Evidence of a High Rate of Selective Sweeps in African *Drosophila melanogaster*

Sylvain Mousset,* Lionel Brazier,*,[1] Marie-Louise Cariou,†,[1] Frédérique Chartois,†,[1] Frantz Depaulis*,[1] and Michel Veuille*,[1,2]

**Laboratoire d'Ecologie, Ecole Pratique des Hautes Etudes, Université Pierre et Marie Curie, 75005 Paris, France and* †*Laboratoire Populations Génétique et Evolution, CNRS, 91198 Gif-sur-Yvette-cedex, France*

Manuscript received May 25, 2002
Accepted for publication October 31, 2002

ABSTRACT

Assessing the rate of evolution depends on our ability to detect selection at several genes simultaneously. We summarize DNA sequence variation data in three new and six previously published data sets from the left arm of the second chromosome of *Drosophila melanogaster* in a population from West Africa, the presumed area of origin of this species. Four loci [*Acp26Aa*, *Fbp2*, *Vha68-1*, and *Su(H)*] were previously found to deviate from a neutral mutation-drift equilibrium as a consequence of one or several selective sweeps. Polymorphism data from five loci from intervening regions (*dpp*, *Acp26Ab*, *Acp29AB*, *GH10711*, and *Sos*) did not show the characteristic deviation from neutrality caused by local selective sweeps. This genomic region is polymorphic for the *In(2L)t* inversion. Four loci located near inversion breakpoints [*dpp*, *sos*, *GH10711*, and *Su(H)*] showed significant structuring between the two arrangements or significant deviation from neutrality in the inverted class, probably as a result of a recent shift in inversion frequency. Overall, these patterns of variation suggest that the four selective events were independent. Six loci were observed with no *a priori* knowledge of selection, and independent selective sweeps were detected in three of them. This suggests that a large part of the *D. melanogaster* genome has experienced the effect of positive selection in its ancestral African range.

YEARS after the neutralist challenge to the Darwinian theory (Kimura 1968), the interpretation of molecular variation still rests on an unresolved combination of the two main mechanisms of genetic change, selection and drift. Our difficulty in integrating these mechanisms bars us from properly using genomic data for assessing the level of natural selection involved in species evolution. An illustration of this inability can be found in the disagreement among studies on the scale at which hitchhiking events must be investigated in the genome (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989). A study of the extent of a selective sweep on flanking regions in the Drosophila genome found a significant reduction in variation over ~20 kb (Hudson *et al.* 1997). Thus the selection event affected several neighboring genes. However, a recent study of the *NF-κB/IκB* region in the same species concluded that the expected window of reduced polymorphism caused by selective fixation of a beneficial mutant is only 200 bases (Begun and Whitley 2000), thus suggesting that selection can potentially be independently detected at virtually any Drosophila locus. The study of a selective sweep at the *Sdic* locus (Nurminsky *et al.* 2001) detected the signature of a unique selection event over a very large region spanning two divisions (~2%) of the *Drosophila melanogaster* genome. No such pattern was found in the corresponding region for the related *D. simulans*, thus providing a neutral standard. This implicitly assumed that a selective event modifies variation patterns over a very large region of the genome. With such large effects, independent selection events could be detected in no more than 25 fruitfly genes.

We previously surveyed *D. melanogaster* molecular variation in Africa, where this species is thought to have originated (Lachaise *et al.* 1988). In a sample from the Lamto ecological station (Ivory Coast), the signature of selective sweep events was previously recorded at four genes: the *Accessory gland-specific peptide 26Aa* gene (*Acp-26Aa*; Aguadé 1998; Fay and Wu 2000), the *Fat body protein-2* gene (*Fbp2*; Bénassi *et al.* 1999), the *vacuolar ATPase 68 kD* gene (*Vha68-1*; Depaulis *et al.* 2000), and the *Suppressor of Hairless* gene [*Su(H)*; Depaulis *et al.* 1999]. These genes are located on the left arm of the second chromosome. Given this, one may ask whether these sweeps were driven by a single selection event or were causally independent. The *In(2L)t* inversion occurs on the same chromosome arm. Contrary to inversions found in the third chromosome of *D. pseudoobscura* (Dobzhansky and Queal 1938; Aquadro *et al.* 1991), the *In(2L)t* inversion of *D. melanogaster* is probably recent (Andolfatto *et al.* 1999) and has reached a high frequency in the Lamto population (61.4%, $n = 88$;

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF459524–AF459586.

[1]These authors are listed in alphabetical order.

[2]*Corresponding author:* Laboratoire d'Ecologie Cc 237, Université Pierre-et-Marie Curie, 75252 Paris Cedex 05, France.
E-mail: mveuille@snv.jussieu.fr

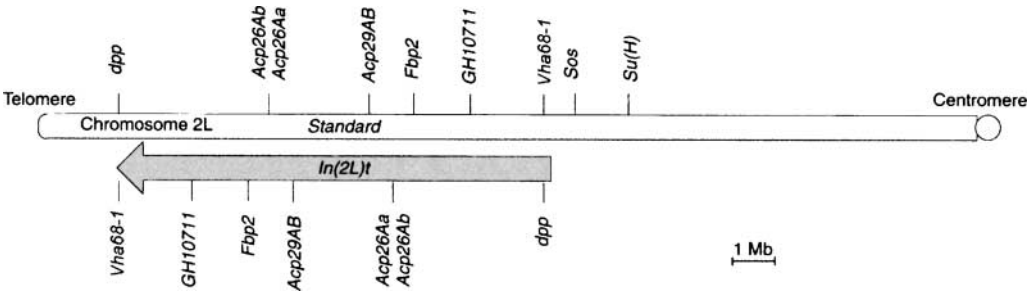

FIGURE 1.—Map of studied loci on the left arm of chromosome 2. Inversion *In(2L)t* (shaded arrow).

