Selective Sweep at the *Drosophila melanogaster Suppressor of Hairless* Locus and Its Association With the *In(2L)t* Inversion Polymorphism

Frantz Depaulis, Lionel Brazier and Michel Veuille

Biologie Intégrative des Populations, Laboratoire d'Ecologie, Paris 6 University, 75005 Paris, France

Manuscript received September 4, 1998
Accepted for publication April 8, 1999

ABSTRACT

The hitchhiking model of population genetics predicts that an allele favored by Darwinian selection can replace haplotypes from the same locus previously established at a neutral mutation-drift equilibrium. This process, known as "selective sweep," was studied by comparing molecular variation between the polymorphic *In(2L)t* inversion and the *standard* chromosome. Sequence variation was recorded at the *Suppressor of Hairless* (*Su[H]*) gene in an African population of *Drosophila melanogaster*. We found 47 nucleotide polymorphisms among 20 sequences of 1.2 kb. Neutrality tests were nonsignificant at the nucleotide level. However, these sites were strongly associated, because 290 out of 741 observed pairwise combinations between them were in significant linkage disequilibrium. We found only seven haplotypes, two occurring in the 9 *In(2L)t* chromosomes, and five in the 11 *standard* chromosomes, with no shared haplotype. Two haplotypes, one in each chromosome arrangement, made up two-thirds of the sample. This low haplotype diversity departed from neutrality in a haplotype test. This pattern supports a selective sweep hypothesis for the *Su(H)* chromosome region.

THE theory of genetic hitchhiking (Maynard-Smith and Haigh 1974) predicts that natural selection on favored genes can be revealed by its driving effect on allele frequencies at neighboring loci (Aguadé *et al.* 1989; Stephan and Langley 1989). Attempts to use this effect in genome-wide surveys of molecular polymorphism as evidence of Darwinian selection have failed due to a strong technical constraint. In highly recombining parts of the genome, the effect of such "selective sweeps" (Kaplan *et al.* 1989; Begun and Aquadro 1992) is limited to small regions and cannot be easily detected. Low recombining regions, such as Drosophila telomeres and pericentromeric regions, show an absence of variation that can be explained either by selective sweeps or by "background selection" (Charlesworth *et al.* 1993). The latter is the loss of diversity caused by the recurrent elimination of chromosomes bearing deleterious mutations. Comparing predictions of realistic quantitative models with empirical data shows that background selection can easily explain the contrast in polymorphism levels between "highly recombining" and "low recombining" regions at a genome-wide scale in Drosophila (Hudson and Kaplan 1995). Even though this global pattern conforms to neutral theory predictions, the burden of Darwinian selection undergone by genomes in natural conditions can still be substantial, and its study remains an essential objective of population genetics. Genetic hitchhiking can more easily be identified in highly recombining regions, where background selection would reduce segregating neutral variation, *e.g.*, in Drosophila, by only ~33% (Hudson and Kaplan 1995). In other words, any departure from neutrality at a specific locus could be ascribed to a selective sweep event, provided that we could compare the actual distribution of polymorphism to its expected value under the neutral mutation-drift equilibrium.

To answer this question, we studied the *In(2L)t* polymorphic inversion. The underlying rationale was that chromosome inversions can reveal selective sweeps because they strongly inhibit recombination between chromosomal types (see Ashburner 1989). They thus divide up a sample of chromosomes into two partially isolated subpopulations exchanging genetic variation through rare recombination events in inversion heterozygotes. A favorable mutation can appear and go to fixation in one of them. The other chromosomal type remains unaffected until the favored allele recombines into it. In the meantime, a strong contrast in variation pattern differentiates the two chromosomal types. This contrast is more easily detected in highly recombining regions, which have more preexisting segregating variation. This contrast can be expected to occur around inversion breakpoints (internally or externally) rather than in the middle of long inversions, where double crossover (Krimbas and Powell 1992) and gene conversion (Chovnik *et al.* 1977) can take place.

