . . . . . . . . . . . . . . . .  . . . . . . . g . . . . .  . t . c . g . g g g c c . . .

m1  . . . c a t g c c c a g . . . . . t  . . c c . . t g . t g c a  g . . . . g . g g g c c - t t
    -----------------------------------  -------------------------  -----------------------------
                  janA                              janB                         ocn
```

FIGURE 2.—Segregating sites in the original sample of eight *D. simulans* lines (s1–s8) from a worldwide distribution. m1 represents the *D. melanogaster* outgroup sequence. Dots indicate a match to the s1 sequence. Gaps are indicated by dashes. The dash at position 2188 represents a 3-bp deletion beginning at 2188.

RESULTS

Nucleotide polymorphism in the *janA-ocn* region: We initially surveyed levels of nucleotide polymorphism in the *janA-ocn* region of eight *D. simulans* lines from a worldwide distribution. This survey revealed a total of 44 single nucleotide polymorphisms (SNPs) and two insertion/deletion (indel) polymorphisms (Figure 2). All of the polymorphisms were at silent or noncoding sites, with the exception of a C/A polymorphism at position 93. Interestingly, this polymorphism changes the sperm-specific downstream initiation codon of *janA* from AUG (Met) to CUG (Leu). This change presumably eliminates the sperm-specific form of the *janA* polypeptide in line s2. However, s2 males appear phenotypically normal and are completely fertile (our unpublished results).

The most striking feature of the data is the distribution of variants at segregating sites over the *janA-janB* region. In this region there are a total of 31 segregating sites. Variants at 30 of these sites are unique within the sample (singletons), and 29 of the singletons are found in line s3. The *ocn* gene also shows an unusual distribution of variation. Five of the eight alleles are identical and match the s1 sequence. Of the three remaining alleles, two are identical to each other (s5 and s7) but differ from s1 at 11 sites. The final allele, s8, differs from s1 at 8 sites but also differs from s5 and s7 at 7 sites. Interestingly, the rare variants in *ocn* occur in different lines than the rare variants in *janA* and *janB*. Because many of these rare variants match the *D. melanogaster* sequence, we can discount the possibility that they are all new mutations and can infer a recombination event within the intergenic region between *janB* and *ocn* (Figure 2).

Several statistical tests were applied to the data to determine whether the observed distribution of variants at segregating sites differed from the neutral expectation (Table 1). Only SNPs were considered in the calculation of statistics presented in Table 1 and in the analyses below. All tests were applied separately to each gene and to the region as a whole. In addition, due to the strong linkage disequilibrium between variants in the *janA* and *janB* genes, we applied the tests to the combined *janA-janB* region. First, we tested for a departure

TABLE 1

Summary statistics for original sample (eight alleles)

Gene	sites	S^a	π^a	θ^a	D	Nhap	Hdiv	Sub(i, j)	H
janA	720	18	0.0063	0.0096	−1.82*	3*	0.46**	(7, 1)**	−11.57**
	(411)	(17)	(0.0103)	(0.0160)					
janB	973	13	0.0036	0.0052	−1.49*	3	0.68	(7, 1)**	−9.21**
	(649)	(13)	(0.0054)	(0.0077)					
janA+B	1693	31	0.0048	0.0071	−1.74*	4*	0.79	(7, 2)**	−20.78**
	(1060)	(30)	(0.0073)	(0.0109)					
ocn	569	13	0.0103	0.0088	0.85	3	0.61*	(5, 0)	−2.57
	(238)	(13)	(0.0246)	(0.0210)					
All	2262	44	0.0061	0.0075	−0.97	5	0.86*	(4, 2)	−23.36**
	(1298)	(43)	(0.0105)	(0.0128)					

S, number of segregating sites; π, average number of pairwise nucleotide differences (NEI 1987); θ, WATTERSON's (1975) estimator of 4*N*μ; *D*, TAJIMA's (1989) *D* statistic; Nhap, number of haplotypes; Hdiv, haplotype diversity (NEI 1987); Sub (*i*, *j*), the most extreme subset of the sample, where *i* and *j* are the number of alleles and number of segregating sites in the subsample, respectively (HUDSON *et al.* 1994); *H*, FAY and WU's (2000) *H* statistic. Significance levels were determined by 10,000 random coalescent simulations on the basis of the number of alleles and the observed number of segregating sites. *$P < 0.05$; **$P < 0.01$.

[a] Values in parantheses are for silent + noncoding sites only.

