Genet. Res., Camb. (2000), **76**, *pp.* 149–158. *With* 3 *figures.* Printed in the United Kingdom © 2000 Cambridge University Press

Selective sweep near the *In(2L)t* inversion breakpoint in an African population of *Drosophila melanogaster*

FRANTZ DEPAULIS*, LIONEL BRAZIER, SYLVAIN MOUSSET, ANNE TURBE AND MICHEL VEUILLE

Laboratoire d'Ecologie CNRS-UMR 7625, EPHE, Université Pierre-et-Marie-Curie, 7 quai St Bernard, case 237, 75252 Paris cedex 05, France

(*Received 22 July 1999 and in revised form 3 December 1999*)

Summary

Chromosomal inversions largely inhibit recombination and may be associated with selective forces, such as hitch-hiking effects: the effect of positive selection on linked loci. A West African population of *Drosophila melanogaster* showed a high frequency (0·61) of the *In(2L)t* inversion. Departure from neutrality statistically associated with the inversion polymorphism was previously recorded at *Su(H)*, a locus distant from the proximal breakpoint of the inversion. These results were consistent with hitch-hiking effects with recombination. The present sequence polymorphism survey involves a 1 kb fragment of the *Vha68-1* locus located closer to the proximal breakpoint of the inversion. It shows a significant deficit of polymorphism with respect to divergence when compared with other loci studied in the same population, thus suggesting selective effects. Only 11 polymorphic sites are present in a sample of 20 chromosomes and these sites present a significant excess of rare-frequency variants. The major haplotype shows an unexpectedly high frequency. Our estimate of the background selection effect is not sufficient to account for the observed reduction of polymorphism. Intraspecific variation is structured between inverted and *standard* chromosomes; there are no shared polymorphisms but also no fixed differences between them. This pattern, together with that found on other loci previously studied near this inversion breakpoint, suggests hitch-hiking effects enhanced by the inversion.

1. Introduction

A major issue in population genetics is distinguishing neutralist versus selectionist hypotheses. Demographic hypotheses (e.g. population subdivision, bottlenecks) can be distinguished from selective ones because the former should affect the whole genome to a similar extent whereas the latter are expected to be more restricted, and could show different effects depending on the distance between the marker considered and the selected locus. Selective events can be detected through their effect on linked neutral variation. Hitch-hiking and background selection are two models describing the effect of selected mutations on linked loci (Charlesworth *et al.*, 1993; Maynard-Smith & Haigh, 1974). Both models predict a reduction in the level of neutral polymorphism around selected loci. The lower the recombination rate, the more extended the effect of selected mutations and the larger the reduction of polymorphism. Background selection describes the effect of the recurrent elimination of deleterious mutations on linked neutral variation (Charlesworth *et al.*, 1993). The hitch-hiking effect involves the suppression of neutral variation near an advantageous mutation that has recently been fixed (Maynard-Smith & Haigh, 1974; Stephan *et al.*, 1992). Consistent with these two models, a reduction of the polymorphism level has been found for genes located in low recombining regions such as telomeres, pericentromeric regions and the fourth chromosome of the *Drosophila melanogaster* genome (Berry *et al.*,

* Corresponding author. Present address: Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. Tel: +44 (0)131 650 5513. Fax: +44 (0)131 650 6564. e-mail: frantz.depaulis@ed.ac.uk

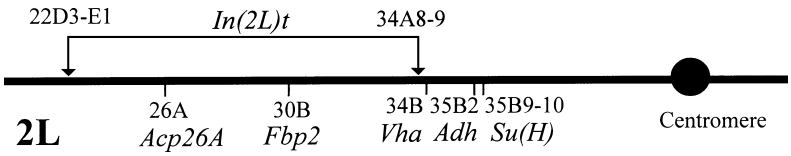

Fig. 1. Left arm of chromosome 2. The physical positions of the markers studied in the Ivory Coast population are indicated.

1991; Begun & Aquadro, 1992). No corresponding reduction is found at the divergence level (Begun & Aquadro, 1992). This pattern is not predicted under a neutral model, which predicts proportionality between polymorphism and divergence, but both background selection and hitch-hiking can account for observed data (Charlesworth *et al*., 1995; Charlesworth, 1996; Hudson & Kaplan, 1995). In principle it is possible to distinguish the two hypotheses. After a selective sweep (i.e. a hitch-hiking event without recombination between the observed marker and the selected site during the selective stage, hereafter 'hitch-hiking effects without recombination'), there is a period of recovery of polymorphism during which the distribution of variation should depart from a neutral distribution (Braverman *et al*., 1995). In contrast, background selection has little effect on the distribution of remaining variation (Charlesworth *et al*., 1995; Hudson & Kaplan, 1995, appendix A; Fu, 1997). However, in regions showing permanently low recombination rate, recurrent selective events remove most of the variability and make it difficult to distinguish the two hypotheses.