We chose to carry out this study on the *Suppressor of Hairless* gene (*Su(H)*) after surveying length variation in several trinucleotide microsatellites from chromosome

Corresponding author: Michel Veuille, EPHE and Laboratoire d'Ecologie UMR 7625, Cc 237, Université Paris 6, 7 quai Saint-Bernard, 75005 Paris, France. E-mail: mveuille@snv.jussieu.fr

2 (Michalakis and Veuille 1996) in natural populations showing inversion polymorphism (Bénassi *et al.* 1993; Veuille *et al.* 1998). We found highly significant linkage disequilibrium between a microsatellite located in the *Su(H)* coding region and the *In(2L)t* chromosome polymorphism in a west African population. The *Su(H)* gene (Furukawa *et al.* 1991; Schweisguth and Posakony 1992) is a transcription factor that is involved in the Notch signalling pathway (Artavanis-Tsakonas *et al.* 1995). It presents an OPA-repeat, consisting of tandemly arranged glutamine codons (CAG or CAA). The relative positioning of *Su(*H) and *In(2L)t* is known. On the *Drosophila melanogaster* physical map, this gene maps to position 35B9-10 (Schweisguth 1995), which is between the proximal breakpoint of *In(2L)t* (in 34A8–9) and the centromere (in 40F). The *Su(H)* gene is not very far from the *Adh* locus (in 35B2), which shows a reduced recombination rate (5×10^{-4} recombination events per generation in females) with *In(2L)t* in heterokaryotypic females (Malpica *et al.* 1987). This suggested that, although distant from the inversion (about one-hundredth of the Drosophila genome), the *Su(H)* gene could be maintained in linkage disequilibrium with the inversion for many generations. The difference in microsatellite allele frequency could indicate a selective sweep had occurred in one or another chromosomal arrangement. We therefore recorded sequence variation at this locus in this population, and found it to agree with the hypothesis of a recent selective sweep.

MATERIALS AND METHODS

A random sample of 85 isochromosomal lines for chromosome 2 was established by Benassi *et al.* (1993) from a natural population of *D. melanogaster* from the Ivory Coast (Lamto Ecological Station). In this population, the frequency of the *In(2L)t* inversion (0.62, SD = 0.05) is particularly high. We used a random sample of 47 lines for the microsatellite survey and a random subsample of 20 lines for the sequence survey. The proportion of *In(2L)t* and *standard* chromosomes in the two samples (25 *vs.* 22 for the microsatellite survey, respectively, and 9 *vs.* 11 for the sequence survey) did not significantly depart from their initial frequencies. The *D. simulans* sequence was obtained from a line bearing multiple recessive markers for chromosome 2 and therefore was expected to be largely homozygous for this chromosome at the molecular level, as previously observed for the *Fbp2* gene (Benassi *et al.* 1999).

Su(H) microsatellite variation was studied according to Michalakis and Veuille (1996). To maximize genetic information in the sequence survey, we studied a region of the gene overlapping a long intron. We amplified a fragment including a 712-bp intron and 314 coding sites plus a varying number of indels and microsatellite repeats. Genomic DNA was amplified through standard PCR conditions between coordinates 142 and 1198 of the published *Su(H)* transcription unit (Furukawa *et al.* 1991), using primers AACCGTAGTTCGTAGG CAAT and GAACGCAGGCGATTGAACAG. They were sequenced in both orientations with these and four intervening primers (AGGGGTGAGCGGTTGGGGGATT, CTTCCGAA CAGATAAATGCA, TTGCTCAATTTGCGGGC, and GAAA GAAAATCTGAA). The reference sequence is 1056 bp long, while our data extend over 1196 bp to account for indels. The sequences are available from the GenBank database under accession nos. AF088255–AF088275.

Sequences were manually aligned. Their phylogeny was analyzed using MEGA (Kumar *et al.* 1993). Descriptive statistics were derived using the DNAsp2 program (Rozas and Rozas 1997). The effective numbers of synonymous and nonsynonymous sites were calculated using Nei and Gojobori's (1986) method.

The *K*-test and *H*-test of haplotype diversity were run according to Depaulis and Veuille (1998). These coalescent-based simulation tests assess whether the association of *S* polymorphic sites into *K* haplotypes in a sample of *n* sequences is in neutral mutation-drift equilibrium or not. The *K*-test is based on the observed haplotype number, and the *H*-test on the sample haplotype diversity. These tests were used under two models. The first is a no-recombination model, for which confidence intervals are available (Depaulis and Veuille 1998). This model is conservative when the number of observed haplotypes is lower than expected, because recombination increases its expectation. The second model takes into account the recombination rate at the studied locus. This rate can be estimated according to Hudson's (1987) method, which derives a value of Nr (where N is the diploid population size and r is the recombination rate per generation per nucleotide) from the variance of pairwise differences in natural populations under the assumption of neutrality. This estimate has a large variance. Hudson *et al.* (1994) compare this estimate to the value ($r = 10^{-8}$/bp, yielding $Nr = 10^{-2}$, assuming $N = 10^6$) directly derived from genetic experiments for highly recombining regions (Chovnik *et al.* 1977). We obtained a lower value ($Nr = 5.45\ 10^{-3}$) for *Su(H)*, which is probably an underestimate.