TABLE 2

Restriction survey of a worldwide sample of *D. simulans*

Line	Origin	*Bst*YI (246)	*Rsa*I (1291)	*Rsa*I (1452)	*Mlu*I (1836)	*Fok*I (2164)	Hap[a]
2	Japan	−	+	+	+	−	1
5	Seychelles	−	+	+	−	+	1
7	Peru	−	+	+	−	+	1
8	Kenya	−	+	+	+	+	1
9	France	−	+	+	+	−	1
10	France	−	+	+	−	+	1
11	France	−	+	+	−	+	1
13	Tunisia	−	+	+	+	−	1
14	Tunisia	−	+	+	+	−	1
15	Tunisia	−	+	+	+	−	1
16	Tunisia	−	+	+	+	−	1
18	Congo	−	+	+	−	+	1
21	South Africa	−	+	+	+	−	1
22	South Africa	−	+	+	−	+	1
23	South Africa	−	+	+	+	−	1
26	Australia	−	+	+	+	−	1
29	Australia	−	+	+	−	+	1
30	Japan	−	+	+	−	+	1
27	St. Martin	−	+	+	+	−	1
1	United States	−	+	−	+	−	1
4	Haiti	−	+	−	+	−	1
6	South Africa	−	+	−	+	−	1
12	France	−	+	−	+	−	1
20	South Africa	−	+	−	+	−	1
28	Australia	−	+	−	+	−	1
32	Japan	−	+	−	+	−	1
33	Japan	−	+	−	+	−	1
3	South Africa	+	−	+	+	−	2
17	Congo	+	−	+	−	−	2
19	Congo	+	−	+	−	−	2
25	Australia	+	−	+	−	−	2
31	Japan	+	−	+	+	−	2
34	Polynesia	+	−	+	−	−	2
36	Zimbabwe	+	−	+	−	−	2
24	Australia	+	+	+	−	−	2/1
35	Seychelles	+	+	−	+	−	2/1

Restriction sites are given at the top, with the coordinates in parentheses. +, cut; −, uncut.

[a] Haplotype is defined by restriction pattern over *janA-janB* region (sites 247 and 1242). Putative recombinants are labeled as both types (*i.e.,* 2/1).

from neutrality in the frequency distribution of variants at segregating sites using Tajima's (1989) *D* statistic. We obtained a significantly negative value of *D* for *janA*, *janB*, and the combined *janA-janB* region (Table 1). This indicates an excess of low-frequency variants and can be explained by the large number of sites at which s3 differs from the other alleles. Tajima's *D* is positive, although not significantly so, for *ocn,* where many of the variants are in intermediate frequency.

Two haplotype tests implemented in the DnaSP computer program (Rozas and Rozas 1999) were applied to our data. These test for either a reduction in the number of haplotypes or in haplotype diversity by comparing the observed data to the results of random coalescent simulations and are similar to the tests proposed by Depaulis and Veuille (1998). Our results indicate a significant paucity of haplotypes at *janA* and in the combined *janA-janB* region (Table 1). We also find a significant reduction in haplotype diversity at both *janA* and *ocn* and in the region as a whole (Table 1). In addition, the haplotype test of Hudson *et al.* (1994) revealed a highly significant departure from the neutral expectation for *janA*, *janB*, and the combined *janA-janB* region (Table 1). This indicates that there are fewer segregating sites within the common haplotype than would be expected under a neutral equilibrium model.

Finally, we compared the frequency spectrum of derived variants at polymorphic sites to the neutral expectation using the *H* statistic of Fay and Wu (2000). A significantly negative value of *H* indicates that derived variants are in higher frequency than expected under a neutral mutation-drift model. Since derived mutations

TABLE 3

Linkage disequilibrium between polymorphic restriction sites

	*Bst*YI(246)	*Rsa*I(1291)	*Rsa*I(1452)	*Mlu*I(1836)	*Fok*I(2164)
*Bst*YI(246)	—	4.3×10^{-6}	0.40	0.11	0.08
*Rsa*I(1291)	1.00	—	0.16	0.08	0.16
*Rsa*I(1452)	0.56	1.00	—	0.01	0.08
*Mlu*I(1836)	0.45	0.53	1.00	—	7.2×10^{-4}
*Fok*I(2164)	1.00	1.00	1.00	0.82	—

P values (Fisher's exact test) are shown above the diagonal. The ratio of the observed linkage to its theoretical maximum (given the observed allele frequencies) is shown below the diagonal.

are expected to increase in frequency when linked to a positively selected site, this can be used as a test for genetic hitchhiking (Fay and Wu 2000). Our results indicate that *H* is significantly negative for *janA* and *janB* and also for the combined *janA-janB* region and the entire *janA-ocn* region (Table 1). Although *H* is not significantly negative for the *ocn* gene by itself, the common *ocn* haplotype does show the derived state at 8 of 13 SNP sites and at both of the indel sites (Figure 2). These results suggest that the haplotype structure observed in this region may be explained by previously rare variants being driven to high frequency due to their linkage with a positively selected site.

Restriction survey: To further investigate the haplotype structure in this region of the genome, we surveyed restriction site polymorphism in an additional 28 lines collected from a worldwide distribution. On the basis of our initial sequencing, we chose five polymorphisms that resulted in either the gain or loss of a restriction site. The first two polymorphisms [*Bst*YI(246) and *Rsa*I(1291)] span the *janA-janB* region and distinguish the s3 allele from all others. Furthermore, these were sites at which the s3 allele matched the *D. melanogaster* outgroup sequence, suggesting that the rare variant represents the ancestral state. We also surveyed an *Rsa*I polymorphism at site 1452. This is the only site over the *janA-janB* region that is not a singleton and represents a derived polymorphism segregating within the common haplotype (Figure 2). The final two restriction site polymorphisms [*Mlu*I(1836) and *Fok*I(2164)] span the *ocn* gene. The former polymorphism distinguishes the s5 and s7 alleles from all others, while the latter distinguishes s5, s7, and s8 from all others. In both cases, the less frequent variant matches the *D. melanogaster* sequence. A summary of restriction site polymorphism is shown in Table 2. Over the *janA-janB* region, we found 27 alleles matching the common haplotype and 7 alleles matching the rare haplotype (*i.e.*, a restriction pattern identical to s3). Two alleles (s24 and s35) appear to be recombinants. Overall, the strongest linkage disequilibrium was between site 246 in *janA* and site 1291 in *janB* (Table 3). The restriction pattern at these two sites was used as a diagnostic to classify the alleles as either haplotype 1 or 2 (Table 2).