When an inversion arises in a region of high recombination, it strongly inhibits recombination in heterokaryotypes (individuals heterozygous for an inversion). Subsequent hitch-hiking and (or) background selection effects can differentially affect the pre-existing level of polymorphism at linked loci. Inversions are thus a potential tool for finding independent evidence for background selection or hitch-hiking effects, and possibly distinguishing between these hypotheses (Depaulis *et al*., 1999). Classic genetics studies of inversions have shown that recombination in heterokaryotypes is inhibited near breakpoints, both within and outside the inverted region (for a review see Krimbas & Powell, 1992). Chromosome inversion polymorphism has also been the focus of numerous population genetic studies in *Drosophila*, and several of these studies have claimed to have found evidence of selection (Krimbas & Powell, 1992). The general idea is that the fate of inversions depends both on their effect on recombination and on their genetic content. Balancing selection has generally been proposed as a mechanism maintaining inversion polymorphism. However, ancient balanced polymorphism cannot account for observed linkage disequilibrium with allozymes unless the recombination rate is very low ($N_e r < 1$; Strobeck, 1983). Under a balancing selection model, an increase in the level of polymorphism is expected around the selected locus (Hudson & Kaplan, 1988), whereas hitch-hiking and background selection models predict a reduction of polymorphism (Stephan *et al*., 1992; Charlesworth *et al*., 1993). In general, the study of nucleotide variation for markers closely linked to an inversion provides information on their evolution (i.e. monophyly, age, phylogenies) and on their differentiating effect on the neighbouring variation (Aguadé, 1988; Popadic *et al*., 1995; Babcock & Anderson, 1996; Hasson & Eanes, 1996; Veuille *et al*., 1998).

We used the cosmopolitan *In(2L)t* inversion of *D. melanogaster* as a model system of the evolution of markers linked to inversions. *In(2L)t* is located in the middle of the left arm of chromosome 2 (22D3-E1; 34A8-9, Fig. 1) and strongly suppresses recombination around its breakpoints in heterokaryotypes (Malpica *et al*., 1987). Previous studies have shown a high frequency (61·4%) of *In(2L)t* in a West African population and non-random association with restriction enzyme polymorphism at the *Adh* gene, which is located outside its proximal breakpoint (at 35B2; Bénassi *et al*., 1993). These studies have suggested that the *In(2L)t* frequency has recently been affected by a selective event, but this conclusion is based on restriction-site polymorphism and may be obscured by the selective history linked to the Fast-Slow polymorphism (Aguadé, 1988; Veuille *et al*., 1998). In addition, a sequence polymorphism study of the *Su(H)* locus, located in 35B9-10, proximal to the *Adh* locus, has indicated a strong departure from neutrality associated with *In(2L)t* polymorphism (Depaulis *et al*., 1999). This pattern is consistent with a hitch-hiking model involving recombination between an advantageous mutation and a partially linked locus (Barton, 1998; hereafter 'hitch-hiking effects with recombination'). This hypothesis implies that some recombination events have occurred during the selective stage between the inversion and the selected loci as well as between the selected locus and *Su(H)*. These recombination events allow the maintenance of some pre-existing neutral polymorphism at both loci (the inversion and *Su(H)*), but were not frequent enough to recover the expected number of haplotypes under a neutral model (Depaulis *et al*., 1999). Under the hypothesis of hitch-hiking effects amplified by the

inversion, markers located closer to inversion breakpoints are good candidates for finding hitch-hiking effects of larger magnitude. The *Vha68-1* (hereafter *Vha*) gene codes for the A subunit of a vacuolar ATPase (68 kDa; Guo *et al*., 1996). This locus maps at 34B, outside *In(2L)t* and close to the proximal breakpoint. The estimate of the distance between these two positions is 121–326 kb (after Sorsa, 1988). This locus is appropriate for testing the hypothesis of hitch-hiking effects amplified by the presence of nearby inversion.

In the present study, a 1 kb fragment of the *Vha* locus was sequenced in 20 chromosomes randomly sampled from the West African population of *D. melanogaster* and from a line of *D. simulans* used as a reference outgroup. This locus presents a significant deficit of polymorphism with respect to divergence when compared with other loci from the same population. The frequency spectrum of mutations showed a significant excess of rare polymorphisms and an unexpectedly frequent major haplotype. None of the polymorphisms was shared or fixed between the two chromosomal types. This pattern, together with that found at loci more distant from the inversion, is suggestive of hitch-hiking effects amplified by the nearby inversion.

2. Material and methods

(i) *Chromosome sample*

The random sample of 20 isogenic second chromosomes from the Ivory Coast population of Lamto was the same as for the previous *Su(H)* study (Depaulis *et al*., 1999). They were extracted from freshly collected individuals and surveyed for inversion polymorphism as described previously (Bénassi *et al*., 1993; Bénassi & Veuille, 1995). Sequencing in *D. simulans* was carried out on a strain multiple recessive chromosome 2 markers, provided by Jerry Coyne.