RESULTS

Association between the inversion and the *Su(H)* microsatellite: Four size variants differing by one repeat unit (one codon) were observed at the *Su(H)* microsatellite (Table 1). The alleles were 251, 254, 257, and 260 nucleotides long and were named after their size. The 251 allele was present on 24 of the 25 inverted chromosomes, but only on 4 of the 22 *standard* chromosomes. The *standard* arrangements were more variable, with four alleles and a larger sample heterozygosity ($H = 0.615$) than the other class ($H = 0.076$). The linkage disequilibrium between the microsatellite and the inver-

TABLE 1

Linkage disequilibrium between the *Su(H)* microsatellite and the *In(2L)t* inversion

Chromosome	Total	*In(2L)t*	*Standard*
N	47	25	22
Allele length in base pairs			
251	28	24	4
254	5	0	5
257	2	1	1
260	12	0	12

Fisher's exact test for multiple classes (Raymond and Rousset 1995), $P < 10^{-4}$.

```
                                                                111
sequence      233333444444444444444455555555556666666666667777777789011  micro-
coordinates  1045669123356677777889125777788001122778990011223209 01  satellite
             799315577796491246947224406783489192448018030705687721  length
             **
reference    GGCGTGCGATCCCATTCAAGCATTGTTCGCCAAAATCTGAGATTTGTCGGGCTG

In-L3        ...C..T......T....................G.............A..GCA   251
In-L4        ...C..T......T....................G.............A..GCA   251
In-L14       ...C..T......T....................G.............A..GCA   251
In-L18       ...C..T......T....................G.............A..GCA   251
In-L21       ...C..T......T....................G.............A..GCA   251
In-L31       ...C..T......T....................G.............A..GCA   251
In-L101      ...C..T......T....................G.............A..GCA   251
In-L12       ........G.ATT.A.-GT...AGAAAGAGTGGG.A.AT...........C...   251
In-L19       ........G.ATT.A.-GT...AGAAAGAGTGGG.A.AT...........C...   251
St-L13       .....T.TG.ATT.A.-GT.A.AGA.AGAG.GGG.A.AT...........C...   254
St-L111      .A......G.ATT.A.-GT...AGA.AGAG.GGG.A.AT...........C...   251
St-L26       A.T......A.....A-..A.G.........-....AA...........C....   260
St-L27       A.T......A.....A-..A.G.........-....AA...........C....   260
St-L28       A.T......A.....A-..A.G.........-....AA...........C....   260
St-L106      A.T......A.....A-..A.G.........-....AA...........C....   260
St-L120      A.T......A.....A-..A.G.........-....AA...........C....   251
St-L5        A.T......A.....A-..A.G.........-....AA...........C....   260
St-L124      ....G........T....................G....C..........C...   257
St-L7        ...C..T......T....................G...---CCC-----.C...   254
St-L127      ...C..T......T....................G...---CCC-----.C...   254
D. simulans  ..G.G.T.G..TT..-AG.AGG..-....GTGG..A.C..A...AACA..C.-.
```

Figure 1.—Alignment of nucleotide changes at *Su(H)* in 20 *D. melanogaster* and one *D. simulans* sequences. The names of *In(2L)t* lines are in boldface; the reference sequence is a consensus. dot, same as in the consensus; dash, indel; star, synonymous change. The last column indicates the length of the fragment amplified for studying the microsatellite; repeats are multiples of 3 bp.

sion was assessed using Fisher's exact test for multiple classes (Raymond and Rousset 1995) and was found to be highly significant ($P < 10^{-4}$).

Sequence polymorphism: An alignment of polymorphisms is shown in Figure 1. We found nucleotide polymorphism at 47 sites, of which 17 were polymorphic only in *standard*, 6 only in *In(2L)t*, and the remaining 24 in both arrangements. These polymorphisms involved two changes out of 58.67 effective synonymous sites ($\pi = 0.00924$), no change out of 241.33 effective nonsynonymous sites, 42 substitutions out of 712 intron sites ($\pi = 0.0206$), and three intronic positions where both indels and substitutions occurred. The level of nucleotide variation, as estimated by π (Nei 1987) and θ (Watterson 1975) is shown in Table 2. These values were within the range of values observed for other genes located in highly recombining regions in this species, both for synonymous and noncoding variation (Charlesworth *et al.* 1995; Moriyama and Powell 1996). The level of variation did not depart from selective neutrality using Tajima's (1989) *D* test, Fu and Li's (1993) *D* tests, McDonald and Kreitman's (1991) test, or the HKA test (Hudson *et al.* 1987) against 5′-*Adh* variation, with *D. simulans* contributing as an outgroup (Table 3).

