The lack of recombination drives the fixation of transposable elements on the fourth chromosome of *Drosophila melanogaster*

CAROLINA BARTOLOMÉ* AND XULIO MASIDE
Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, UK

(*Received 4 November 2003 and in revised form 20 January 2004*)

Summary

In regions of suppressed recombination, where selection is expected to be less efficient in removing slightly deleterious mutations, transposable element (TE) insertions should be more likely to drift to higher frequencies, and even to reach fixation. In the absence of excision events, once a TE is fixed it cannot be eliminated from the population, and accumulation of elements thus should become an irreversible process. In the long term, this can drive the degeneration of large non-recombining fractions of the genomes. Chromosome 4 of *Drosophila melanogaster* has very low levels of recombination, if any, and this could be causing its degeneration. Here we report the results of a PCR-based analysis of the population frequencies of TE insertions in a sample from three African natural populations. We investigated 27 insertions from 12 TE families, located in regions of either suppressed or free recombination. Our results suggest that TE insertions tend to be fixed in the non-recombining regions, particularly on the fourth chromosome. We have also found that this involves all types of elements, and that fixed insertions are significantly shorter and more divergent from the canonical sequence than those segregating in the sample (28·1 % vs 86·3 % of the canonical length, and average nucleotide divergence $(D_{XY}) = 0·082$ vs $0·008$, respectively). Finally, DNA-based elements seem to show a greater tendency to reach fixation than retrotransposons. Implications of these findings for the population dynamics of TEs, and the evolutionary forces that shape the patterns of genetic variation in regions of reduced recombination, are discussed.

1. Introduction

The selfish DNA hypothesis proposes that the abundance of transposable elements (TEs) in the genome is the consequence of a balance between two antagonistic forces: their tendency to increase in number by transposition and natural selection against deleterious effects on their hosts (Doolitle & Sapienza, 1980; Orgel & Crick, 1980). These effects are generally attributed to the disruption by TEs of genes or regulatory sequences (Charlesworth & Charlesworth, 1983), to chromosomal rearrangements due to ectopic recombination between elements inserted in non-homologous chromosomal locations (Langley *et al.*, 1988), and/or to deleterious effects directly caused by their transpositional activity (Brookfield, 1996). An expected consequence of this equilibrium is that TEs should be more abundant in regions of the genome where (1) they are more likely to cause fewer or

weaker deleterious effects, and/or (2) selection is less efficient in removing them. In agreement with these predictions, recent analyses of the distribution of TEs in different organisms whose genome sequences are available have shown that TEs are rarely inserted into coding sequences, and that accumulation of elements takes place almost exclusively in the non-coding DNA, where their phenotypic effects are expected to be relatively slight (Adams *et al.*, 2000; Duret *et al.*, 2000; Bartolomé *et al.*, 2002; Kaminker *et al.*, 2002; Wright *et al.*, 2003). In addition, in the *Drosophila melanogaster* genome there is a negative association between TE abundance and rate of recombination (Bartolomé *et al.*, 2002). This observation has been explained by the combined effects of two processes: the expected reduction in the efficacy of selection against weakly deleterious mutations in regions of reduced recombination (the Hill–Robertson effect) (Hill & Robertson, 1966), and a low frequency of ectopic exchange in such regions (Charlesworth *et al.*,

* Corresponding author. e-mail: Carolina.Bartolome@ed.ac.uk

1992*b*; Bartolomé *et al.*, 2002). An ultimate consequence of the irreversible and continuous accumulation of TEs in non-recombining regions is the inactivation of genes and, in the long term, the irremediable deterioration of these fractions of the genome (Steinemann *et al.*, 1993).