(ii) *DNA amplification and sequencing*

A 1163 bp sequence of the *Vha* coding region was PCR amplified in *D. simulans* and *D. melanogaster* using the following primers (5′ to 3′): CTTGAGGAAATTCAAAGACG, CATAATGGACTCGGTGACGC. Amplified DNA was reamplified using internal primers (5′ to 3′): GAATATGGCCGTGTCTACGC, CAGGTCAGTTCGGGAAAGTC. Amplification was carried out following the supplier's instruction (Promega). Amplified DNA was sequenced using the Amersham PCR-Sequenase kit. The fragment was sequenced over 1062 nucleotides. One-third of it was intronic, thus putatively showing low constraints and providing a high level of information in terms of variability. According to the reference sequence (GenBank accession number DMU19742), the sequenced region includes the last 251 bp (position [235,485]) of the 262 bp intron 1 and 726 nucleotides corresponding to the 5′ part of exon 2 ([486–1211]). An additional 85 bp intron was found between positions 1181 and 1182 of the DMU 19742 sequence. Sequencing was carried out in both directions. Sequences are available from the GenBank database under accession numbers AF087489–AF087509. Sequences were visually aligned with no ambiguity.

(iii) *Data treatment*

Statistical analysis was performed using DNAsp3 (Rozas & Rozas, in press) and Fu's (1997) program. Nucleotide polymorphism was estimated using Tajima's (1989) π and Watterson's (1975) θ statistics. Departure from neutrality was tested using the HKA test (Hudson *et al*., 1987), Tajima's (1989) *D* test, Fu & Li's (1993) *D* test, and McDonald & Kreitman's (1991) test using Fisher's exact test. Differentiation between the two arrangements was estimated using Hudson *et al*.'s (1992) F_{st} and tested by performing permutations over individuals. Divergence was estimated using MEGA (Kumar *et al*., 1993), without correction for multiple hits.

3. Results

(i) *Divergence between* D. simulans *and* D. melanogaster

The *D. simulans Vha* sequence differs from the *D. melanogaster* consensus sequence by two insertions and three deletions (lengths between 1 and 19bp) in the first intron. A total of 58 fixed differences are found, either synonymous (29) or intronic (29). Divergence levels are 57·9 nucleotides per kilobase (nuc/kb) for total DNA, 150 nuc/kb for silent sites and 98·9 nuc/kb for intronic sites (Table 1).

(ii) *Sequence polymorphism in* D. melanogaster

No length polymorphisms are found. A total of 11 polymorphic sites are found (Fig. 2). Six occur in introns, four are synonymous and one is a conservative replacement (serine–threonine). Seven polymorphisms are unique. In the four remaining polymorphisms, the rare allele is at most present in three sequences out of 20. Estimates of nucleotide diversity are 1·52 nuc/kb for total DNA, 4·46 nuc/kb for silent sites and 2·03 nuc/kb for intronic sites. The π/D ratio between *D. melanogaster* and *D. simulans* (2·6%) for the total DNA is the lowest found so far among for regions

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

Table 1. *Evolutionary statistics at the* Vha *locus*†

	Sample size	Polymorphic sites	Nucleotide polymorphism		Divergence[a]	Tajima's D^b	Fu & Li's D^b	Fu's F_s^b	HHT^c	H^c	McDonald–Kreitman[d]
			Tajima	Watterson							
Total	20	11	1·52	2·92	57·9	−1·503*	−1·648†	−4·562**	0·025*	0·026*	0·149
Inverted	9	3	0·78	1·04	57·9	−0·825	−0·890	−1·417	0·257	0·260	0·033*
Standard	11	8	1·98	2·57	57·9	−0·853	−1·183	−1·419	0·186	0·180	/
Synonymous	20	4	4·46	6·02	150·0						
Intronic	20	6	2·03	5·01	98·9						

†$P < 10\%$; *$P < 5\%$; **$P < 1\%$. All polymorphism and divergence values are expressed per kilobase.
[a] Divergence between *D. melanogaster* and *D. simulans* calculated as the average number of differences.
[b] Significance assessed after Fu's (1997) program.
[c] *P* value of Hudson *et al.*'s (1994) haplotype test and Depaulis & Veuille's (1998) *H* test with $Nr = 0{\cdot}0035$ per base pair.
[d] *P* value from Fisher's exact test.