Nucleotide polymorphism in the rare haplotype: The six additional alleles matching the restriction pattern of haplotype 2 were completely sequenced over the *janA-ocn* region and compared to the original s3 sequence. This revealed a total of 71 SNPs and 3 indel polymorphisms (Figure 3). All of the polymorphisms occurred at silent or noncoding sites. Levels of nucleotide polymorphism within haplotype 2 were typical of *D. simulans* autosomal loci (Moriyama and Powell 1996; Table 4), and we found no significant departure from neutrality by any of the statistical tests described above when applied to each gene separately, the entire region, or to the combined *janA-janB* region (data not shown). The amount of polymorphism in the *janA* and *janB* genes within haplotype 2 is nearly 30-fold greater than that within haplotype 1 (Table 4), suggesting that haplotype 2 is ancestral. The ancestral state of haplotype 2 is also supported by the neighbor-joining tree shown in Figure 4, where the haplotype 1 alleles form a single clade within the haplotype 2 alleles. The separate evolutionary histories of these alleles over the *janA-janB* and *ocn* genes further indicate recombination between the *janB* and *ocn* genes. For example, the s3 allele differs substantially from the common haplotype over *janA* and *janB* but is identical to the most common haplotype at *ocn* (Figure 4).

DISCUSSION

Our survey of nucleotide polymorphism in the *janA-ocn* region of *D. simulans* reveals two noteworthy features. First, we detect strong linkage disequilibrium between variants in the *janA* and *janB* genes, which results from the segregation of variation in two divergent haplotypes. There are 13 fixed differences between the two haplotypes. The common haplotype, designated here as haplotype 1, has an estimated frequency of 75% and is present in worldwide populations. The rare haplotype, designated as haplotype 2, is present in several geographically distinct regions but appears to be more frequent in populations from central/southern Africa and the Pacific rim (Table 2). A recent report of nucleotide variation in the *janus* region of three *D. simulans* lines from Kenya (Kliman *et al.* 2000) is consistent with our

```
           1 2 2 2 3 3 3 3 3 4 5 5 5 5 5 6 6 6 6 6 6 6 6 7
     3 8 9 4 4 6 6 0 0 3 8 9 5 0 4 7 8 9 4 4 5 5 5 6 6 6 0
     5 3 8 0 2 0 6 0 1 2 6 2 9 0 9 0 8 1 4 6 2 7 8 5 7 9 7
     -----------------------------------------------------
s3   a c c a g a - g g c t t g c c g g c g t g g a t c t t
s17  c g . t a . i . . . c . . . . . c a t c a a g c t . c
s19  c g . . . c . a t a . . . . . . . . t . . a a g c t a .
s25  c g t . . . . . . . . . a t t a c . . . . . . . . . c
s31  c g . . . . . . . . . . . . . . . . . . . . . . . . c
s34  c g . . . c . a t a . c a t t a c . . . . . . . . . c
s36  c g . . . c . a t a . . a t t a c . . . . . . . . . c

m1   c g . . . . . . . . . . . . . . . . a . a a g c t a .
     -----------------------------------------------------
                               janA

             1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
     7 8 8 9 0 1 1 1 1 1 1 1 1 1 2 2 3 3 3 4 4 4 4 4 5 5 6 6 6 6 6 7   7 8 8 9 9 0 0 0 0 1 1 1 1 1 2
     7 4 9 0 9 0 2 3 3 5 5 6 6 8 8 8 0 2 8 4 6 6 7 9 1 6 1 3 5 6 8 1   7 3 3 2 9 1 2 2 9 3 4 5 7 8 5
     1 3 2 9 7 6 5 0 2 3 6 2 5 9 2 4 5 2 8 3 5 9 9 9 6 8 4 3 5 1 2 5   9 3 6 6 4 1 0 8 1 3 5 7 6 8 1
     ---------------------------------------------------------------   -----------------------------
s3   a a t t c g t c t t c g t c a c c c t g a g g - g - t t c t c t   g t a t c g t g c a c a a c c
s17  . g . c . a . . a . . . . . . . . g c . . . . . . i g . . a t c   . . g . . . g a . t . . c t .
s19  . g c . . . . . a . . . c . t . a g . . . . . i t . g . . . t .   . c g . . a g a t t . g c t .
s25  t g . . g . . . a c t c . a . . . g c . . . a . . . g . . . . .   a . g a . . g . . t . . c t t
s31  t . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .   . . . . . . . . . . . . . . .
s34  t g . . . . . . a . . . . . . . . g c . . . . . . i g . . a t .   . . g . a . g . . . . . c t .
s36  . g . . . . g t a . . . . . t t g g c a t c . . . . g c t . t .   . c g . . a g a . . t . c t .

m1   . g . . . . . . a c . . . . . . . g . . g . . . t . g . . . . .   . . g . . . g . . . . g c - .
     ---------------------------------------------------------------   -----------------------------
                                   janB                                             ocn
```

Figure 3.—Segregating sites in seven *D. simulans* lines matching the restriction pattern of haplotype 2. m1 represents the *D. melanogaster* outgroup sequence. Dots indicate a match to the s3 sequence. Gaps are indicated by dashes. Insertions are indicated by "i" and represent sequences of 23, 2, and 7 bp beginning at sites 266, 1499, and 1568, respectively.