Linkage disequilibrium between nucleotide polymorphisms: Although levels of variation were thus unremarkable, nucleotide polymorphism was clumped into a small number of haplotypes, as shown in Figure 1. Only two haplotypes were found in *In(2L)t* and five in *standard*. The two arrangements shared no haplotype. Two haplotypes, one in the inversion and one in standard, made up two-thirds of the sample (13 out of 20 sequences). Linkage disequilibrium was studied by calculating the *P* value of Fisher's exact test in all pairwise associations of substitution polymorphisms. Of 741 possible pairwise combinations between informative sites, 290 were significant at the 0.05 level, of which 182 were significant at the 0.001 level. Bonferroni's correction was used to correct for multiple testing, and 35 tests remained significant at the 0.05 level. To our knowledge, this very high proportion by far exceeds the results of any similar study carried out on this species. The large number of tests that are significant at the 0.001 level results from the fact that an excess of chromosomes belonging to the same haplotypes repeatedly give the same result along the sequence. These observations rule out the null hypothesis of random association. They are in agreement with the structuring of basic nucleotide variation into haplotypes.

TABLE 2

Molecular diversity at *Su(H)*

Chromosomes	*n*	π	*SE*(π)	θ	*SE* (θ)	*C.I.* (θ)
Total	20	0.0136	0.0064	0.0111	0.0040	[0.0055–0.0244]
In(2L)t	9	0.0097	0.0049	0.0092	0.0041	[0.0039–0.0272]
Standard	11	0.0128	0.0062	0.0119	0.0050	[0.0054–0.0317]
Synonymous	20	0.0092	0.0019	0.0096	0.0072	[0.0011–0.0334]
Noncoding	20	0.0206	0.0027	0.0166	0.0060	[0.0082–0.0367]

π was calculated according to Nei (1987) and θ according to Watterson (1975); standard errors (SE) were calculated on the basis of the total variance, including the stochastic and the sampling variance according to Nei (1987); confidence intervals (C.I.) were calculated according to Kreitman and Hudson (1991).

TABLE 3

Neutrality tests in *Su(H)*

Sample	Tajima's *D*	Fu and Li's *D*	McDonald-Kreitman[a]	HKA[b]
Total	0.92	1.37	0.476	0.55
Inverted	0.28	1.93[c]	0.587	0.22
Standard	0.36	1.38	0.519	0.64

[a] Probability of Fisher's exact test.
[b] Probability using the *5′Adh* dataset (Kreitman and Hudson 1991) as a reference.
[c] $P < 0.05$.

Haplotype tests: We tested the probability of observing $k \leq 7$ haplotypes and a haplotype diversity of $H = 0.76$, given a sample of $n = 20$ sequences showing $S = 44$ diallelic polymorphisms using the *K*-test and the *H*-test (Depaulis and Veuille 1998). Results are shown in Table 4. The tests were significant in all cases. The most intuitive test is the *K*-test, which is based on the observed number of haplotypes. Its significance in the no-recombination test ($P < 0.011$) is remarkable because this test is conservative. A conservative estimate of the recombination rate was used in the recombination test, which was highly significant ($P < 10^{-4}$). These tests assess whether or not the haplotypes originated under a neutral coalescent process. A lower than expected number of haplotypes indicates that an event of reduced variation has recently impoverished the haplotype diversity of the gene. The fact that the haplotype tests are significant even under a no-recombination model further means that this conclusion is true irrespective of the inhibitory effect of the inversion on recombination.

Phylogenetic analysis of the haplotypes: To illustrate the structuring of variation, we carried out a phylogenetic analysis of the haplotypes. This approach was validated by the fact that the minimum number of recombination events in the sample, as estimated after Hudson and Kaplan's (1985) method, was only one. A neighbor-joining tree is shown in Figure 2. The phylogenetic analysis provided the same topology under the unweighted pair-group method using arithmetic averages (UPGMA) and maximum parsimony methods (data not shown) and was supported by high bootstrap values. The *In(2L)t* and *standard* chromosomes do not represent completely isolated lines. It was mentioned earlier that the two types of chromosomes shared 24 out of 47 polymorphic sites, suggesting a substantial genetic exchange of material between the two arrangements in the past. The phylogeny comprises a small cluster of four related sequences, consisting of two inverted and two *standard* haplotypes. This cluster substantially diverges from the other sequences, which make up a larger cluster. The latter involves the two major haplotypes (one inverted and one standard) and several intermediates belonging to standard. The only recombination event substantiated by the four-gamete rule (Hudson and Kaplan 1985) differentiates the two clusters, as is apparent in the alignment (Figure 1). There is no fixed difference between *In(2L)t* and *standard*.