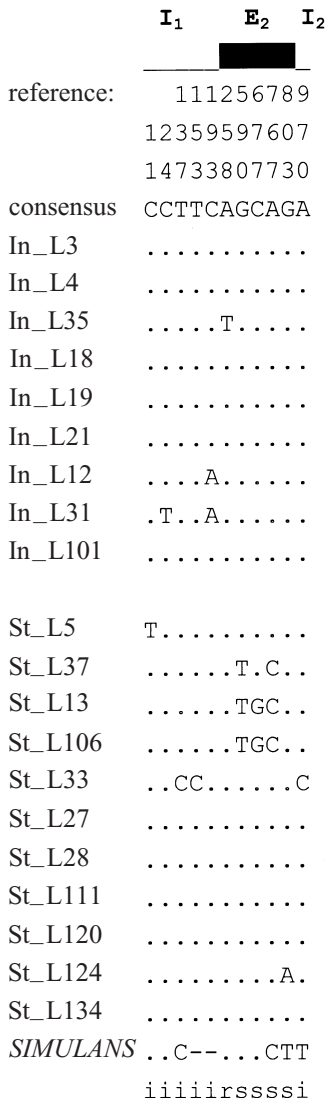

Fig. 2. Alignment of polymorphic sites in the *Vha* gene of *D. melanogaster* and divergence from *D. simulans*. The reference sequence is a consensus. Gaps are indicated by a dash. The bottom line indicates whether the polymorphism is intronic (i), replacement (r) or synonymous (s). The karyotype associated with each sequence is given by *St* (*Standard*) or *In* (inverted). I_i and E_i indicate intron and exon number respectively.

with substantial recombination rates ($N_e r > 10^{-3}$/bp, see below; after Moriyama & Powell, 1996).

(iii) *Recombination*

Given the low number of polymorphisms present in the data and given that their distribution departs from neutrality (see below), we prefer to use a recombination rate estimate that is independent of these data. A rough estimate can be obtained from the relation between genetic and physical distances (Adjusted Coefficient of Exchange). Disregarding the inversion, the recombination rate shows a middle range value (Adjusted Coefficient of Exchange =

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

0·01314; Kindahl, 1994). This leads to an estimate of the recombination parameter of $r = 7{\cdot}00\ 10^{-9}$ per base pair per generation (Sorsa, 1988). This value can be divided by 2 to obtain the approximate mean recombination rate in the population ($r = 3{\cdot}50\ 10^{-9}$), assuming that the contribution of heterokaryotypes to recombination is negligible and assuming Hardy–Weinberg proportions (in which case the expected frequency of heterokaryotypes is 0·48). According to Malpica *et al*. (1987), the recombination rate between the marker *b* close to the inversion and *Adh* gene located further away (at 35B2) is reduced in heterokaryotypes by roughly a factor 42 compared with homokaryotypes. The previous estimates consider the mean recombination rate in the population, which is relevant for an equilibrium model such as a neutral or a background selection model. However, an important parameter in this study is the recombination rate between the inversion and *Vha* in heterokaryotypes (hereafter $N_e r_b$). The recombination inhibitor effect of the inversion probably increases closer to the inversion tip. An upper bound of this recombination rate is provided by that found between the *Adh* locus and the marker *b* (Malpica *et al*., 1987). This leads to an estimated range for $N_e r_b$ between 9·4 and 48·5, assuming an effective population size of 10^6 and taking into account the imprecise distance estimate between *Vha* and the proximal breakpoint (see above; after Malpica *et al*., 1987; Sorsa, 1988; Kindahl, 1994).

(iv) *Association with the inversion*

Nine chromosomes from our sample show the *In(2L)t* rearrangement (see top of Fig. 2). The two arrangements share no polymorphisms (maximal linkage disequilibrium with the inversion), but there are no fixed differences between them suggesting a recent age of the inversion. Three polymorphisms occur in *In(2L)t* and eight in the *standard* chromosomes. Accordingly, the estimated level of polymorphism is lower in the inverted subsample than in *standard* (Table 1). A summary statistic describing the association of *Vha* with the inversion is the F_{st} between the two arrangements. The F_{st} value is 0·095 and is close to significance ($P = 0{\cdot}085$) in a permutation test. The two subsamples are thus considered both separately and pooled in further analyses.

(v) *Tests of selective neutrality of the frequency spectrum of mutations*

Tajima's *D* and Fu & Li's *D* statistics are negative, and Fu's F_s value is large, revealing an excess of rare and unique polymorphisms. Tajima's *D* is significant and Fu's F_s is highly significant. The first test is known to lack power, especially with low polymorphism values (Simonsen *et al*., 1995). Fu's F_s is not conservative against recombination, and therefore we used haplotype tests based on coalescent simulations, with the recombination rate estimate described above. The *H* test (Depaulis & Veuille, 1998) and Hudson *et al*.'s (1994) haplotype test are significant (haplotype diversity lower than expected and major haplotype more frequent than expected). These significant results are remarkable given the little information available, and suggest strong selective effects. The McDonald-Kreitman (1991) test is non-significant on the whole sample (deviating towards an excess of replacement versus synonymous polymorphism). Taking each chromosome arrangement separately, this test is significant on the inverted subsample (deviation in the same direction). It seems unlikely that the subsampling would bias the results of this test as it affects synonymous and non-synonymous sites to the same extent.