BÉNASSI *et al.* 1993). A parsimonious interpretation would be that selection linked to this inversion has affected allele frequency at several loci simultaneously. Using available data on DNA variation at *Fbp2*, *Vha68-1*, and *Su(H)*, GALTIER *et al.* (2000) could exclude a demographic interpretation as a unique cause of departure from equilibrium. However, they could not exclude a single selective sweep involving all loci. The purpose of this study is to resolve this question. We recorded DNA sequence variation at genes located at intervening positions in this linear arrangement of genes. We also examined genetic structuring between chromosome arrangements. We show that independent selective events were involved. Overall, one-half of randomly examined genes were affected by selection. This is the first estimate of the number of selective events having occurred in the genome of *D. melanogaster* in its ancestral range. This proportion is much higher than was previously believed.

MATERIALS AND METHODS

Data acquisition: The isolation of 84 chromosome 2 isogenic lines from an Ivory Coast sample from the Lamto ecological station and the extraction of genomic DNA were previously described (BÉNASSI *et al.* 1993). Sequence polymorphism was recorded in the same random subsample of 20 lines as for other genes (BÉNASSI *et al.* 1993; DEPAULIS *et al.* 1999). This sample includes 11 lines with *standard* chromosomes and 9 lines with *In(2L)t* chromosomes. A line of *D. simulans* was used as an outgroup.

Three loci were amplified according to standard polymerase chain reaction techniques and sequenced (primer sequences are given as supplemental data in Table S1, available at http://www.genetics.org/supplemental/). The following were sequenced: a fragment of the *decapentaplegic* (*dpp*) gene consisting of the last 441 bp of intron 2 and the first 656 bp of exon 3; a fragment of the *Son of sevenless* (*Sos*) gene sequence (SIMON *et al.* 1991) consisting of the last 397 bp of exon 5, the 69 nucleotides of intron 5, the 634 bp of exon 6, and the first 49 bp of intron 6; and a fragment of the *GH10711* putative gene consisting of 188 bp of exon 4, 794 bp of intron 4, and the first 159 bp of exon 5. They were aligned using ClustalX software (THOMPSON *et al.* 1997) and checked manually.

Polymorphism at these loci was compared to sequence polymorphism data previously obtained from the same sample for *Su(H)* (DEPAULIS *et al.* 1999) and for *Vha68-1* (DEPAULIS *et al.* 2000). It was also compared to other sequence polymorphism data obtained from different subsamples of our Ivory Coast sample: *Fbp2* ($n = 10$; BÉNASSI *et al.* 1999), *Acp26Aa* and *Acp26Ab* ($n = 24$; AGUADÉ 1998), and *Acp29AB* ($n = 14$; AGUADÉ 1999). Figure 1 shows the position of these genes on the chromosomal map. Their genetic properties are summarized in Table 1. The whole area covers 46.5 cM of genetic map.

Data treatment: The effective number of synonymous and nonsynonymous sites, descriptive population statistics, and some statistical tests were performed using the DnaSP3.53 program (ROZAS and ROZAS 1999). Phylogenies were obtained using MEGA 2.1 (KUMAR *et al.* 1994). The fixation index F_{ST} was computed according to HUDSON *et al.* (1992). Recombination rates were computed from the comparisons of adjusted coefficients of exchange (KINDAHL 1994) and the DNA content of chromosomal bands (SORSA 1988).The multiple Hudson-Kreitman-Aguadé (HKA) test was run using a program obtained from Jody Hey (http://lifesci.rutgers.edu/~heylab/). Neutral expectations for haplotype diversity, haplotype number (DEPAULIS and VEUILLE 1998; DEPAULIS *et al.* 2001), and FAY and WU's (2000) *H*-statistic were obtained by generating coalescents following standard procedures (HUDSON 1993). Coalescent simulations using the observed number of segregating sites were run with and without recombination. Recombination was then set at the value of $4N_e r = 0.008$ events/bp. This value is much lower than its estimated value in any of these genes and resulted in conservative tests. The program used for simulations (Allelix) can be obtained from S. Mousset (http://www.snv.jussieu.fr/mousset/).

RESULTS

Sequence alignments for newly sequenced loci are shown in Figure 2. Sequence alignments for *Vha68-1* and *Su(H)* loci obtained from the same sample are shown in the supplemental data Table S2 at http://www.genetics.org/supplemental/. Summary statistics of gene polymorphism are shown in Table 2.

Sequence polymorphism at *dpp*, *GH10711*, and *Sos*: Sequence polymorphism data from 20 *D. melanogaster* and 1 *D. simulans* chromosomes were collected at *dpp*, *GH10711*, and *Sos*. In *D. melanogaster* at *dpp*, a total of 30 polymorphic sites (corresponding to 31 mutations) and five length polymorphisms were identified over the 1082 bp examined (excluding sites with alignment gaps) and formed 15 haplotypes. At *GH10711*, 60 polymorphic sites (61 mutations) and six length polymorphisms were identified over the 1074 bp examined and formed 14 haplotypes. At *Sos*, 23 polymorphic sites (23 mutations) were identified over the 1146 bp examined and formed 13 haplotypes. A minimum number of four recombination events between informative sites was inferred using the four-gamete rule (HUDSON and KAPLAN 1985) at both *Sos* and *dpp*, whereas at least nine events were

TABLE 1

Genetic properties of studied loci

Locus	Polytene map position	Recombination rate (10^{-8})	Type of data	Distance from telomere (Mb)	*n*	Reference
In(2L)t (distal)	22D3-E1		Inversion	1.6		
dpp	22F1-F2	1.8	Sequence	1.7	20	This study
Acp26Aa	26A5	1.5	Sequence	5.3	24	Aguadé (1998)
Acp26Ab	26A5	1.5	Sequence	5.3	24	Aguadé (1998)
Acp29AB	29C1	3.1	Sequence	7.7	14	Aguadé (1999)
Fbp2	30B5	3.1	Sequence	8.8	10	Bénassi *et al.* (1999)
GH10711	32A4-5	3.1	Sequence	10.1	20	This study
Vha68-1	34A5	0.6	Sequence	12.4	20	Depaulis *et al.* 2000
In(2L)t (proximal)	34A8-9		Inversion	12.5		
Sos	34D1-D2	0.6	Sequence	13.1	20	This study
Su(H)	35C1	1.5	Sequence	14.4	20	Depaulis *et al.* (1999)

inferred at *GH10711*. Replacement variants were usually found to occur at low frequency, except at *GH10711* where two replacement variants occurred, respectively, in 4 and 14 of the 20 sequenced lines.