results, as all three matched haplotype 2 at the diagnostic sites. The second noteworthy feature of our data is that levels of nucleotide polymorphism among alleles of the common haplotype are greatly reduced in comparison to alleles of the rare haplotype. The latter show levels of polymorphism typical for *D. simulans* autosomal genes, while the former show a nearly 30-fold reduction in polymorphism. Previous surveys of DNA sequence variation in *D. melanogaster* have revealed nonneutral haplotype structures at a number of loci, including *Sod* (Hudson *et al.* 1994), *white* (Kirby and Stephan 1995, 1996), *Suppressor of Hairless* (Depaulis *et al.* 1999), *Fbp2* (Bénassi *et al.* 1999), and the region spanning the proximal breakpoint of the chromosomal inversion *In(2L)t* (Andolfatto *et al.* 1999). In *D. simulans*, nonneutral haplotype structures have been reported for *Pgd* (Begun and Aquadro 1994), *runt* (Labate *et al.* 1999), *G6pd* and *vermilion* (Hamblin and Veuille 1999), and the

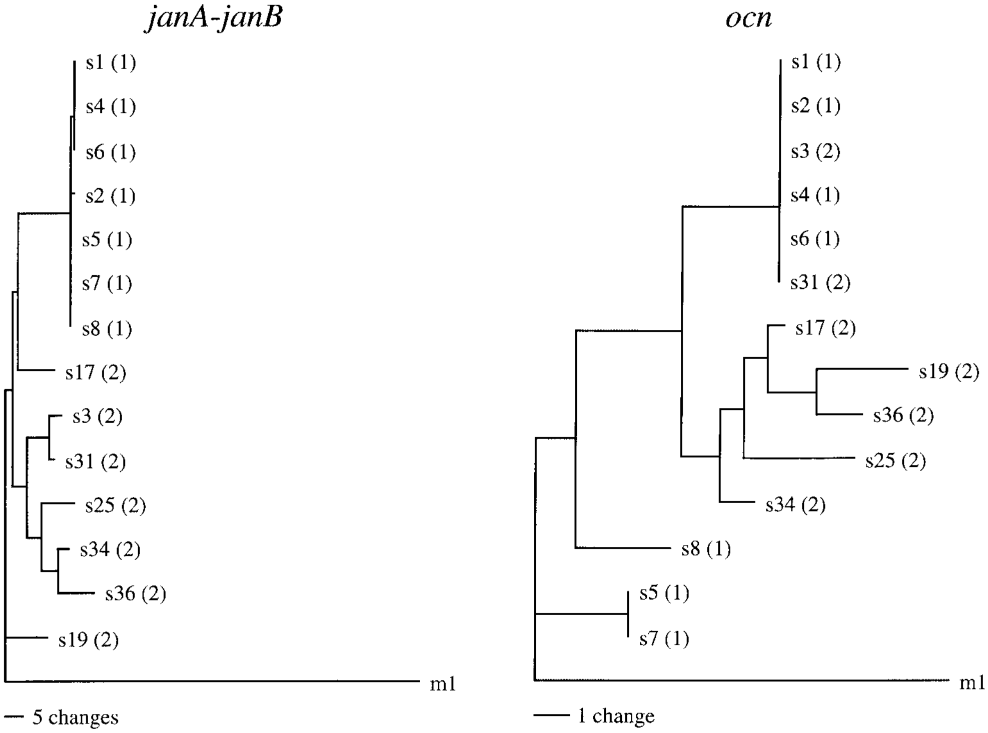

Figure 4.—Neighbor-joining trees of the 14 completely sequenced *D. simulans* alleles (Figures 2 and 3). The *D. melanogaster* sequence was used as an outgroup. Separate trees were constructed for the combined *janA-janB* genes and for the *ocn* gene. Numbers in parentheses indicate the haplotype class of each allele according to restriction pattern (Table 2).

TABLE 4

Summary statistics for haplotypes 1 and 2 (seven alleles each)

Haplotype	Gene	S^a	π^a	θ^a	Div^a	Nhap	Hdiv	R_m
1	*janA*	1	0.0004	0.0006	0.052	2	0.29	0
		(0)	(0.0000)	(0.0000)	(0.085)			
1	*janB*	1	0.0006	0.0004	0.104	2	0.57	0
		(1)	(0.0009)	(0.0006)	(0.152)			
1	*janA* + *B*	2	0.0005	0.0005	0.081	3	0.71	0
		(1)	(0.0006)	(0.0004)	(0.124)			
1	*ocn*	13	0.0112	0.0093	0.032	3	0.81	0
		(13)	(0.0268)	(0.0223)	(0.070)			
1	All	15	0.0032	0.0027	0.068	4	0.67	0
		(14)	(0.0054)	(0.0044)	(0.114)			
2	*janA*	26	0.0158	0.0153	0.053	7	1.00	3
		(26)	(0.0277)	(0.0258)	(0.087)			
2	*janB*	30	0.0111	0.0130	0.103	7	1.00	3
		(30)	(0.0166)	(0.0189)	(0.150)			
2	*janA* + *B*	56	0.0131	0.0140	0.081	7	1.00	6
		(56)	(0.0209)	(0.0216)	(0.124)			
2	*ocn*	15	0.0105	0.0107	0.036	6	0.95	1
		(15)	(0.0251)	(0.0257)	(0.080)			
2	All	71	0.0125	0.0132	0.069	7	1.00	7
		(71)	(0.0217)	(0.0223)	(0.116)			

Symbols and abbreviations are the same as in Table 1; Div, average pairwise divergence from *D. melanogaster*; R_m, the minimum number of recombination events (Hudson and Kaplan 1985) inferred from the *D. simulans* polymorphism data.