To evaluate the importance of this process, data on the population frequencies of TE insertions, particularly in regions of reduced recombination, are needed. The few relevant large-scale studies reported so far suggest that TEs are usually found at low frequency at each site (Charlesworth *et al.*, 1992*a*; Biémont *et al.*, 1994). However, these studies have been carried out by means of *in situ* hybridization of TE probes on polytene chromosomes extracted from natural populations, and this method is probably not the most adequate to estimate the insertion frequencies of TEs (see Section 3). The application of techniques with higher resolution should overcome the intrinsic problems of *in situ* hybridization. Here we report on the results of a PCR-based analysis of the frequency of TE insertions in three different natural populations of *D. melanogaster*. We investigated 27 TEs from 12 families, representing retroposons and transposons, located in the regions of null (chromosome 4, the base of 2R and the tip of the X) and free recombination (middle section of 2R).

2. Materials and methods

(i) *Fly strains*

We analysed a sample of 28 isofemale lines of *D. melanogaster* from three different locations in Zimbabwe – *Z* (Victoria Falls Hotel grounds), *ZH* (Harare city) and *ZS* (Sengwa Wildlife Reserve) – and a *D. simulans* strain (MW23) from Malawi, kindly provided by P. Andolfatto. The reason for using African populations of flies is that this should avoid the effect of demographic problems, such as the existence of population bottlenecks originated by the colonization of non-African areas (David & Capy, 1988; Andolfatto, 2001).

(ii) *Selection of the TE insertions and genomic regions for the analysis*

The criteria for choosing the insertions were as follows. (1) Insertions should represent regions of suppressed and high recombination. (2) With its approximately 1·2 Mb (Adams *et al.*, 2000), the fourth chromosome is the largest single block of non-recombining euchromatin in the *D. melanogaster* genome (Hochman, 1976). Therefore, we decided that the TE insertions should represent its entire length, plus a sample from other regions of low recombination such as the base of an autosome and the tip of

the X. The euchromatin of chromosome 4 of *D. melanogaster* harbours about 102 TEs (Kaminker *et al.*, 2002). Considering the low levels of nucleotide polymorphism reported for this chromosome (Berry *et al.*, 1991; Jensen *et al.*, 2002), little variation is expected along its sequence. We selected 12 insertions (one out of every nine) evenly distributed along the chromosome, which should be representative of the whole of its euchromatic fraction. (3) Insertions in introns and intergenic regions of the fourth chromosome should be equally represented. Individual insertions that met these criteria were picked at random. As a result, 16 of the 27 elements were located in regions of null recombination (Table 1): 12 on chromosome 4 (102A3–102F8), 3 at the base of 2R (41E3–41F3) and one at the tip of the X (1A2). The remaining elements corresponded to insertions in highly recombining portions of chromosomes X (4C15–13A9) and 2R (43A2–47B4).

These TEs represented 12 different families: five retroelements with long terminal repeats (LTRs) (*297*, *Burdock*, *roo*, *mdg1*, *Blastopia*), three retroelements without LTRs (*jockey*, *jockey2*, *Doc*) and four transposons (*1360*, *HB1*, *S-element*, *hopper*). The insertions of *jockey*, *Doc*, *Burdock*, *hopper* and the *S-element* were present in regions of both high and suppressed recombination, while *Blastopia*, *1360*, *jockey2* and *HB1* were located only in non-recombining regions.

(iii) *Estimates of the population frequency of the insertions*

A direct estimate of the allelic frequency of each insertion in the sample was obtained by means of PCR with pairs of primers specific to the sequences flanking the elements (Petrov *et al.*, 2003), using genomic DNAs from a single male from each line as templates. This allowed us to assess the presence/absence of autosomal insertions in both chromosomes of each individual examined. The PCR primers were designed at an average of 500 base pairs upstream and downstream of each predicted TE location, which was determined using the Release 3 annotation of the *Drosophila melanogaster* genome (http://www.fruitfly.org/cgi-bin/annot/query). When the insertion was present a band of expected length (flanking sequences + TE) was obtained, and when the insertion was absent a shorter PCR product (representing only the flanking sequences) was detected. In cases were individuals were heterozygous we observed the presence of both amplicons.