Neutrality tests: HKA tests (Hudson *et al.* 1987) were run between all pairs of loci (Table 3). One locus, *Vha68-1*, consistently showed significant deviation from neutrality with all other loci, except *Acp26Aa* and *Acp29AB*. All other pairwise comparisons, except the one between *Acp26Ab* and *Acp29AB*, were nonsignificant. A multilocus HKA test, including all loci, did not show significant deviation from neutrality in the nine-loci sample (result not shown).

The *H*- and *K*-haplotype tests, as proposed by Depaulis and Veuille (1998), when run with a no recombination assumption (Table 4), were significant for *Su(H)* as previously shown (Depaulis *et al.* 1999). Adding recombination made the *H*-test (based on haplotype diversity) also significant for *Vha68-1*. These tests were also performed separately for each chromosomal class. Significant deviation from neutrality was found using the *H*- and *K*-tests for the *inverted* and the *standard* class at the *Su(H)* locus. The *H*-test was significant for inverted chromosomes at the *GH10711* locus, and the *K*-test was also significant for these chromosomes when using a conservative recombination rate. However, *inverted* chromosomes are not a random subsample and deviation from neutrality may result from sampling data within an allelic class (Innan and Tajima 1997, 1999). Hudson *et al.*'s (1994) haplotype test showed significant deviation from neutrality at *Fbp2*, as previously shown (Bénassi *et al.* 1999). This test uses a sliding window approach, which was inappropriate for the other loci, due to their short fragment length (see discussion).

Fay and Wu's (2000) test was not significant for any loci when run without recombination. However, it became significant with recombination at *Acp26Aa* and *Sos* (Table 5). Significant non-neutrality had previously been found at *Acp26Aa*, using a mixed African population sample from Ivory Coast and Malawi (Fay and Wu 2000), whereas our study includes only the Ivory Coast subset for consistency with our other data. The *P* values indicated marginally significant non-neutrality ($P < 0.06$) for three other loci: *dpp*, *Vha68-1*, and *Su(H)*. This test was run independently on each chromosomal class and showed significant departure from neutrality at the *Acp26Aa* locus for *standard* lines and at *Su(H)* for *In(2L)t* lines.

Tajima's *D* (Tajima 1989) and Fu and Li's *D* (Fu and Li 1993) neutrality tests were always nonsignificant (results not shown). However, high positive values were found for the two genes that were significant for haplotype tests (Fu and Li's $D = 0.682$ for *Fbp2* and $D = 1.181$ for *Su(H)*).

Genetic structuring between chromosomal arrangements: Three loci [*dpp*, *Sos*, and *Su(H)*] showed significant genetic structuring between chromosome arrangements (Table 6). These genes are the closest to inversion breakpoints, excluding *Vha68-1*, which shows too little variation for homogeneity tests to be applied. Genetic exchange between *In(2L)t* and *standard* lines was detected at six loci using Betrán *et al.*'s (1997) procedure.

Phylogenetic analysis of linkage with the inversion: The association of haplotypes with chromosome arrangements is illustrated in Figure 3 using neighbor-joining trees. Since recombination events or gene conversion was found to have occurred between *In(2L)t* and *standard* lines in these genes (see above), internal branches do not represent true descent relationships and, similarly, bootstrap values should be interpreted with caution. Except where gene conversion was found, the inverted chromosomes of *dpp* and *GH10711* were clustered within one or two subsections of the tree. This pattern was less obvious at *Sos*. The result for *dpp* strongly suggests that a single recombination event occurred after the divergence between *standard* and *In(2L)t*, and this is responsible for the origin of a new family of *standard* haplotypes. This is compatible with indications