(vi) *Comparison with other loci*

Three loci were previously studied in the same population: the *Acp26A* (Aguadé, 1998) and *Fbp2* (Bénassi *et al*., 1999) genes are located within the inversion (26A and 30B, respectively), far from breakpoints, whereas *Su(H)* is located proximally outside *In(2L)t*, closer to the inversion breakpoint than the two previous loci, but further away than *Vha* (Fig. 1). The *Su(H)* study has shown strong departure from neutrality associated with the inversion (Depaulis *et al*., 1999). The other two loci show no significant association with the inversion polymorphism. The *Fbp2* locus shows departure from neutrality, but this is limited to one region of the locus. No significant linkage disequilibrium is found between any pairwise comparison of the four loci studied in the Ivory Coast population. An HKA test (Hudson *et al*., 1987) performed between *Vha* and each of the three genes studied in this population gives significant *P* values in all pairwise comparisons involving *Vha* (Table 2), considering either the whole sample or each chromosomal type. The deviation is in all cases towards a deficit of polymorphism at the *Vha* locus. It seems unlikely, then, that the other genes present an excess of polymorphism, especially considering that two of them show evidence of hitch-hiking effects with recombination (Bénassi *et al*., 1999; Depaulis *et al*., 1999), and that none of the other pairwise HKA tests between these genes are significant (results not shown). This suggests a deficit of polymorphism at the *Vha* locus.

Two models can account for significant HKA tests: hitch-hiking and background selection. The magnitude of the hitch-hiking effect is difficult to evaluate and largely speculative as it depends on many

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

Table 2. *HKA neutrality tests between* Vha *and other loci*[a]

Vha	*Su(H)*	*Fbp2*	*Acp26Ab*[b]
Total	5·218*	5·478*	8·981**
Inverted	5·214*	4·767*	7·969**
Standard	4·407*	4·925*	8·336**

[a] χ^2 values: $*P < 5\%$; $**P < 1\%$. All inverted and *standard* comparisons were applied within karyotypes for the two genes. All deviations are in the direction of a deficit of polymorphism with respect to divergence at the *Vha* locus.
[b] Analysis was performed on *Acp26Ab* since previous study has found evidence of directional selection acting on *Acp26Aa* (Aguadé, 1998).

unknown parameters: the selection coefficient, the age of the advantageous mutation and the genetic distance between the (unknown) selected locus and *Vha*. This effect can thus take any value between 0 and 100%. In contrast, background selection allows predictions using current estimates of recombination and deleterious mutation rates (Hudson & Kaplan, 1995). We have no *a priori* estimate of the population neutral mutation parameter $\theta = 4N_e u$, and this might largely depend on the locus under consideration, its neutral mutation rate and the studied population. But we can make predictions about the relative effect of background selection on different loci studied in the same population. Background selection is an equilibrium model, which depends on the mean recombination rate in the population ($3{\cdot}5 \times 10^{-9}$ per base pair per generation): under this model the inversion is subject only to genetic drift and the frequency does not change rapidly. Considering two loci with neutral mutation rates μ_1 and μ_2, and local recombination rates r_1 and r_2. Hudson and Kaplan's (1995) equation 9 predicts a reduction of the diversity value at the first locus:

$$\frac{\hat{\pi}_1}{4N_e\mu_1} = \exp\left[\frac{-u}{r_1}\right], \tag{1}$$

with u the deleterious mutation rate per unit of length, and similarly for locus 2. We use Tajima's estimate rather than that of Watterson as the former should be less affected when a departure from neutrality is detected in the frequency spectrum of mutations. The ratio of neutral mutation rates at locus 1 and 2 can be estimated by their ratio of divergence with a close species. Thus an estimate of expected diversity at locus 1 is:

$$\hat{\pi}_1 = \pi_2 \frac{D_1}{D_2} \exp\left[-u\left(\frac{1}{r_1} - \frac{1}{r_2}\right)\right], \tag{2}$$

where D's are divergence values and where subscripts indicate the genes considered. This equation was applied using *Vha* as locus 1 and *Acp26Ab, Fbp2* or *Su(H)* as locus 2, with the corresponding diversities, divergences and recombination rates derived from genetic and physical maps, and a deleterious mutation rate of 1 per genome (i.e. $5{\cdot}88 \times 10^{-9}$ per base pair), which may be overestimated (Keightley, 1996). This predicts a variation level at least 2 times larger than the observed values (4·08, 3·81 and 3·13 per kilobase respectively, compared with 1·52). Conversely, reversing (2) to estimate the recombination rate necessary for the background selection to account for the observed *Vha* diversity value provides $2{\cdot}20 \times 10^{-9}$, $2{\cdot}26 \times 10^{-9}$ and $2{\cdot}44 \times 10^{-9}$ as opposed to the estimated value of 3.50×10^{-9} for the recombination per base pair per generation. The recombination rate estimates used can be imprecise, but we control for any possible source of bias since they were derived from the same method for the three loci and since the different reference loci give similar results. Moreover, *Acp26Ab* is our only neutral-like reference.