Neutrality tests within each chromosomal arrangement: We recorded linkage disequilibrium within each chromosomal arrangement. Linkage disequilibrium between the 31 polymorphisms from the two inverted haplotypes was significant ($P = 0.028$, Fisher's exact test). Among the 34 polymorphic sites from standard chromosomes, 185 of the 561 comparisons were significant ($P < 0.05$). They corresponded to three of the comparisons between the five haplotypes. This confirms the clustering of variation into a few combinations, as was already apparent from Figure 1. Neutrality tests were nonsignificant, except for Fu and Li's test, which was marginally significant among inversions (Table 3), meaning

TABLE 4

Significance of haplotype tests in *Su*(*H*)

Recombination model	*K*-test on the number of haplotypes[a]	*H*-test on haplotype diversity[a]	Haplotype test[b]
Pooled data ($n = 20$, $S = 44$, $K = 7$, $H = 0.760$)			
No recombination	$P = 0.0126$	$P = 0.0170$	$P = 0.0756$
With recombination[c]	$P < 0.0001$	$P = 0.0004$	$P = 0.0062$
Inversion ($n = 9$, $S = 31$, $K = 2$, $H = 0.346$)			
No recombination	$P < 0.0001$	$P = 0.0001$	$P = 0.0005$
With recombination[c]	$P < 0.0001$	$P = 0.0001$	$P = 0.0001$
Standard ($n = 11$, $S = 39$, $K = 5$, $H = 0.645$)			
No recombination	$P = 0.0199$	$P = 0.0058$	$P = 0.0097$
With recombination[c]	$P = 0.0002$	$P = 0.0001$	$P = 0.0002$

K, number of haplotypes; *H*, sample haplotype diversity calculated as $H = 1 - \Sigma_{i=1}^{k} p_i^2$, where p_i is the frequency of the *i*th allele.
[a] Tests based on 10,000 simulations as in Depaulis and Veuille (1998).
[b] As in Hudson *et al.* (1994).
[c] Recombination rate, $Nr = 5.45 \times 10^{-3}$.

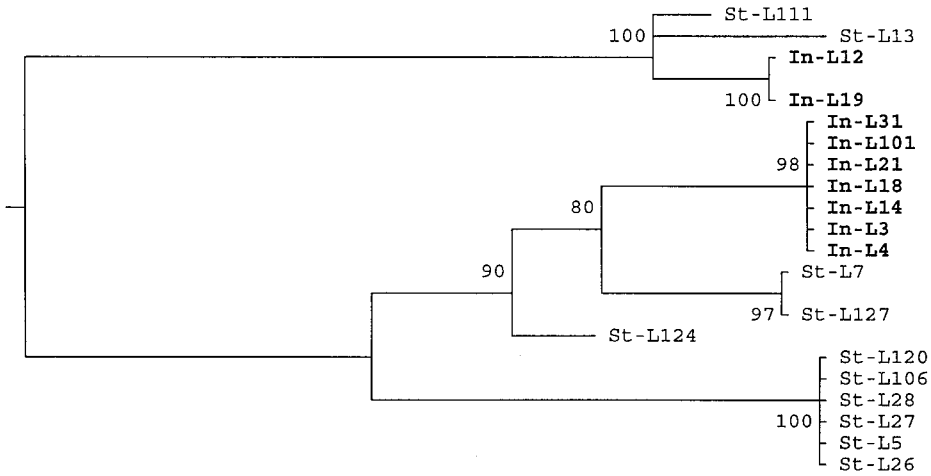

Figure 2.—Neighbor-joining tree of the 20 *Su(H)* sequences, showing bootstrap values (percentage over 500 replicates) for each node. *In(2L)t* sequences are in boldface; the *D. simulans* sequence contributes as an outgroup.

that singletons were not equally distributed in haplotypes from this subsample. However, because its nine sequences make up only two haplotypes, this result is not very informative. Haplotype tests deserve more interest, because they were significant on pooled data. If we apply these tests to each chromosome class separately, they are significant in all cases. This indicates that the significance was not due to a heterogeneity of the data caused by the inversion.