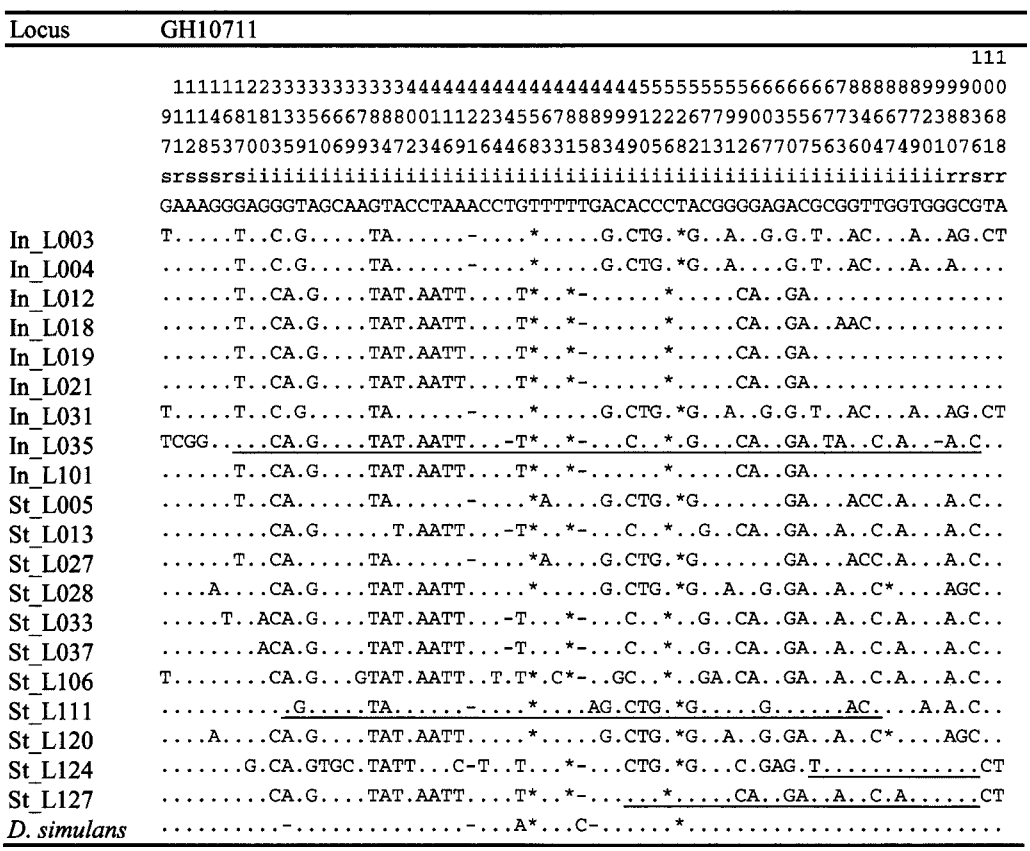

FIGURE 2.—Alignment of polymorphic sites. The ancestral state (when available) inferred from *D. melanogaster/D. simulans* comparison was used as reference. –, 1-bp deletion; *, deletion >1 bp; i, intronic mutation; s, silent mutation; r, replacement mutation. Gene conversion tracts between *inverted* and *standard* chromosomes are underlined.

TABLE 2

Summary statistics of sequence polymorphism data

Locus	*dpp*	*Acp26Aa*	*Acp26Ab*	*Acp29AB*	*Fbp2*	*GH10711*	*Vha68-1*	*Sos*	*Su(H)*
Sample size	20	24	24	14	10	20	20	20	20
Length	1082	1053	567	1755	2156	1074	1063	1146	1014
No. of silent sites[a]	592	438	351	1197	1529	826	525	372	768
No. of polymorphic sites	30	40	37	56	65	60	11	23	44
Total no. of mutations	31	41	39	56	67	61	11	23	44
No. of singletons	10	11	8	12	15	23	7	4	5
No. of silent mutations	30	21	35	52	62	56	10	20	44
π (10^{-2})[b]	0.684	0.962	1.548	1.050	1.165	1.675	0.152	0.571	1.504
θ_w (10^{-2})[c]	0.808	1.043	1.842	1.003	1.098	1.601	0.292	0.566	1.223
π silent (10^{-2})	1.232	1.106	2.124	1.413	1.566	2.030	0.289	1.651	1.984
θ_w silent (10^{-2})	1.428	1.283	2.664	1.365	1.433	1.911	0.537	1.512	1.613
Tajima's *D*	−0.601	−0.298	−0.612	0.205	0.253	0.186	−1.698	0.038	0.918

[a] Synonymous and noncoding sites.
[b] Average heterozygosity per nucleotide (Tajima 1989).
[c] Mutation parameter estimate (Watterson 1975).

that the age of the *In(2L)t* inversion is much less than the coalescence time of autosomes in highly recombining regions (Andolfatto *et al.* 1999). This is also compatible with previous observations showing an unusual haplotype structure of this gene in an American population, even though no inversions were present (Richter *et al.* 1997).

DISCUSSION

As *D. melanogaster* probably originated in Africa (Lachaise *et al.* 1988), the genes considered here provide an opportunity to examine the effect of natural selection on DNA variation in an ancestral population living in its natural environment. Under these circumstances, it is more likely that this population is at equilibrium for demographic and genetic factors. Nine genes were examined in a population in which the *In(2Lt)* inversion shows a high frequency. The signature of selection was observed in four of these loci [*Acp26Aa*, *Fbp2*, *Vha68-1*, and *Su(H)*].

Detectability of selective sweep events: Detecting selective sweeps depends on the significance of neutrality tests. Given the depletion of nucleotide polymorphism after a complete sweep, variation is scarce and evidence of selection is difficult to obtain from tests such as Tajima's (1989) or Fu and Li's (1993) that use a single distribution of variation. Power is presumably greater for tests such as the HKA test (Hudson *et al.* 1987) that use another locus as a reference. In partial selective sweeps, recombination has occurred between the selected and the surveyed loci during the selective phase, resulting in the preservation of a large amount of variation in recombining chromosomes. This leaves aberrant haplotype patterns (Kaplan *et al.* 1989), thus leading to increased power. Interestingly, most selective sweep events detected in Drosophila concern partial sweeps [*e.g.*, at *Sod*, *White*, *Est-6*, *Acp70A*, *Fbp2*, *Su(H)*, and *In(2L)t*; Kirby and Stephan 1995; Cirera and Aguadé 1997; Andolfatto *et al.* 1999; Balakirev *et al.* 1999; Bénassi *et al.* 1999; Depaulis *et al.* 1999]. This may indicate that selective sweeps affect polymorphism at many loci in

TABLE 3

Pairwise HKA tests between loci

Locus	*dpp*	*Acp26Aa*	*Acp26Ab*	*Acp29AB*	*Fbp2*	*GH10711*	*Vha68-1*[a]	*Sos*
Acp26Aa	0.719							
Acp26Ab	1.07	3.536						
Acp29AB	0.966	0	5.373*					
Fbp2	0.044	0.483	2.011	0.507				
GH10711	0.194	1.685	0.416	2.107	0.446			
Vha68-1[a]	5.045*	2.182	10.553***	2.221	4.193*	6.658**		
Sos	0.14	1.461	0.376	2.059	0.401	0.002	7.047**	
Su(H)	0.035	1.096	0.799	1.406	0.164	0.066	5.697*	0.042

χ^2 and *P* value when significant. Only silent sites were used. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$.
[a] All significant tests for *Vha68-1* indicated a defect of polymorphism (or an excess of divergence).