[a] Values in parentheses are for silent + noncoding sites only.

In(2L)t breakpoint (Andolfatto and Kreitman 2000). In addition, Andolfatto and Przeworski (2000) report that for an unexpectedly high number of *D. melanogaster* and *D. simulans* loci that have been surveyed, there is a discordance between the recombination rate inferred from population surveys and that from experimental mapping, with the experimental rates being higher. This can be interpreted as a genome-wide excess of linkage disequilibrium, although note that this criterion for detecting linkage disequilibrium is not as strict as many of the haplotype tests that assumed no recombination. Below we consider some genetic and evolutionary forces that may affect the distribution of variants at segregating sites in a population and discuss whether or not they can explain the patterns observed in the *janA*-*ocn* region.

One potential cause of strong linkage disequilibrium over an extended region of DNA sequence is a severe reduction in the rate of recombination. For example, recombination rates are known to be reduced in regions containing chromosomal inversions. To eliminate this possibility, we performed *in situ* hybridizations of a *janA*-*ocn* probe to polytene chromosomes of several representative lines from each haplotype to ensure that the genes were at the same cytological location in both haplotypes. In addition, we examined Giemsa-stained polytene chromosomes of the remaining lines of each haplotype and saw no evidence for inversion polymorphism in this region of the genome. Recombination rates are also known to be reduced in certain chromosomal regions, particularly those flanking centromeres and telomeres (Lindsley and Sandler 1977). Since the *janA*-*ocn* region is near the tip of chromosome arm 3R, this is a possibility. However, several observations argue against reduced recombination being the cause of the observed haplotype structure. First, a comparison of the physical and genetic distances for loci on the third chromosome (Hamblin and Aquadro 1996) indicates that there is not a great reduction in recombination in this region in *D. simulans*. Second, a worldwide sample of nine alleles of the *Tpi* locus (Hasson *et al.* 1998), which is located distal to *ocn* on chromosome 3R, shows no departure from neutrality by the haplotype number, haplotype diversity, or Hudson's haplotype test as implemented in this article. Third, there is ample evidence for recombination within the *janA*-*ocn* region in haplotype 2 (Figure 3). From the distribution of segregating sites in the seven alleles of haplotype 2 we infer a minimum of seven recombination events by the method of Hudson and Kaplan (1985), six of which fall within the *janA*-*janB* region (Table 4). Finally, the probabilities associated with all of the statistical tests in Table 1 were estimated from coalescent simulations under the conservative assumption of no recombination. Thus, even in the extreme case of zero recombination, the observed data differ significantly from the neutral expectation.

The presence of two divergent haplotypes could also be explained by demographic factors, such as popula-

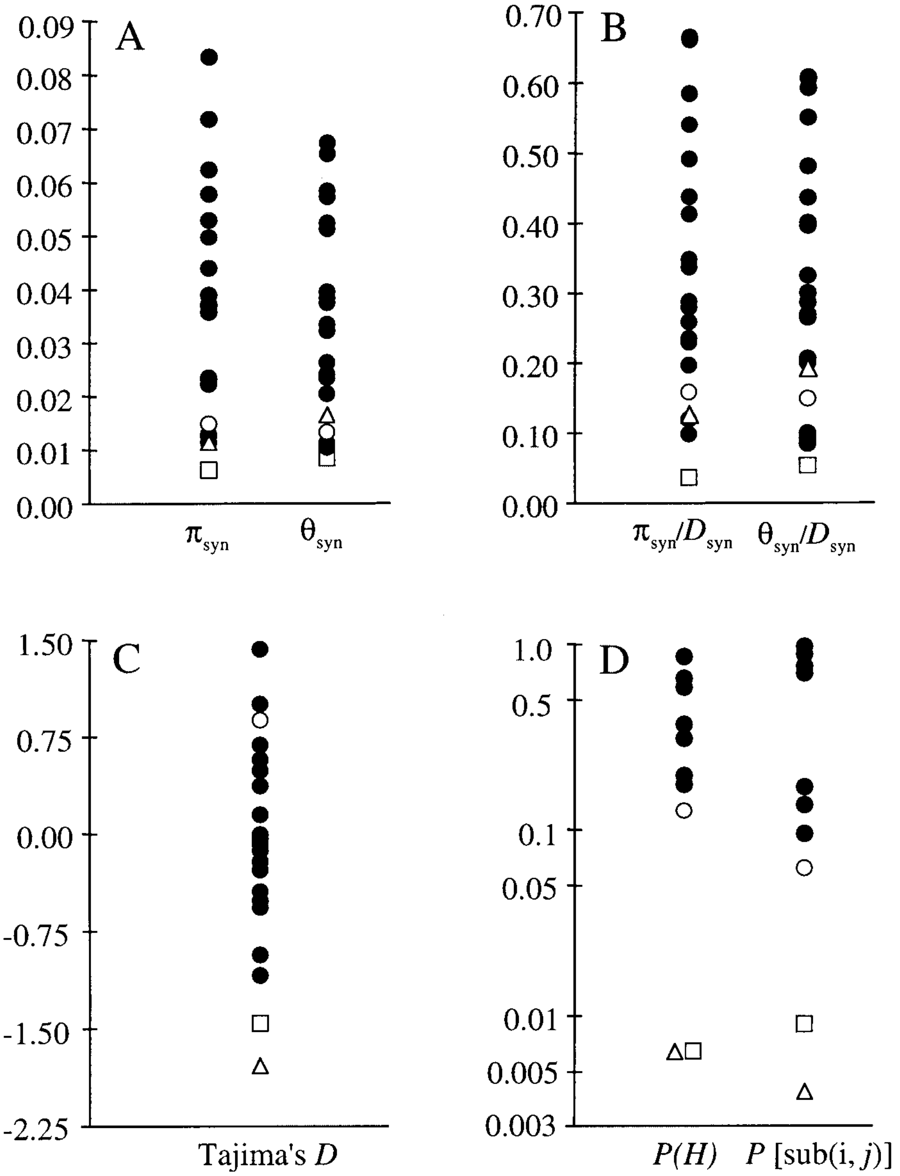

Figure 5.—Comparison of DNA polymorphism at *janA* (open triangles), *janB* (open squares), and *ocn* (open circles) with 19 other *D. simulans* loci (solid circles) spanning chromosome arm 3R. (A) Two measures of nucleotide polymorphism (π_{syn} and θ_{syn}) at synonymous sites. (B) Synonymous nucleotide polymorphism divided by interspecific divergence, where D_{syn} is the average pairwise divergence at synonymous sites of the original eight *D. simulans* alleles from the outgroup *D. melanogaster* sequence. (C) Tajima's (1989) D statistic. (D) Probability of H (Fay and Wu 2000) and sub(i, j) (Hudson *et al.* 1994) shown on a log scale. In D, only loci with a sample size ≥8 are shown.