It is very unlikely that a different TE of the same size could be present at a given site. However, to ensure that we were amplifying the right TE, we performed TE-specific PCRs in eight of the insertions. These were carried out using one primer internal to

Table 1. *Transposable element insertions under analysis*

Rec.	Chrom.	Map	TE (no. in Release 3)	Length (bp)	% of canonical sequence	Location Introns	Location Intergenic (between genes)
Null	4	102A3	*1360* (1481)	447	13	*plexB*	
,,	4	102A4	*1360* (1482)	894	26	*ci*	
,,	4	102B1	*S-element* (1493)	214	12		CG31999 and *yellow-h*
,,	4	102B5	*1360* (1502)	470	14	*Syt7*	
,,	4	102C1	*1360* (1514)	499	15		*caMKI* and *bip2*
,,	4	102C4	*HB1* (1517)	1233	75		CG11533 and *zfh2*
,,	4	102D1	*1360* (1520)	1010	30	*Eph*	
,,	4	102D3	*jockey2* (1521)	272	8		*mav* and CG1732
,,	4	102D6–E1	*jockey* (not annotated)	861	17	*ey*	
,,	4	102F6	*S-element* (1551)	221	13		CG32018 and *RfaBp*
,,	4	102F8	*1360* (1558)	1082	32		*Kif3C* and *pho*
,,	4	102F8	*jockey2* (1562)	1256	37	*pho*	
,,	2R	41E3	*Doc* (590)	4301	91		centromere and *TpnC41C*
,,	2R	41F1	*hopper* (634)	1119	78		CG2905 and *d4*
,,	2R	41F3	*Burdock* (664)	2575	40	CG30437	
,,	X	1A2	*Blastopia* (3)	5031	100		*Or1a* and *y*
High	2R	43A2	*jockey* (765)	5007	100		*Or43A* and *Ady43A*
,,	2R	44B1–B2	*297* (774)	6523	93		*cul-4* and CG30372
,,	2R	44D1	*S-element* (778)	388	22		CG8693 and CG30359
,,	2R	45A13–B1	*Burdock* (783)	6413	100		CG30345 and CG8008
,,	2R	45C9–D1	*roo* (785)	8976	99		*Or45a* and CG13954
,,	2R	47B4	*mdg1* (802)	6773	91		CG12934 and *stan*
,,	X	4C15–C16	*jockey* (51)	5018	100		CG6986 and CG12683
,,	X	4F5	*jockey* (60)	1153	23		*CanB* and *SK*
,,	X	5B3	*Doc* (64)	4721	100		CG33080 and CG15773
,,	X	7C1	*hopper* (82)	1431	100		CG1402 and CG10920
,,	X	13A9	*S-element* (135)	543	31		CG9057 and CG33177

the element and another located in the flanking region. All these TE identification tests were proved correct. As a positive control we employed genomic DNA extracted from the strain used in the *Drosophila* Genome Project (*y bw*[1] *cn*[1] *sp*[1]). A single band of the expected size was obtained in all cases (details of the primers and PCR conditions can be obtained from the authors on request).

(iv) *Cloning and DNA sequencing*

The PCR amplification product of the *jockey* insertion that was found in *D. simulans* was cloned (TOPO-TA, Invitrogen) and sequenced on an ABI377 automatic sequencing machine using Dyenamic (Amersham). To minimize the errors in the sequencing procedure at least three plasmids of the cloning reaction were sequenced. The read-outs were checked for accurate base calling and assembled using Sequencher (Gene Codes Corporation).

(v) *Sequence analysis*

Sequences of all TE insertions, as well as their flanking regions, were obtained from the published *D. melanogaster* genome (release 3), including the 69 *jockey* elements identified in the euchromatin (Adams *et al.*, 2000; Kaminker *et al.*, 2002).

All the alignments were performed with McAlign (Keightley & Johnson, 2004), which implements a statistical method based on an evolutionary model of the frequency distribution of gaps and substitutions observed in *Drosophila*.