TABLE 4

***H*- and *K*-haplotype tests**

Locus	*dpp*	*Acp26Aa*	*Acp26Ab*	*Acp29AB*	*Fbp2*	*GH10711*	*Vha68-1*	*Sos*	*Su(H)*
Sample size (n)	20	24	24	14	10	20	20	20	20
No. of mutations (S)	31	41	39	56	67	61	11	23	44
No. of haplotypes (K_{obs})	15	22	21	11	10	14	9	13	7
Sample haplotype diversity (H_{obs})	0.92	0.95	0.95	0.90	0.90	0.91	0.67	0.91	0.76
$P(K \leq K_{obs} \mid S, n)$[a]	0.983	1	1	0.670	1	0.624	0.934	0.961	0.012[c,d]
$P(H \leq H_{obs} \mid S, n)$[a]	0.961	1	0.999	0.546	1	0.506[c]	0.168	0.980	0.017[c,d]
$P(K \leq K_{obs} \mid S, n)$[b]	0.928	1	1	0.461	1	0.420[c]	0.652	0.813	0.001[c,d]
$P(H \leq H_{obs} \mid S, n)$[b]	0.856	1	0.999	0.336	1	0.314[c]	0.049	0.895	0.003[c,d]

After Depaulis and Veuille (1998), run after (Depaulis *et al.* 2001).
[a] Assuming no recombination, computed from 50,000 simulations.
[b] Assuming recombination parameter $4N_e r = 0.008$/pb, computed from 20,000 simulations.
[c] This test was significant for the *In(2L)t* subsample.
[d] This test was significant for the *standard* subsample.

addition to the actual targets of selection or that selective sweeps are more readily detected in partial sweeps. The haplotype tests that reveal partial selective sweeps are designed to detect linkage disequilibrium immediately after the sweep and are thus efficient only for recent events.

Using computer simulations Fay and Wu (2000; Figure 3) estimated that their test has no power to detect selective sweeps when the c/s ratio is >0.1 (where c is the rate of recombination between the observed and the selected loci and s is the haploid selection coefficient). The smallest physical distance between loci showing evidence of selection in our sample is 2×10^6 nucleotides (Table 1). The lowest recombination rate at studied loci is $r = 6 \times 10^{-9}$ event/bp. Assuming this value over the whole region, a favored gene occurring between two observed loci should have a selective advantage of ≥ 0.06 for the selective sweep to be detected at one of these loci with at least a 5% probability. The power of the other haplotype tests is probably not very different. It is therefore highly improbable that the partial selective sweeps observed at *Acp26Aa*, *Fbp2*, *Vha68-1*, and *Su(H)* result from the same selective event. The only two possibilities are either a shift in the frequency of the *In(2L)t* inversion, which extends over a large region, or population structuring.

Population structure was previously found in African *D. melanogaster* (Michalakis and Veuille 1996). Local genetic drift combined with a low level of migration could produce the observed pattern of variation. Population structure, however, increases the length of internal branches of a coalescent and thus should lead to positive Tajima's D values. This is not observed at the studied loci since four of the nine Tajima's D are negative (Table 2).

Below we examine whether a recent increase of *In(2L)t* frequency, as suggested by Andolfatto *et al.* (1999), could have caused a general deviation from neutral expectation. We then try to identify independent selective events and to assess whether selection was suspected before the study or found at random.

Linkage with the *In(2L)t* inversion: The frequency of

TABLE 5

Neutrality test based on derived mutations

Locus	*dpp*	*Acp26Aa*	*Acp26Ab*	*Acp29AB*	*Fbp2*	*GH10711*	*Vha68-1*	*Sos*	*Su(H)*
Sample size (n)	20	24	24	14	10	20	20	20	20
No. of oriented mutations (M)[a]	27	35	37	43	52	57	7	21	40
θ_π (per gene)	6.62	8.62	8.66	14.16	20.22	16.97	1.13	6.18	13.84
θ_H (per gene)	12.23	15.82	14.12	14.6	21.11	23.55	3.51	11.29	21.53
$P(H \leq H_{obs} \mid M, n)$[b]	0.088	0.082[d]	0.121	0.305	0.297	0.153	0.062	0.071	0.091[e]
$P(H \leq H_{obs} \mid M, n)$[c]	0.057	0.047[d]	0.104	0.379	0.369	0.123	0.051	0.047	0.057[e]

After Fay and Wu (2000).
[a] Mutations with a known ancestral state (*D. simulans* was used as an outgroup).
[b] Assuming no recombination, computed from 50,000 simulations.
[c] Assuming recombination parameter $4N_e r = 0.008$/pb, computed from 20,000 simulations.
[d] This test was significant for the *standard* subsample.
[e] This test was significant for the *In(2L)t* subsample.

TABLE 6

Genetic structuring between chromosomal arrangements

Locus	*dpp*	*Acp26Aa*	*Acp26Ab*	*Acp29AB*	*Fbp2*	*GH10711*	*Vha68-1*	*Sos*	*Su(H)*
No. of *In(2L)t* chromosomes	9	10	10	10	6	9	9	9	9
No. of *standard* chromosomes	11	11	11	4	4	11	11	11	11
No. of informative sites[a]	23	29	25	46	60	46	4	19	57
No. of shared polymorphisms[b]	7	22	23	30	23	32	0	6	23
No. of genetic exchange events[c]	1	1	1	0	1	4	0	0	2
In(2L)t to *standard* ratio for π at silent sites	0.618	0.985	0.851	0.951	0.565	0.869	0.289	1.218	0.760
F_{ST}	0.190	−0.018	−0.034	−0.011	0.152	0.076	0.095	0.175	0.314
P value	0.002	0.662	0.768	0.466	0.082	0.107	0.083	0.006	0.002

[a] Including length polymorphisms.
[b] No fixed differences were found between *In(2L)t* and *standard* lines at any locus.
[c] Computed after Betrán *et al.* (1997).

In(2L)t varies widely in Africa (Veuille *et al.* 1998). This may result from selection or from drift. The absence of nucleotide variation at its breakpoints suggests that the inversion is very young (Andolfatto *et al.* 1999). Old world populations of *D. melanogaster* are genetically differentiated for silent variation in *Adh slow* haplotypes, which are in linkage disequilibrium with this inversion (Bénassi and Veuille 1995). No such structuring was observed for microsatellites from other regions of the second chromosome (Michalakis and Veuille 1996). These observations suggest that the very high frequency of this inversion in Ivory Coast results from selection. In this study, genetic structuring between inverted and standard chromosomes was found for loci close to *In(2L)t* breakpoints [*dpp*, *Sos*, and *Su(H)*].