tion subdivision, changes in population size, or founder effects. A survey of nucleotide polymorphism in the *vermilion* and *G6pd* genes detected significant population subdivision within *D. simulans*, particularly among African populations (Hamblin and Veuille 1999). Population subdivision is also indicated by the presence of distinct mitochondrial races within the species (Baba-Aissa *et al.* 1988). Although these results indicate that *D. simulans* likely departs from the standard demographic assumption of panmixis, it is unlikely that demographics alone can explain the pattern of nucleotide polymorphism in the *janA*-*ocn* region. This is because demographic forces are expected to affect the entire genome, not just particular loci, and when compared with previously sequenced loci the *janA*-*ocn* region appears to be unusual (see below). Furthermore, a demographic explanation must account for both features of the observed data (two divergent haplotypes and very low polymorphism within the common haplotype). A comparison of DNA polymorphism at *janA*, *janB*, and *ocn* with other *D. simulans* third chromosome loci studied by Begun and Whitley (2000) is relevant to these two points. Figure 5 shows the distribution of several statistics for 19 loci on chromosome arm 3R (see Table 1 of Begun and Whitley 2000) as well as for *janA*, *janB*, and *ocn*. All three of these genes show low levels of silent DNA polymorphism (Figure 5A). For example, the *janB* gene has lower values of π_{syn} and θ_{syn} than any of the other genes. *janA* and *ocn* are also among the lowest for these values (Figure 5A). The low variability observed in these genes cannot be caused by reduced mutation rates in this region of the genome or by unusually strong selective constraints, as measures of nucleotide variability are low in these genes even when standardized by interspecific divergence (Figure 5B). Differences in sampling schemes between our data and those of Begun and

Whitley (2000) strengthen support for the conclusion that the *janA-ocn* region is unusual among chromosome 3R loci. Most of the genes in Begun and Whitley's study were sampled from a single California population (although two notable exceptions are discussed below), whereas our data are from a worldwide sampling of *D. simulans.* If there is population subdivision, one would expect that increasing the geographical scope of a sample would increase, not decrease, the observed level of nucleotide polymorphism. In this respect, comparison of our data to Begun and Whitley's is conservative. In addition, none of the other third chromosome loci show a significant departure from neutrality by any of the statistical tests used in this study. This is illustrated by the distributions of Tajima's D (Figure 5C), the probability of Fay and Wu's H, and the probability of Hudson's haplotype test (Figure 5D). For the latter two quantities, only loci with a sample size ≥ 8 are shown because the power to detect significant results with these tests increases with sample size. However, it should be noted that none of the loci with sample size <8 showed a significant departure from neutrality. Of course, if there is population subdivision then the comparisons in Figure 5, C and D, could be misleading due to the different sampling schemes used for the different loci. However, if one invokes population subdivision to explain the divergent haplotypes it becomes difficult to explain the low levels of polymorphism within haplotype 1. The most obvious way to divide the sample under a population subdivision model would be to consider all of the haplotype 1 alleles as a single population. When this is done the level of polymorphism in *janA* and *janB* drops to nearly zero (see Table 4). Thus, the comparison of variability among 3R loci would be even more extreme than shown in Figure 5, A and B, and would argue against a purely demographic explanation of the data. Although the *janA-janB* region shows the most extreme haplotype structure of loci on chromosome 3R, some loci on the X chromosome show a similar pattern (Labate *et al.* 1999; Begun and Whitley 2000). In general, it appears that the amount of variation on the X chromosome is lower than that on the autosomes, while the amount of linkage disequilibrium on the X is higher (Andolfatto and Przeworski 2000; Begun and Whitley 2000). This difference is larger than expected even after correcting for the different effective population sizes of the X and the autosomes and may be explained by stronger selection on X-linked loci (Begun and Whitley 2000).