A rough estimate of the population neutral mutation parameter $\theta = 4N_e\mu$ at the *Vha* locus is:

$$\hat{\theta}_1 = \theta_2 \frac{D_1}{D_2}, \tag{3}$$

where the D's are divergence values and the subscripts indicate the genes considered. Using *Fbp2* and *Acp26Ab* as reference loci, this leads to estimates of the expected θ at the *Vha* locus of 13·94 and 15·71 per kilobase (respectively). Assuming a star phylogeny after a putative hitch-hiking event without recombination between *Vha* and the selected locus during the selective stage (this is realistic if the sweep is recent compared with the effective population size and *Vha* close to the selected site), the expected number of polymorphic sites in a sample of size n is

$$S = n\mu t, \tag{4}$$

where μ, the mutation rate per gene, and t, the age of the sweep, are expressed per generation. An underestimate of the age of the sweep expressed in units of N_e generations is thus

$$\tau = \frac{4S}{n\theta}. \tag{5}$$

This leads to estimates of the corresponding age of the putative sweep of 0·15 and 0·13 N_e generations (respectively) using the two previous estimates of θ. Assuming a Poisson distribution of mutation with time, and solving $P(S \leqslant 11) = 0{\cdot}05$ numerically for τ, provides an upper bound for τ: 0·24 and 0·22 N_e (respectively). Though the star phylogeny assumption is not conservative, this suggests a relatively recent putative sweep.

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

4. Discussion

Any attempt to explain the evolution of the *Vha* locus must account for several features of the data: (*a*) a significant deficit of polymorphism with respect to divergence, (*b*) a significant excess of rare polymorphisms in the frequency spectrum of mutations and a major haplotype unexpectedly frequent, and (*c*) non-random association with the inversion.

(i) *Unsatisfactory hypotheses*

We should consider first neutralist hypothesis. The significant HKA tests obtained when the *Vha* data set is compared with other loci indicate a departure from neutrality on at least one of the two genes involved in each comparison. This result does not favour a hypothesis involving a low neutral mutation rate at the *Vha* locus, as this should be apparent at the divergence level. Similarly, a bottleneck, as well as any hypothesis involving population dynamics, is unlikely since polymorphism data are derived from the same population. Such events should affect both loci to the same extent.

Balancing selection is frequently considered an explanation for the maintenance of inversion polymorphisms (Wright & Dobzhansky, 1946), but these models predict an equilibrium pattern opposite to our data: an excess of polymorphism with respect to divergence and an excess of intermediate-frequency polymorphisms.

According to our estimate of the diversity level under background selection, this effect does not seem sufficient to account for the observed reduction of polymorphism. Furthermore, background selection is unlikely to lead to a significant excess of rare polymorphisms in the frequency spectrum of mutations and haplotypes, especially with a low variability (Charlesworth *et al.*, 1995; Hudson & Kaplan, 1995; Fu, 1997).

(ii) *Hitch-hiking associated with the inversion*

In contrast, hitch-hiking can have a large effect on the frequency distribution of mutations in the population, and is a non-equilibrium model. Fixation of an advantageous haplotype leads to a decrease in the ratio of polymorphism to divergence (Stephan *et al.*, 1992). During the subsequent recovery of polymorphism, the genealogy of samples is characterized by a star genealogy and an excess of low-frequency variants. This leads to negative values of Tajima's *D* and Fu & Li's *D* and to large negative Fu's F_s statistics (Braverman *et al.*, 1995; Fu, 1997). All these expectations are observed at the *Vha* locus.

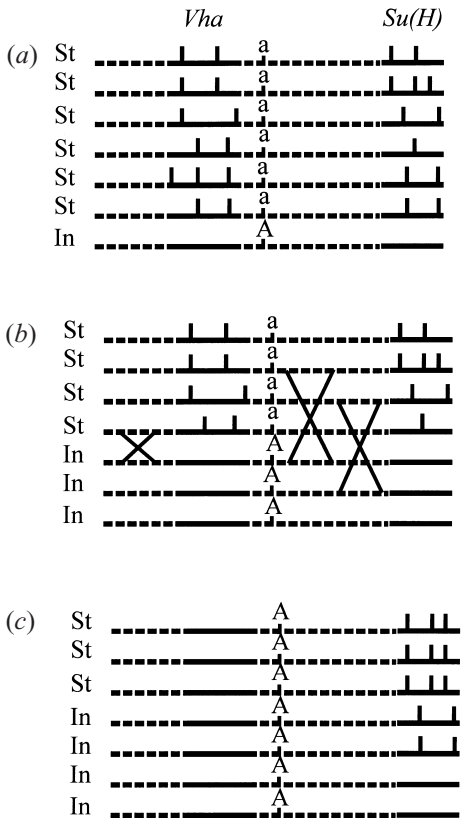

Fig. 3. Effect of a selective sweep on different loci located near an inversion. (*a*) Neutral polymorphism, an advantageous mutation (A) appears proximally and close to *Vha* on an inverted haplotype. (*b*) The inverted haplotype increases in frequency and recombinations occur between the inversion and *Vha*, and between the selected locus and *Su(H)*. (*c*) All haplotypes carrying the advantageous mutation increase in frequency until the advantageous mutation is fixed, but the inversion and *Su(H)* remain polymorphic, with strong haplotype structure. The mutations will then start to recover polymorphism at the *Vha* locus and, together with recombination, will alter the haplotype structure on *Su(H)* locus and the inversion breakpoint.