In a comparison of nucleotide diversity between *In(2L)t* and *standard*, inverted chromosomes were less polymorphic in eight out of nine loci (Table 6). This result was significant in a sign test ($P = 0.02$). The trend (nonsignificant) is the same for Watterson's estimator of nucleotide diversity, which should be equal to nucleotide diversity under a neutral model, since only seven out of nine loci are less polymorphic within the inverted than within the *standard* chromosomal class ($P = 0.11$). However, this trend suggests that the *In(2L)t* chromosomal class is less polymorphic. This is consistent with a recent increase of *In(2L)t* frequency (Andolfatto *et al.* 1999). The significant lack of haplotype diversity for *In(2L)t* at *GH10711* (Table 4) could also result from a recent shift in *In(2L)t* frequency. However, these results should be interpreted with caution since a chromosomal class isolated from the rest of the population does not constitute a random sample of chromosomes, and applying neutrality tests on this subsample may lead to invalid conclusions (Innan and Tajima 1997, 1999).

A recent shift in *In(2L)t* frequency may have skewed the frequency spectrum of mutations at loci where significant structuring is observed with respect to the inversion. This may explain the low *P* values obtained from Fay and Wu's test for *dpp*, *Vha68-1*, *Sos*, and *Su(H)* (Table 5), which are always associated with significant F_{ST} values (Table 6). At *Vha68-1* and *Su(H)*, however, there is independent evidence of a selective sweep, suggesting that patterns of variation in each of these genes were altered by different events: a selective sweep and a shift in the inversion frequency. A single selective sweep occurs at a given time and extends over a given area. Considering the time parameter, Galtier *et al.* (2000) showed that a single bottleneck could not have caused the deviation from neutrality observed at *Fbp2*, *Vha68-1*, and *Su(H)*. Polymorphism data from the three newly sequenced intervening loci (*dpp*, *GH10711*, and *Sos*) enable us to partition the deviations from neutrality previously detected at other loci [*Acp26Aa*, *Fbp2*,*Vha68-1*, and *Su(H)*] into effects from different selective events.

Selective sweep at *Su(H)*: Sequence variation in this gene was substantial, and yet was distributed among few haplotypes, thus departing significantly from neutral equilibrium in a haplotype test (Depaulis *et al.* 1999). This departure from neutrality was also observed in each chromosomal class (Table 4), suggesting that the shift in inversion frequency was not responsible for the observed pattern of variation at this locus. There were few singletons (5 out of 44 polymorphisms), suggesting that the sweep is recent and that few mutations have accumulated since. This polymorphic pattern makes it unlikely that the observed 1-kb DNA fragment was the focus of the sweep. More likely, selection occurred in a nearby region, within or outside *Su(H)*, and some haplotypes were rescued from the sweep through recombination. Linkage disequilibrium with *In(2L)t* at *Su(H)* was first noticed during a microsatellite survey (Depaulis *et al.* 1999). Sequence variation then showed that two phenomena were superimposed; there was structuring with the inversion, but there was also a selective sweep within each chromosomal arrangement. This means that this gene was taken at random with respect to the selective sweep, but not with respect to linkage disequilibrium with the inversion.