Another potential explanation for the maintenance of two divergent haplotypes in a population is balancing selection. However, the very low levels of nucleotide polymorphism within haplotype 1 are inconsistent with this being an old balanced polymorphism. It is also possible that strong linkage disequilibria are maintained by epistatic selection favoring combinations of segregating sites (Kimura 1956; Lewontin 1974). This explanation seems unlikely because there are many linked sites over *janA* and *janB* and all of the segregating sites that distinguish the two haplotypes are at silent or noncoding positions. One possibility is that epistatic selection is acting at silent or noncoding sites to maintain mRNA or pre-mRNA secondary structures. Such interactions have been proposed to explain patterns of linkage disequilibria in the *Adh* gene of *D. pseudoobscura* (Kirby *et al.* 1995). However, we find no evidence for strongly conserved RNA secondary structures in either *janA* or *janB*, using the comparative method of Parsch *et al.* (2000).

Finally we consider a model of genetic hitchhiking (Maynard Smith and Haigh 1974), in which neutral variants are driven to high frequency in a population due to their linkage with a positively selected variant. If selection is strong or recombination is low, hitchhiking will result in a "selective sweep" that removes variation in the region flanking the selected site (Kaplan *et al.* 1989; Stephan *et al.* 1992). A reduction in polymorphism is also expected under a model of "background selection" (Charlesworth *et al.* 1993), in which neutral variants are removed from the population due to the recurrent action of purifying selection at tightly linked sites. Several features of our data are consistent with the selective sweep hypothesis. For example, genetic hitchhiking is expected to affect the frequency distribution of variants at segregating sites such that derived variants will be in higher frequency than expected under a neutral equilibrium model (Fay and Wu 2000; Kim and Stephan 2000). The significantly negative value of Fay and Wu's H in the *janA-janB* region (Table 1) agrees with this prediction. Genetic hitchhiking is also expected to skew the frequency distribution of variants at segregating sites toward rare alleles, resulting in a significantly negative value of Tajima's D (Braverman *et al.* 1995). We find that Tajima's D is significantly negative for *janA* and *janB* (Table 1). This is due to the large number of unique variants in our sample (Figure 2). However, many of these singletons can be inferred to be ancestral from the *D. melanogaster* outgroup sequence, so the negative Tajima's D is not caused by new mutations occurring after a complete selective sweep as modeled by Braverman *et al.* (1995). This is confirmed by the tests of Fu and Li (1993) that use an outgroup to identify derived singletons and do not produce a significant result when applied to our *janA-janB* data ($D = -0.34$, $P > 0.10$; $F = -0.88$, $P > 0.10$).

Although a selective sweep can explain the low level of polymorphism within haplotype 1 and the high frequency of many derived variants, it does not explain the presence of the highly divergent haplotype 2, which shows a normal level of polymorphism. One possibility is that the selective sweep is either temporally or spatially incomplete (Hudson *et al.* 1994, 1997), perhaps due to limited gene flow into ancestral African populations. It is also possible that there are positively selected variants at different sites in the two haplotypes, and fixation of a single haplotype is delayed until a recombination event brings the two variants together on the same chromosome (the "traffic" model; Kirby and Stephan 1996). A final possibility is that the positively selected site responsible for the hitchhiking event lies proximal to the *janA* gene, and there has been limited recombination between this unknown site and *janA*. Our results indicate that distally the haplotype structure is broken in the intergenic region between *janB* and *ocn*. However, the strong linkage disequilibrium in the *janA-janB* region does not allow us to define the proximal limit of the haplotype structure. Two of the chromosome 3R loci included in Begun and Whitley's (2000) study, *Rh3* (Ayala *et al.* 1993) and *boss* (Ayala and Hartl 1993), were sampled from a worldwide distribution and included some of the same lines used in our study. Both of these loci show high levels of variation relative to other loci on chromosome 3R and show no sign of haplotype structure. In particular, the line s3 (designated as "f" in those articles), which shows a quite divergent haplotype over the *janA-janB* region in our survey (Figure 2), does not show this same pattern at *Rh3* or *boss*. Both of these loci lie proximal to *janA*, with *boss* being relatively close at 96F. The observation of high variation and no haplotype structure at *boss* again argues against a purely demographic explanation for our data and indicates that the unusual haplotype structure does not extend over a large portion of the chromosome. It is tempting to speculate that the *janB* gene may be a target of positive selection. Our previous work showed that within the *D. melanogaster* species subgroup *janB* has a faster rate of evolution than *janA* or *ocn* (Parsch *et al.* 2001). *janB* is also the most divergent of these genes in a comparison between *D. simulans* and *D. melanogaster* (Table 4). In addition, *janB* shows the lowest level of within-species polymorphism (Table 1). This combination of high divergence and low polymorphism is suggestive of positive selection. However, further surveys of DNA sequence polymorphism in this region of the genome are required to identify the selected site or sites and allow estimation of the selection coefficient associated with the hitchhiking event.

We thank J. Braverman and J. Fay for providing computer programs and P. Capy and Y. Tao for providing *D. simulans* stocks. We are grateful to W. Stephan and members of the D. Hartl and J. Wakeley labs for valuable suggestions throughout the course of our work. We also thank M. Aguadé and two anonymous reviewers for comments on an earlier version of the manuscript. This work was supported by National Institutes of Health grants GM60035 and HG01250 to D.L.H.