Molecular divergence between chromosomal arrangements: The genetic divergence between the two karyotypes can be assessed using a fixation index (Hudson *et al.* 1992), the value of which was $F_{st} = 0.252$ (significance as estimated by permutations, $P = 0.0170$). This means that one-quarter of the distance between arrangements is due to the structuration into chromosomes. This is a considerable level of differentiation, as can be emphasized by noting that this value is higher than that observed between *D. melanogaster* populations from different continents. For nine polymorphic genes [excluding *Su(H)*] located at distant positions on the same chromosome, the average F_{st} was 0.105 in a group of European and African populations (Michalakis and Veuille 1996). This value was also higher than that observed between the two chromosome arrangements in Lamto for another gene, *Adh*. Four-cutter restriction site polymorphism at this locus (Veuille *et al.* 1998) showed a relative divergence of $D_a = 0.63$ *vs.* $D_{xy} = 5.93$ using Nei's (1987) distances, giving a ratio $G_{st} = 0.106$, which is equivalent to Hudson *et al.*'s (1992) F_{st} (Charlesworth 1998).

DISCUSSION

Departure from neutrality in a highly polymorphic gene: The *Su(H)* gene shows a strong contrast between the distribution of polymorphisms at the nucleotide level and at the haplotype level. Nucleotide variation may be briefly described as a normal neutral polymorphism. It presents the composition expected for this species, both in the number of polymorphisms and in the proportion of rare and frequent variants. It therefore appears neutral under available neutrality tests. On the contrary, there is a drastic deficit of haplotypes. All tests lead to the conclusion that the number of haplotypes has recently been substantially reduced. A theoretical model by Barton (1998) shows that a sudden reduction of effective population size (through, *e.g.*, a selective sweep or a bottleneck) splits preexisting neutral variation at a given locus into distantly related families of closely-related lineages. Haplotype distribution at *Su(H)* seems a remarkable illustration of this (see Figures 1 and 2).

Comparing selective *vs.* demographic explanations: This effect could have been caused by a selective sweep or by a population bottleneck. The two processes have the same effect at the level of single genes, but the first applies to part of the genome, whereas the second applies to the whole of it. Molecular variation is known for other genes from the Lamto sample. Restriction site polymorphism has been observed in 85 chromosomes in Lamto for 2.4 kb of the *Adh* gene (Benassi *et al.* 1993). The observed haplotype diversity ($H = 0.936$) did not depart from that observed for populations from France ($H = 0.89$) and from Malawi ($H = 0.99$; Benassi and Veuille 1995). The genetic variability of these three populations has also been compared using a set of 10 polymorphic microsatellites spread over the whole length of chromosome 2 (Michalakis and Veuille 1996). The overall heterozygosity in Lamto ($H = 0.421$) was in the range observed for the other populations ($H = 0.321$–0.514).

Comparison with variation at the *Acp26Aa* and *Acp26Ab* loci (Aguadé 1998) is in agreement with a selective sweep at the *Su(H)* locus. These genes, like *Su(H)* and *Adh*, are located in a highly recombining region of the left arm of chromosome 2. They lie between positions 25D7 and 26A8-9, in the middle of the region covered by *In(2L)t*, and show no linkage disequilibrium with the inversion (Aguadé 1998). Sequence polymorphism at

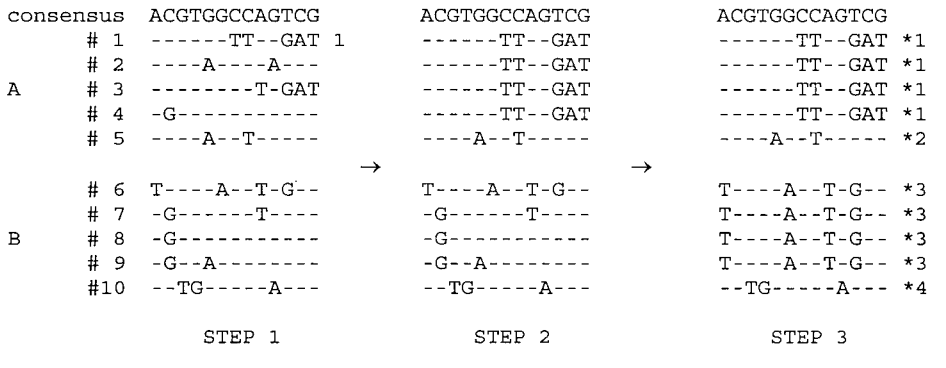

Figure 3.—Hypothetical scheme of selective sweep in *Su(H)*. Step 1: the *Su(H)* locus is in neutral mutation-drift equilibrium in two partially divergent chromosomal arrangements (A) inverted and (B) standard (or the letters could be reversed); a favored mutation at another locus appears in the same chromosome as haplotype 1. Step 2: a complete selective sweep at the selected locus occurs in (A) resulting in loss of variation at *Su(H)*, for the A subsample; the loss is partial, due to intra-arrangement recombination. Step 3: a recombination between chromosome arrangements imports the favored mutation into haplotype 6, and the selective sweep event continues in B. Former haplotypes are numbered on the left and surviving haplotypes on the right.