Selective sweep at *Vha68-1*: Variation in this gene was

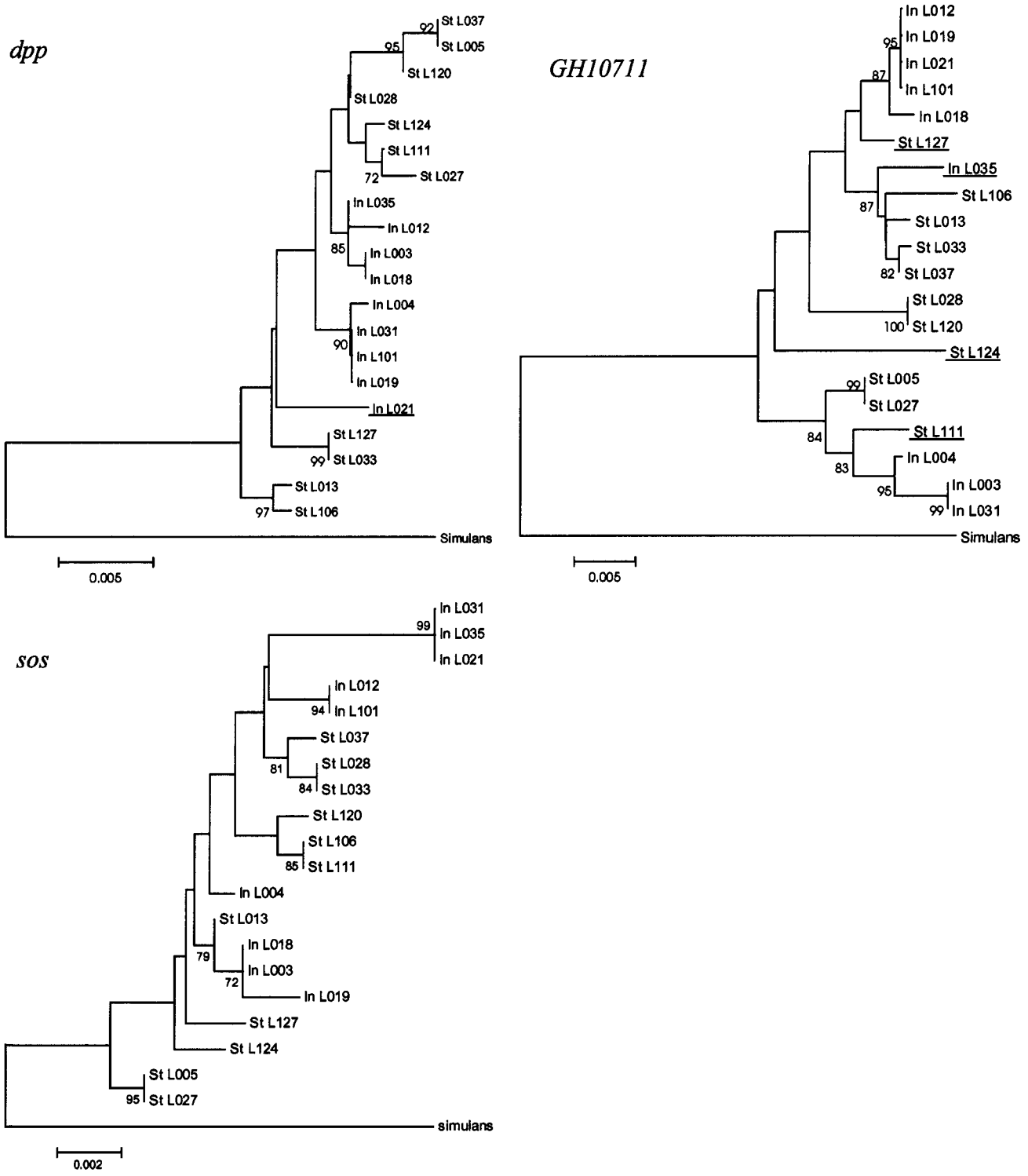

FIGURE 3.—Neighbor-joining trees of haplotypes at *dpp, GH10711*, and *Sos*. Bootstrap percentages were calculated among 5000 replicates; sequences for which genetic exchange between *In(2L)t* and *standard* lines was identified are underlined.

very low and significantly departed from neutrality in a HKA test (HUDSON *et al.* 1987) when other genes were used as a reference. An HKA test performed against the *GH10711* locus led to significant results both within *In(2L)t* ($\chi^2 = 8.441$, $P = 0.004$) and within the *standard* class ($\chi^2 = 5.474$, $P = 0.019$), suggesting that the selective sweep was not driven by the inversion shift only. Moreover, the 11 variants found in this gene were not very frequent; 8 of them are singletons and the other 3 are rare variants. A selective sweep possibly occurred

across this gene or very close to it, and observed polymorphisms probably result from mutations occurring afterward. In the standard arrangement, this gene is located between two neutral loci (*GH10711* and *Sos*). This suggests that the selective sweep at this locus is independent of those found elsewhere. Departure from neutrality in this gene was initially unexpected when first observed, as the authors' purpose was to study population structuring at a gene close to *In(2L)t* (Depaulis *et al.* 2000).

Selective sweep at *Fbp2*: Sequence variation in this gene deviated from equilibrium using a haplotype test based on the frequency of the major haplotype (Hudson *et al.* 1994; Kirby and Stephan 1995). This gene lies within the *In(2L)t* inversion. Of 10 sequences, 5 showed the same haplotype over the entire coding unit, 3 of them belonged to the *In(2L)t* class, and 2 belonged to the *standard* class. The *Fbp2* gene probably underwent hitchhiking due to selection at a nearby locus (Bénassi *et al.* 1999). The driving gene of this selective sweep probably lies close to *Fbp2* within the inversion. Nucleotide variation was first investigated at *Fbp2* to examine amino acid variation as this gene encodes a protein that is methionine rich (Rat *et al.* 1991; Meghlaoui and Veuille 1997). There was no segregating methionine in the data set and no methionine was fixed in the *D. melanogaster* lineage since its speciation with *D. simulans*. However, inspection of the silent sites showed evidence of hitchhiking due to an undetermined selection event. After this study, we changed our sampling method and recorded polymorphism >1 kb from intronic regions in 20 lines rather than 2 kb from coding regions in 10 lines. This change aimed to use *a priori* tests like *H*- and *K*-tests rather than *a posteriori* tests that use a sliding window approach. Despite these changes in the sampling design, *Fbp2* significantly departed from neutrality. Therefore, it is valid to compare *Fbp2* results with those of other loci, even though more satisfying tests were applied to them.

Selective sweep at *Acp26Aa*: In this study, we confirmed that sequence variation in this gene deviated from neutral equilibrium using Fay and Wu's test (Fay and Wu 2000) for our population sample. This test remained significant when run on *standard* chromosomes, showing that the shift in *In(2L)t* frequency is unlikely to be the cause of departure from neutrality. Fay and Wu (2000) previously showed this result to be associated with the absence of polymorphism >200 bp in this gene (Aguadé 1999). As Drosophila *Acp* genes encode sex peptides that are thought to contribute to sperm competition, hypotheses on selection clearly motivated population genetics studies in these genes (Aguadé *et al.* 1992; Aguadé 1998, 1999), meaning that *Acp26Aa* was not chosen at random in the Drosophila genome.

Proportion of selective events in a random sample of genes: In this study, selection was shown to have independently affected variation patterns at four genetic loci [*Acp26Aa*, *Fbp2, Vha68-1*, and *Su(H)*]. The contrasting patterns of polymorphism observed between loci in our African sample make alternative demographic explanations, such as bottleneck, founder effects, or population structure unlikely, although these events may have increased the variance over loci. In addition, a selective sweep was probably transmitted to several genes through a change in *In(2L)t* frequency. An obvious consequence of this shift in inversion frequency is the lower level of polymorphism in inverted chromosomes, as previously observed by Andolfatto *et al.* (1999) at the proximal breakpoint of this inversion. As the four regions lying proximally and distally to breakpoints cosegregate as a single unit, an inversion can also be considered a "locus."

However, this sample of loci was not taken at random, since selection was already suspected for *Acp* genes, and since inversion polymorphisms are considered potential targets for selection in many population genetics studies. Sequence variation was blindly examined at only six loci. Three loci [*Fbp2*, *Vha68-1*, and *Su(H)*] out of the six deviated from neutrality. Some of these deviations may be due to a demographic event. For instance, we would expect about three genes out of six to depart significantly from neutrality using a test with 50% power to detect an event. Although this proportion is imprecise, it suggests that "footprints" of positive selection are present in a substantial proportion of genes.