Note added in proof: Following the submission of this manuscript, Rozas *et al.* (Rozas, J., M. Gullaud, G. Blandin and M. Aguadé, 2001, DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. Genetics **158:** 1147–1155) reported unusual haplotype structure in the *rp49* gene region in both a European and an African population of *D. simulans*. The *rp49* gene lies ~7 kb proximal to *janA* on chromosome 3R.

LITERATURE CITED

Andolfatto, P., and M. Kreitman, 2000 Molecular variation at the *In(2L)t* proximal breakpoint site in natural populations of *Drosophila melanogaster* and *D. simulans*. Genetics **154:** 1681–1691.

Andolfatto, P., and M. Przeworski, 2000 A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics **156:** 257–268.

Andolfatto, P., J. D. Wall and M. Kreitman, 1999 Unusual haplotype structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. Genetics **153:** 1297–1311.

Ayala, F. J., and D. L. Hartl, 1993 Molecular drift of the *bride of sevenless* (*boss*) gene in *Drosophila*. Mol. Biol. Evol. **10:** 1030–1040.

Ayala, F. J., B. S. Chang and D. L. Hartl, 1993 Molecular evolution of the *Rh3* gene in *Drosophila*. Genetica **92:** 23–32.

Baba-Aissa, F., M. Solignac, N. Dennebouy and J. R. David, 1988 Mitochondrial DNA variability in *Drosophila simulans*: quasi absence of polymorphism within each of the three cytoplasmic races. Heredity **61:** 419–426.

Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the *6-phosphogluconate dehydrogenase* locus in natural populations of Drosophila: selection and geographic differentiation. Genetics **136:** 155–171.

Begun, D. J., and P. Whitley, 2000 Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **97:** 5960–5965.

Bénassi, V., F. Depaulis, G. K. Meghlaoui and M. Veuille, 1999 Partial sweeping of variation at the *Fbp2* locus in a west African population of *Drosophila melanogaster*. Mol. Biol. Evol. **16:** 347–353.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Civetta, A., and R. S. Singh, 1998 Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. **15:** 901–909.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol. Biol. Evol. **15:** 1788–1790.

Depaulis, F., L. Brazier and M. Veuille, 1999 Selective sweep at the *Drosophila melanogaster Suppressor of Hairless* locus and its association with the *In(2L)t* inversion polymorphism. Genetics **152:** 1017–1024.

Fay J. C., and C .-I Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Hamblin, M. T., and C. F. Aquadro, 1996 High nucleotide sequence variation in a region of low recombination in *Drosophila simulans* is consistent with the background selection model. Mol. Biol. Evol. **13:** 1133–1140.

Hamblin, M. T., and M. Veuille, 1999 Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics **153:** 305–317.

Hasson, E., I. N. Wang, L. W. Zeng, M. Kreitman and W. F. Eanes, 1998 Nucleotide variation in the *triosephosphate isomerase* (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **15:** 756–769.

Hudson, R. R., 1990 Gene geneologies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, Vol. 7, edited by D. Futuyma and J. Antonovics. Oxford University Press, Oxford.

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

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Hudson, R. R., A. G. Sáez and F. J. Ayala, 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. Proc. Natl. Acad. Sci. USA **94:** 7725–7729.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kimura, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10:** 278–287.

Kirby, D. A., and W. Stephan, 1995 Haplotype test reveals departure from neutrality in a segment of the *white* gene of *Drosophila melanogaster*. Genetics **141:** 1483–1490.

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **144:** 635–645.

Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA **92:** 9047–9051.

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.

Labate, J. A., C. H. Biermann and W. F. Eanes, 1999 Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **16:** 724–731.

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lindsley, D. L., and L. Sandler, 1977 The genetic analysis of meiosis in female *Drosophila melanogaster*. Philos. Trans. R. Soc. Lond. Ser. B **277:** 295–312.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277.

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

Parsch, J., J. M. Braverman and W. Stephan, 2000 Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics **154:** 909–921.

Parsch, J., C. D. Meiklejohn, E. Hauschteck-Jungen, P. Hunziker and D. L. Hartl, 2001 Molecular evolution of the *ocnus* and *janus* genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. **18:** 801–811.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Russo, C. A. M., N. Takezaki and M. Nei, 1995 Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. **12:** 391–404.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

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

Tsaur, S.-C., and C.-I. Wu, 1997 Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol. Biol. Evol. **14:** 544–549.

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

Yanicostas, C., and J.-A. Lepesant, 1990 Transcriptional and translational *cis*-regulatory sequences of the spermatocyte-specific *Drosophila janusB* gene are located in the 3′ exonic region of the overlapping *janusA* gene. Mol. Gen. Genet. **224:** 450–458.

Yanicostas, C., A. Vincent and J.-A. Lepesant, 1989 Transcriptional and posttranscriptional regulation contributes to the sex-regulated expression of two sequence-related genes at the janus locus of *Drosophila melanogaster*. Mol. Cell. Biol. **9:** 2526–2535.

Yanicostas, C., P. Ferrer, A. Vincent and J.-A. Lepesant, 1995 Separate *cis*-regulatory sequences control expression of *serendipity* β and *janus A*, two immediately adjacent *Drosophila* genes. Mol. Gen. Genet. **246:** 549–560.

Communicating editor: M. Aguadé