these loci has been recorded for 24 chromosomes drawn from Benassi *et al.*'s (1993) Lamto sample. Haplotype diversity was much higher than for *Su(H)*. Their number of haplotypes (S) and haplotype diversity (H) were above their expectation under a recombination model (Depaulis and Veuille 1998) and were thus opposite to the result observed for *Su(H)*. This remained true if considering only the 44 first polymorphic sites [the number found in *Su(H)*], where both *Acp26A* genes showed a number of haplotypes $S = 21$ and a sample haplotype diversity $H = 0.968$. We also computed the 0.05 limit intervals of S and H, using a recombination model (Depaulis and Veuille 1998) in combination with estimates of Nr obtained using Hudson's (1987) method. Variation at *Acp26A* genes appeared neutral (data not shown). They thus provide a negative control for *Su(H)*, allowing us to exclude the hypothesis of a population bottleneck in the Lamto sample.

Relation of the inversion to the selective sweep: Our conclusion is thus that the haplotype pattern of *Su(H)* results from a selective sweep that affected both chromosome arrangements at a nearby locus. It is remarkable that we do not need to consider the two chromosome arrangements separately to come to this conclusion. Linkage disequilibrium and haplotype tests yield significant results on pooled data, even though they are also significant for each of the chromosome arrangements. This suggests that the same selective sweep event affected both arrangements. An outline of this process is presented in Figure 3. In a first step, an advantageous allele arises by mutation at an unidentified locus "*U*" on chromosome 2. This new allele of *U* is linked to one of the *Su(H)* haplotypes and to one of the chromosome arrangements, *In(2L)t* or *standard*. If *Su(H)* is at a neutral mutation-drift equilibrium, most haplotypes will be different, as is observed in other *D. melanogaster* genes, and little linkage disequilibrium will be present between polymorphic sites. In a second step, the favored allele of *U* goes to fixation in the first chromosome arrangement. This causes a selective sweep at *Su(H)*, increasing the frequency of the first haplotype associated with the favored allele. Because the recombination rate is not inhibited within chromosome arrangements, this allele is soon linked to several other *Su(H)* haplotypes, which thus survive the selective sweep event. In a third step, the favored allele recombines into the other chromosomal arrangement and links to another *Su(H)* haplotype. The selective sweep process continues in this arrangement, albeit involving different haplotypes. The strong differentiation observed between haplotypes thus conforms to predictions of a hitchhiking model with recombination where different alleles are affected in different populations (Slatkin and Wiehe 1998).

An alternative hypothesis is that two selective sweep events, one on each chromosomal arrangement, occurred independently. This would cause a balanced polymorphism pattern between them. This explanation was put forward by Kirby and Stephan (1996) for small sequences of DNA, where the absence of recombination is due to the small genetic distance. It is known as the "traffic hypothesis." This explanation is probably less parsimonious than the former in the case of *Su(H)*.

This schema is based on simple hypotheses. In addition, an implicit observation is that selection did not sweep away the inversion polymorphism. The focus of this study is not to evaluate the individual contribution of genes to this phenomenon. We cannot say, from this study, to what extent individual genes can affect an inversion frequency. We can only rule out the hypothesis of a balanced polymorphism at *Su(H)*. The possible role of inversions in maintaining genetic polymorphism under balanced selection was put forward by Wright and Dobzhansky (1946) for *D. pseudoobscura* and was an influential model in the development of population genetics (Lewontin *et al.* 1981). Data obtained for another *D. melanogaster* inversion, *In(3L)Payne*, showed no significant departure from neutral equilibrium (Hasson and Eanes 1996) in spite of patterns sometimes suggesting selective sweeps. Similar patterns were observed in Drosophila species from the *obscura* group, which also present inversion polymorphisms (Rozas and Aguadé 1990, 1993; Babcock and Anderson 1996; Popadic *et al.* 1995). Recombination is not completely inhibited between *Su(H)* and *In(2L)t* in heterokaryotypic

females. According to Strobeck (1983), even a small recombination rate is sufficient to generate random associations between anciently coexisting polymorphic sites. An old balanced polymorphism could not therefore have caused the pattern observed in *Su(H)*, first because the two chromosomal arrangements are very divergent at this locus, and second because haplotypic variation is depleted within each arrangement.