The average number of nucleotide substitutions per site (D_{XY}) between the TE insertions in the published genome and their canonical sequences (available at http://www.fruitfly.org/p_disrupt/datasets/ NATURAL_TRANSPOSABLE_ELEMENTS.fa) (see Section 3) was calculated using the Kimura two-parameter method with the Jukes and Cantor correction, as implemented in the software package Mega v 2.1 (Kumar *et al.*, 2001).

3. Results and Discussion

(i) *Population frequency versus recombination rate*

We found that the frequencies of insertions in regions of null recombination were much greater (0.83 ± 0.092, mean $\pm$ SE) than in freely recombining sequences

(0·19±0·121). Thirteen of the 16 TE insertions studied in the non-recombining fraction of the genome were fixed (Table 2). These included the 12 TEs located on chromosome 4 and one insertion on chromosome 2R. Two other TEs, residing on chromosomes 2R and X, were absent in all the lines except the control. The remaining one, a *Doc* insertion, was found at an intermediate frequency (0·32). In contrast, most insertions from the highly recombining fraction of the genome were not present in the sample (Table 3). The exceptions were two *S-element* insertions (one on chromosome 2R and the other on chromosome X) that were fixed in all the populations, and one insertion of *297* that was present in heterozygosis in four of the 10 individuals from the Z population, at an allelic frequency of 0·20.

These results are in good agreement with the expected deleterious effects of TE insertions under the selfish DNA hypothesis. The absence from the African sample of most elements located in the highly recombining genome is consistent with previous evidence that TEs are usually segregating at low frequencies in such regions (Charlesworth *et al.*, 1992*a*), where selection is expected to be more efficient in removing them and/or where gene density is higher. There is, consequently, little chance that the insertions identified in these regions in the single genome represented in the Drosophila Genome Project would also be found here. On the other hand, the observation that most elements in the non-recombining genome are fixed strongly suggests that TEs are building up in numbers in these regions. The causes of this accumulation could either be the fact that there are no deleterious effects due to ectopic exchange, and/or that the effectiveness of natural selection is greatly reduced in these regions. Several population genetic models have been put forward to explain the dynamics of genetic variation in regions of reduced recombination. These models explore various predicted consequences of the reduction in the efficacy of selection associated with the interference between linked selected alleles (the Hill–Robertson effect) (Hill & Robertson, 1966). (1) The background selection model (Charlesworth *et al.*, 1993) proposes that strong deleterious mutations removed from the population by natural selection will carry away all associated variants, imposing a reduction of the effective population size (N_e), with the consequent increase in the chance that mildly deleterious mutations can become fixed by random drift. (2) The rise in frequency of a favourable mutation through a population will cause a hitchhiking effect carrying along to fixation all associated mildly deleterious mutations, and sweeping away most unlinked variation (Maynard Smith & Haigh, 1974). Finally, (3) Muller's ratchet proposes that a progressive accumulation of deleterious mutations in a finite population could be a consequence of the stochastic loss of the class of chromosomes carrying the smallest number of deleterious mutations (Muller, 1964). The loss of the least-loaded class is followed by the fixation of a deleterious allele in the entire population. In the absence of recombination and back-mutation (excision of elements), these mechanisms are irreversible.

All these processes cause a reduction in neutral nucleotide variation. Indeed, a 20- to 30-fold reduction in the levels of nucleotide variation with respect to freely recombining regions has been reported at the site of the ci^D, and *ankyrin* genes on the fourth chromosome of *Drosophila melanogaster* (Berry *et al.*, 1991; Jensen *et al.*, 2002) and on the neo-Y chromosome of *Drosophila miranda* (Bachtrog & Charlesworth, 2002). However, there is recent evidence that the level of polymorphism varies between different regions of the chromosome 4 of *D. melanogaster* (Wang *et al.*, 2002), suggesting the existence of different chromosomal domains. In the present study, we covered most sections across the euchromatic sequence of chromosome 4 (from 102A3 to 102F8), finding no variation (the 12 TE insertions were homozygous in all the lines; Table 2). One way to explain these different findings is to assume that certain regions of this chromosome are experiencing some recombination, probably involving gene conversion, for which there is some evidence (Jensen *et al.*, 2002; Wang *et al.*, 2002).