D. melanogaster is a reference organism that has been used in many evolutionary studies. Most of these studies were conducted in "derived" populations that may have been affected by adaptation to new environments and to founding events. Such a high proportion of nonneutral loci has been previously recorded in a sample of 20 genes from highly recombining regions in *D. melanogaster* (Moriyama and Powell 1996). However, this initial review considered data from studies carried out on a nonrandom sample of genes. Furthermore, all of these elementary studies consisted of data for derived populations of *D. melanogaster*. The present study is the first to consider sequence polymorphism in several genes from the same African population of *D. melanogaster*, thus providing a rough idea of selection pressure in this organism in its original habitat. The shift in *In(2L)t* frequency was observed both at neutral loci (*dpp*, *Sos*, *GH10711*) and at *Su(H)*. In *dpp*, a family of haplotypes recombined away from the inversion, thus generating a singular haplotype structure of the sample. It is probably the same pattern that was observed in a New Jersey population (Richter *et al.* 1997). However, the low frequency of this inversion in North America did not allow the authors to make any definitive conclusions. Our observations show that haplotype patterns can have a complex historical origin and that populations are likely to be by no means simple. The impact of selection

upon Drosophila genome variation in the recent history of this species appears very substantial. Chromosomal inversions could play an important role in shaping variation along chromosomes. It would be of interest to survey molecular polymorphism from regions where no common inversions occur in the range of the species.

We thank two anonymous reviewers for valuable comments and suggestions, Michèle Huet for technical assistance, and John Ewen and Karen McCoy for help in preparing the manuscript. This research was financially supported by the Centre National de la Recherche Scientifique Groupe de Recherche 1928.

LITERATURE CITED

Aguadé, M., 1998 Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *Drosophila melanogaster* species complex. Genetics **150:** 1079–1089.

Aguadé, M., 1999 Positive selection drives the evolution of the *Acp-29AB* accessory gland protein in Drosophila. Genetics **152:** 543–551.

Aguadé, M., N. Miyashita and C. H. Langley, 1992 Polymorphism and divergence in the Mst26A male accessory gland gene region in Drosophila. Genetics **132:** 755–770.

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.

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

Balakirev, E. S., E. I. Balakirev, F. Rodriguez-Trelles and F. J. Ayala, 1999 Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. Genetics **153:** 1357–1369.

Begun, D. J., and P. Whitley, 2000 Genetics of alpha-amanitin resistance in a natural population of *Drosophila melanogaster*. Heredity **85:** 184–190.

Bénassi, V., and M. Veuille, 1995 Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, west Africa and east Africa. Genet. Res. **65:** 95–103.

Bénassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

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.

Betrán, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146:** 89–99.

Cirera, S., and M. Aguadé, 1997 Evolutionary history of the sex-peptide (*Acp70A*) gene region in *Drosophila melanogaster*. Genetics **147:** 189–197.

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.

Depaulis, F., L. Brazier, S. Mousset, A. Turbé and M. Veuille, 2000 Selective sweep near the *In(2L)t* inversion breakpoint in an African population of *Drosophila melanogaster*. Genet. Res. **76:** 149–158.

Depaulis, F., S. Mousset and M. Veuille, 2001 Haplotype tests using coalescent simulations conditional on the number of segregating sites. Mol. Biol. Evol. **18:** 1136–1138.

Dobzhansky, T., and M. L. Queal, 1938 Genetics of natural populations. I. Chromosome variation in populations of Drosophila pseudoobscura inhabiting isolated mountain ranges. Genetics **23:** 463–484.

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.

Galtier, N., F. Depaulis and N. H. Barton, 2000 Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics **155:** 981–987.

Hudson, R. R., 1993 The how and why of generating gene genealogies, pp. 23–36 in *Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.

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., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

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

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. Saez 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.

Innan, H., and F. Tajima, 1997 The amounts of nucleotide variation within and between allelic classes and the reconstruction of the common ancestral sequence in a population. Genetics **147:** 1431–1444.

Innan, H., and F. Tajima, 1999 The effect of selection on the amount of nucleotide variation within and between allelic classes. Genet. Res. **73:** 15–28.

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

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

Kindahl, E. C., 1994 Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. Ph.D. Thesis, Cornell University, Ithaca, NY.

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.

Kumar, S., K. Tamura and M. Nei, 1994 MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput. Appl. Biosci. **10:** 189–191.

Lachaise, D., M. Cariou, J. R. David, F. Lemeunier, L. Tsacas *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup, pp. 159–225 in *Evolutionary Biology*, edited by M. K. Hecht, B. Wallace and G. T. Prance. Plenum, New York.

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

Meghlaoui, G. K., and M. Veuille, 1997 Selection and methionine accumulation in the *fat body protein 2* gene (*Fbp2*), a duplicate of the *Drosophila* alcohol dehydrogenase (*Adh*) gene. J. Mol. Evol. **44:** 23–32.

Michalakis, Y., and M. Veuille, 1996 Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogaster* and its relation to the recombination rate. Genetics **143:** 1713–1725.

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

Nurminsky, D., D. D. Aguiar, C. D. Bustamante and D. L. Hartl, 2001 Chromosomal effects of rapid gene evolution in *Drosophila melanogaster*. Science **291:** 128–130.

Rat, L., M. Veuille and J. A. Lepesant, 1991 *Drosophila* fat body protein *P6* and *alcohol dehydrogenase* are derived from a common ancestral protein. J. Mol. Evol. **33:** 194–203.

Richter, B., M. Long, R. C. Lewontin and E. Nitasaka, 1997 Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in Drosophila. Genetics **145:** 311–323.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated pro-

gram for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Simon, M. A., D. D. Bowtell, G. S. Dodson, T. R. Laverty and G. M. Rubin, 1991 *Ras1* and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the *sevenless* protein tyrosine kinase. Cell **67:** 701–716.

Sorsa, V., 1988 *Chromosome Maps of Drosophila*. CRC Press, Boca Raton, FL.

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

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.

Veuille, M., V. Bénassi, S. Aulard and F. Depaulis, 1998 Allele-specific population structure of *Drosophila melanogaster* alcohol dehydrogenase at the molecular level. Genetics **149:** 971–981.

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

Communicating editor: M. Aguadé