Our study indicates a way to observe selective sweeps in genomes showing many inversion polymorphisms, as in *D. melanogaster* (Lemeunier and Aulard 1992). In the case of *In(2L)t*, the *Su(H)* locus is located outside the inversion, at the substantial distance of one chromosome division. The effect of inversions on variation can thus extend far from inversion breakpoints. Future research should tell whether the observations made in *Su(H)* are characteristic of only this gene or can be replicated for other genes associated with this inversion.

We thank Anne Turbé for contributing to sequencing as a DEUG student, Michèle Huet for technical assistance, and Matthew Cobb for comments on the manuscript. This work was supported by Centre National de la Recherche Scientifique, Université Pierre-et-Marie Curie, and École Normale Supérieure.

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., N. Miyashita and C. F. Langley, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

Artavanis-Tsakonas, S., K. Matsuno and M. E. Fortini, 1995 Notch signalling. Science **268:** 225–232.

Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Babcock, C. S., and W. W. Anderson, 1996 Molecular evolution of the Sex-Ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. Mol. Biol. Evol. **13:** 297–308.

Barton, N. H., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. **72:** 123–133.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. Nature **356:** 519–520.

Benassi, 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.

Benassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *FBP2* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

Benassi, 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.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15:** 538–543.

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

Charlesworth, D., B. Charlesworth and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141:** 1619–1632.

Chovnik, A., W. Gelbart and M. McCarron, 1977 Organization of the *Rosy* locus in *Drosophila melanogaster*. Cell **11:** 1–10.

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.

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

Furukawa, T., M. Kawaichi, N. Matsunami, H. Ryo and Y. Nishida, 1991 The *Drosophila* RBP-J_κ gene encodes the binding protein for the immunoglobulin J_κ recombination signal sequence. J. Biol. Chem. **266:** 23334–23340.

Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144:** 1565–1575.

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

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

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background mutations with recombination. Genetics **141:** 1605–1617.

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 a positive selection in the Superoxide Dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

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

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **114:** 635–645.

Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from pattern of polymorphism and divergence. Genetics **127:** 565–582.

Krimbas, C. B., and J. R. Powell, 1992 *Drosophila Inversion Polymorphism*. CRC Press, Cleveland.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA: Molecular Evolutionary Genetics Analysis version 1.01. The Pennsylvania State University, University Park, PA.

Lemeunier, F., and S. Aulard, 1992 Inversion polymorphism in *Drosophila melanogaster*, pp. 339–405 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Cleveland.

Lewontin, R. C., J. A. Moore, W. B. Provine and B. Wallace, 1981 *Dobzhansky's Genetics of Natural Populations*. Columbia University Press, New York.

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

Malpica, J. M., J. M. Vassallo, A. Frias and F. Fuente-Bol, 1987 On recombination among *In(2L)t*, α-*Gpdh* and *Adh* in *Drosophila melanogaster*. Genetics **115:** 141–142.

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

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. F. 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.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Popadic, A., D. Popadic and W. W. Anderson, 1995 Interchromosomal exchange of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12:** 938–943.

Raymond, M., and F. Rousset, 1995 GENEPOP version 1.2: a population genetics software for exact test and ecumenicism. J. Hered. **86:** 248–249.

Rozas, J., and M. Aguadé, 1990 Evidence of genetic exchange in

the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. Genetics **126:** 417–426.

Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. Proc. Natl. Acad. Sci. USA **90:** 8083–8087.

Rozas, J., and F. R. Rozas, 1997 DNAsp version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Schweisguth, F., 1995 *Suppressor of Hairless* is required for signal reception during lateral inhibition in the *Drosophila* pupal notum. Development **121:** 1–10.

Schweisguth, F., and J. W. Posakony, 1992 *Suppressor of Hairless*, the Drosophila homologue of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. Cell **69:** 1199–1212.

Slatkin, M., and T. Wiehe, 1998 Genetic hitchhiking in a subdivided population. Genet. Res. **71:** 155–160.

Stephan, W., and C. H. Langley, 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

Strobeck, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103:** 545–555.

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

Veuille, M., V. Benassi, 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.

Wright, S., and Th. Dobzhansky, 1946 Genetics of natural populations. XII. Experimental reproduction of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics **31:** 125–156.

Communicating editor: W. Stephan