The high frequencies of TE insertions in the non-recombining fraction of the genome reported here are consistent with other studies describing isolated examples of fixation of TEs on regions that do not recombine like the β-heterochromatin (Carmena & Gonzalez, 1995) and the fourth chromosome of *D. melanogaster* (Jensen *et al.*, 2002) or the neo-Y of *D. miranda* (Bachtrog, 2003). However, this result disagrees with other population surveys of TEs using *in situ* hybridization on polytene chromosomes (Charlesworth *et al.*, 1992*a*; Biémont *et al.*, 1994), whose main conclusion is that euchromatic TE insertions segregate at low frequencies in natural populations of *D. melanogaster*. This includes a recent analysis on the fourth chromosome (Carr *et al.*, 2001) where TE insertions were found at a wide range of frequencies, with only a few of them reaching high values (but not fixation). There are several reasons why *in situ* hybridization on polytene chromosomes may underestimate insertion frequencies. First, the intensity of the hybridization signal is associated with both the length and the degree of homology between the probe and the target sequences. Second, on average, fixed TE insertions are likely to have had more time than polymorphic ones to diverge from extant active copies by accumulating point mutations and, especially, deletions (see below). Given that the *in situ* technique is quite inefficient at detecting short regions of homology and that TEs located in regions of null

Table 2. *Estimates of the population frequencies of the TE insertions* (*excluding the control line*) *located in regions of null recombination*

| | 2R | | | X | 4 | | | | | | | | | | | |
| | 41E | 41F | 41F | 1A | 102A | 102A | 102B | 102B | 102C | 102C | 102D | 102D | 102D-E | 102F | 102F | 102F |
Sample	Doc	hopper	Burdock	blastopia	1360	1360	S-element	1360	1360	HB1	1360	jockey2	jockey	S-element	1360	jockey2
Z139	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z144	O	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z149	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z155		X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z164	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z168	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z184	O	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z185	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z189		X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Z190	O	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH12	O	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH16	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH23	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH26		X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH27	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH32	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH33	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH40	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZH42	X	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS02	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS10	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS11	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS24		X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS29	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS30	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS49	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS53		X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
ZS56	–	X	–	–	X	X	X	X	X	X	X	X	X	X	X	X
Control	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Frequency	0·32	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1

X, O and –, mean that the insertions were homozygous, heterozygous or absent in the individual analysed, respectively.

Table 3. *Estimates of the population frequencies of the TE insertions (excluding the control line) located in regions of high recombination*

| Sample | 2R | | | | | | X | | | | |
	43A *jockey*	44B *297*	44D *S-element*	45A-B *Burdock*	45C-D *roo*	47B *mdg1*	4C *jockey*	4F *jockey*	5B *Doc*	7C *hopper*	13A *S-element*
Z139	–	O	X	–	–	–	–	–	–	–	X
Z144	–	–	X	–	–	–	–	–	–	–	X
Z149	–	–	X	–	–	–	–	–	–	–	X
Z155	–	O	X	–	–	–	–	–	–	–	X
Z164	–	O	X	–	–	–	–	–	–	–	X
Z168	–	–	X	–	–	–	–	–	–	–	X
Z184	–	–	X	–	–	–	–	–	–	–	X
Z185	–	–	X	–	–	–	–	–	–	–	X
Z189	–	–	X	–	–	–	–	–	–	–	X
Z190	–	O	X	–	–	–	–	–	–	–	X
ZH12	–	–	X	–	–	–	–	–	–	–	X
ZH16	–	–	X	–	–	–	–	–	–	–	X
ZH23	–	–	X	–	–	–	–	–	–	–	X
ZH26	–	–	X	–	–	–	–	–	–	–	X
ZH27	–	–	X	–	–	–	–	–	–	–	X
ZH32	–	–	X	–	–	–	–	–	–	–	X
ZH33	–	–	X	–	–	–	–	–	–	–	X
ZH40	–	–	X	–	–	–	–	–	–	–	X
ZH42	–	–	X	–	–	–	–	–	–	–	X
ZS02	–	–	X	–	–	–	–	–	–	–	X
ZS10	–	–	X	–	–	–	–	–	–	–	X
ZS11	–	–	X	–	–	–	–	–	–	–	X
ZS24	–	–	X	–	–	–	–	–	–	–	X
ZS29	–	–	X	–	–	–	–	–	–	–	X
ZS30	–	–	X	–	–	–	–	–	–	–	X
ZS49	–	–	X	–	–	–	–	–	–	–	X
ZS53	–	–	X	–	–	–	–	–	–	–	X
ZS56	–	–	X	–	–	–	–	–	–	–	X
Control line	X	X	X	X	X	X	X	X	X	X	X
Frequency	0	0·07	1	0	0	0	0	0	0	0	1