The effect of a selective sweep can be amplified by the inhibitory effect on recombination of the nearby inversion. The inversion itself could be the target of selection and have been recently driven by selection to an intermediate frequency, thus affecting linked genes on inverted haplotypes. However, this would not explain the significant HKA tests when applied to the *standard* subsample. We therefore propose an alternative hypothesis (Fig. 3).The inversion is initially at a low frequency and thus mostly heterozygous. If an advantageous mutation appears in an inverted chromosome, on a third locus strongly linked to *Vha* but partially linked to the inversion breakpoint, the advantageous mutation and an associated haplotype at the *Vha* locus will go to fixation. Meanwhile, the inversion will increase in frequency by hitch-hiking, though not to fixation because of recombination

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

events between the breakpoints and the region encompassing the selective site and *Vha*. Alternatively other unknown selective forces acting on the inversion polymorphism could prevent its fixation although there is no evidence regarding this hypothesis. During the increase in frequency of the inversion, recombination between inverted chromosomes also increases (due to the appearance of inverted homokaryotypes). However, these recombination events occur mainly between identical copies of the original advantageous haplotype and are thus not able to break the hitch-hiking effect. The effective recombination rate affecting the magnitude of the hitch-hiking effect would be that occurring in heterokaryotypes ($N_e r_b$). This recombination rate is about 42 times lower than that in homokaryotypes (both inverted and *standard* ones). This could lead to an extended hitch-hiking effect over a large portion of the chromosomal arm. This low recombination rate and the recent age of the increase in frequency of the inversion are in agreement with the absence of both fixed difference and shared polymorphisms between the two chromosomal types and with the significant association of the inversion with *Su(H)*. Of course, advantageous mutations can also appear on the most common chromosomal type (i.e. *standard* chromosomes). They are, however, unlikely to be detected as they do not extend the hitch-hiking effect. The significance of the McDonald–Kreitman test could be explained by the fact that even slightly deleterious mutations (such as conservative replacements) can hitch-hike with advantageous ones. Alternatively, purifying selection could be sufficient to explain this result and the excess of rare polymorphism. But this would predict an excess of polymorphism with respect to divergence in an HKA test, in contrast to the deviation observed. Finally, the absence of linkage disequilibrium between the studied loci could simply result from the high recombination level within each arrangement. Alternatively, the observed hitch-hiking effect could not be amplified by the inversion. In this case, the inversion could have appeared after and independently from the sweep observed on *Vha*. Then all selective effects observed at all studied loci would be due to independent events and would have been detected by chance. We consider this hypothesis as non-parsimonious, though the question of the number of hitch-hiking events associated with the inversion remains to be addressed.

(iii) *Generality of these conclusions*

Evidence for differentiation between chromosomal types for loci tightly linked to inversion breakpoints has been found in several instances, but is generally characterized by fixed differences between chromosomal types, suggesting an old chromosomal polymorphism (Rozas & Aguadé, 1993; Babcock & Anderson, 1996; Hasson & Eanes, 1996). Reduced variation was found in only one arrangement and little evidence of selection has been found (see, however, Babcock & Anderson, 1996). A large proportion of unique polymorphisms has, however, already been observed (Rozas & Aguadé, 1993).

The same hitch-hiking event could be involved for *Su(H)* and *Vha*, with a different outcome due to a different linkage with the selected site, thus showing the extended impact of a hitch-hiking effects occurring near an inversion (Fig. 3). Recently, Andolfatto *et al.* (1999) studied the breakpoints of this inversion, though in a different population. Their estimate of the age of the increase in frequency of the inversion is comparable to ours and is too low to be expected under a neutral model in all populations presenting a high frequency of the inversion. No evidence of geographical differentiation of inverted chromosomes was found. This suggests a recent selective spread of the inversion. They also found a haplotype structure similar to that found at the *Su(H)* locus, affecting both chromosomal types, with a shared haplotypic class between the two chromosomal types. We consider these results as evidence for the hitch-hiking effects with recombination suggested here at these two loci (*Su(H)* and the inversion). Evidence for a simultaneous sweep on several loci, possibly associated with an inversion, has already been found for the *Sod* region (Hudson *et al.*, 1997). The estimate of the age of the sweep is comparable: $t = 41\,000$ years, which is equivalent to $0{\cdot}2\ N_e$ (taking 5 generations per year and an effective size of 10^6). This study assumed a hitch-hiking event having not gone to fixation and could not exclude demographic hypotheses. The present case concerns a much wider region, but the recombination suppression effect of the inversion could largely enhance the hitch-hiking effect. Moreover, previous theoretical studies on the extent of the hitch-hiking effects have considered the effect on the level of polymorphism (Kaplan *et al.*, 1989). Hitch-hiking could show other effects on the distribution of polymorphisms, which might extend further away from the selected site. We do not see any reduction in polymorphism at the *Su(H)* locus (hence the significant HKA test with *Vha*), but we observe departure from neutrality in the distribution of haplotypes.