X, O and –, mean that the insertions were homozygous, heterozygous or absent in the individual analysed, respectively.

recombination are generally older (Blumenstiel *et al.*, 2002; Petrov *et al.*, 2003), and therefore shorter and more divergent from the active copies than those found in highly recombining regions, the resolution of this method might not be great enough to detect the presence of the elements. Third, the banding pattern in regions of low recombination (the base of the autosomes and the fourth chromosome) is more difficult to interpret, so fixations might be overlooked. The use of the released sequence of the *Drosophila melanogaster* genome (Adams *et al.*, 2000) for the application of sequence-specific PCR avoids this lack of resolution and allows particular sequences to be identified in a much more accurate way.

According to Petrov *et al.* (2003), TE families with lower copy numbers and/or shorter lengths may experience a reduced strength of selection due to the lower probability of inducing ectopic exchange and are, therefore, more likely to attain high population frequencies. However, our data suggest that this is a general consequence of the lack of recombination rather than specific properties of the different element families. First, all the TEs investigated on chromosome 4 were fixed regardless of the family size. Second, a comparison of the frequencies of insertions in regions of high and null recombination of five families with different copy numbers (*Burdock*, *hopper*, *S-element*, *Doc* and *jockey*, with 13, 15, 51, 55 and 69 copies, respectively; Kaminker *et al.*, 2002) revealed that the representatives of the two most abundant families, *Doc* and *jockey*, were fixed in the fourth chromosome. *Burdock* (the less abundant family) was not present in either region in the sample, there being no hint of an association between family size and insertion frequency.

It should also be pointed out that the four *S-element* insertions were found to be fixed in both fractions of the genome, regardless of the rate of recombination. This agrees with previous reports of fixations of *S-element* insertions in a recombining region of the genome (87C), where there were indications that they could be of functional significance for

the host (Maside *et al.*, 2002). In this context, there are grounds to believe that the positively selected properties might extend to other members of the family, which would provide an explanation for their fixation other than relaxed selection.

Another observation of our study is a different rate of fixation for the two major classes of TEs: DNA-based elements were more likely to be fixed than retrotransposons (with or without LTRs). Twelve of the 13 transposons were fixed (including two of three in regions of high recombination), whereas only three of the 14 retroelements reached the same frequency, despite the fact that six of them were in the non-recombining fraction of the genome. Although the sample size may be too small to draw any definite conclusion, these results are concordant with previous reports describing a higher rate of enrichment of DNA-based elements in regions of low recombination (Rizzon *et al.*, 2002), particularly on the fourth chromosome (Bartolomé *et al.*, 2002, table 5).

(ii) *Population frequency and age of the insertions*

We also investigated a potential association between the length and population frequency of the insertions. Insertions of TEs belonging to families with long canonical sequences were usually found at lower frequencies (data not shown), but given that not all families were equally represented in regions of high and suppressed recombination, and that DNA-based elements tend to be shorter than retrotransposons (Kaminker *et al.*, 2002), it is difficult to determine whether this pattern is related to the canonical lengths of the families, to specific properties of the two major classes of elements, or to both.