The present data, together with the data obtained on other loci close to the inversion breakpoint, do not favour background selection as a sufficient explanation. This does not, however, mean that the background selection effect has to be minimized in general: the present study constitutes a special case and not a general survey of selective effects on the whole genome. More data are needed to answer this question. One difference in our case compared with studies of polymorphism in low recombining regions

https://doi.org/10.1017/S0016672300004626 Published online by Cambridge University Press

is that information is not completely removed by recurrent selective events: we observe only one event leading to the high frequency of the inversion. Finally we suggest that hitch-hiking effects with recombination may be easier to detect as they maintain a large variability, though non-neutrally distributed (Barton, 1998). As an illustration, one can compare the polymorphism level and neutrality test results of *Vha* and *Su(H)* (Depaulis *et al*., 1999).

We thank Matthew Cobb, Nicolas Galtier, Jody Hey, Gilean McVean, Christina Muirhead and anonymous reviewers for significant contributions in improving this manuscript. This research was financially supported by the CNRS-UPMC-ENS-EPHE.

References

Aguadé, M. (1988). Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and non-inverted chromosomes. *Genetics* **119**, 135–140.

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.

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

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

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research* **72**, 123–133.

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

Bénassi, V. & Veuille, M. (1995). Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, West Africa and East Africa. *Genetical Research* **65**, 95–103.

Bénassi, V., Aulard, S., Mazeau, S. & Veuille, M. (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.

Bénassi, V., Depaulis, F., Meghlaoui, G. K. & Veuille, M. (1999). Partial sweeping of variation at the *Fbp2* locus, in a West African population of *Drosophila melanogaster* genome. *Molecular Biology and Evolution* **16**, 347–353.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **126**, 1111–1117.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitch-hiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* **140**, 783–796.

Charlesworth, B. (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical Research* **68**, 131–149.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious alleles on neutral molecular variation. *Genetics* **134**, 1289–1303.

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

Depaulis, F. & Veuille, M. (1998). Neutrality tests based on the distribution of haplotypes under an infinite site model. *Molecular Biology and Evolution* **15**, 1788–1790.

Depaulis, F., Brazier, L. & Veuille, M. (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.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

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

Guo, Y., Dow, J. A. T., Gillan, A., Kiss, I. & Kaiser, K. (1996). Molecular characterisation and inactivation of the 68kDa A-subunit of V-ATPase in *Drosophila*. In *37th Annual Drosophila Research Conference*, p. 91B.

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

Hudson, R. R. & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840.

Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

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

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

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hudson, R. R., Saez, A. A. & Ayala, F. J. (1997). DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. *Proceedings of the National Academy of Sciences of the USA* **94**, 7725–7729.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitch-hiking effect' revisited. *Genetics* **123**, 887–899.

Keightley, P. D. (1996). Nature of mutation load in *Drosophila*. *Genetics* **144**, 1993–1999.

Kindahl, E. C. (1994). Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. PhD thesis, Cornell University, Ithaca.

Krimbas, C. B. & Powell, J. R. (1992). *Drosphila Inversions Polymorphism*. Cleveland, OH: CRC Press.

Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular Evolutionary Genetic Analysis. The Pennsylvania State University.

Malpica, J. M., Vassalo, J. M., Frias, A. & Fuentes-bol, F. (1987). On recombination among *In(2L)t, alpha Gpdh* and *Adh* in *Drosophila melanogaster*. *Genetics* **133**, 141–142.

Maynard-Smith, J. & Haigh, J. (1974). The hitch-hiking effect of favourable gene. *Genetical Research* **23**, 23–35.

McDonald, J. H. & Kreitman, M. (1991). Adaptative protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654.

Moriyama, E. N. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.

Popadic, A., Popadic, D. & Anderson, W. (1995). Interchromosomal exchange of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. *Molecular Biology and Evolution* **12**, 938–943.

Rozas, J. & Aguadé, M. (1993). Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. *Proceedings of the National Academy of Sciences of the USA* **90**, 8083–8087.

Rozas, J. & Rozas, R. (2000). DNAsp version 3.0: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, in press.

Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

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

Stephan, W., Wiehe, T. H. E. & Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology* **41**, 237–254.

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., Bénassi, V., Aulard, S. & Depaulis, F. (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. *Theoretical Population Biology* **7**, 256–276.

Wright, S. & Dobzhanksy, T. (1946). Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics* **31**, 125–156.