In addition, we found that the average length (relative to the canonical sequence) of fixed elements was significantly shorter than that of polymorphic ones ($28 \cdot 1\% \pm 5 \cdot 52$ vs $86 \cdot 3\% \pm 7 \cdot 53$). This pattern is likely to correlate directly with the age of the insertions. If this is the case, shorter elements should also be more divergent from the consensus sequence (Blumenstiel *et al.*, 2002; Petrov *et al.*, 2003). Consistent with this, we found a significant negative association between the relative size of the insertions (sequences were taken from the DGP: see Section 2) and the average nucleotide divergence from the canonical sequences (D_{XY}) (Kendall's $\tau = -0 \cdot 56$, $P < 0 \cdot 001$, one-tailed; Fig. 1). According to their D_{XY} value, fixed elements show a wide distribution (from $0 \cdot 032$ to $0 \cdot 164$). This could be interpreted as TEs becoming fixed at different evolutionary stages. Considering that the average synonymous divergence between *D. melanogaster* and *Drosophila simulans* is about $0 \cdot 11 \pm 0 \cdot 004$ (Betancourt *et al.*, 2002) we can infer that at least four of the 15 fixation events could have occurred prior to the split between the two species (Fig. 1). To test this, we
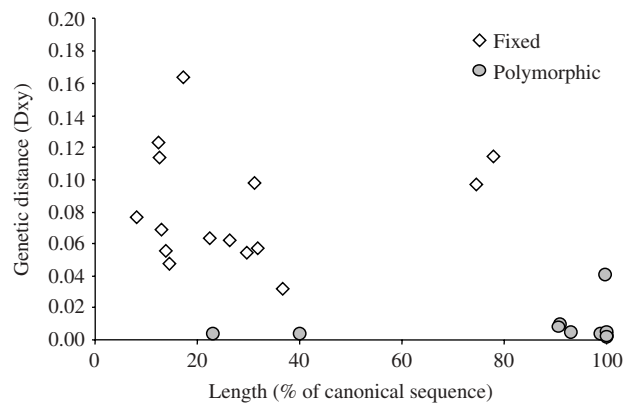

Fig. 1. Association between the genetic distance (D_{XY}, Kimura two-parameter model with JC correction) and the length of the TE insertions, relative to their family's canonical sequence.

investigated the presence in *D. simulans* of any of the 12 insertions fixed on the *D. melanogaster* fourth chromosome (see Section 2), using the *D. melanogaster* primers. Only three of the primer pairs, corresponding to inserts of *1360*, *S* and *jockey* elements (at 102D1, 102F6 and 102D6, respectively), yielded amplification products. This low rate of success is probably due to the above-mentioned level of silent nucleotide divergence which reduces the chance of the primers to work in both species. The sizes of the PCR products indicated that only the *jockey* insertion was present in *D. simulans*. This is not surprising given that the *1360* insertion in *D. melanogaster* is likely to have occurred after the split between the two species ($D_{XY} = 0 \cdot 054$ from the canonical sequence), and that the *S-element* has not been found in any species of the *melanogaster* complex other than *D. melanogaster* (Merriman *et al.*, 1995; Maside *et al.*, 2003). The presence of the *jockey* insertion was further confirmed by obtaining the DNA sequence of the amplicon. It corresponds to a 2257 bp fragment of the second intron of the *eyeless* gene and comprises a 576 bp fragment of the *jockey* element flanked by two fragments of 750 and 931 bp, respectively.

To investigate whether this insertion is a true orthologue of the *D. melanogaster* one, we performed a phylogenetic analysis including these sequences as well as those of the other 69 *jockey* elements identified in the *D. melanogaster* euchromatin (available at http://www.fruitfly.org/p_disrupt/TE.html; one of them did not overlap with the fixed insertions and was excluded from the analysis). D_{XY} between the *D. simulans* and the *D. melanogaster* insertions into the *eyeless* intron (102D) is $0 \cdot 14 \pm 0 \cdot 062$ ($\pm$ SE), and between these and the other euchromatic copies of the *jockey* family is $0 \cdot 26 \pm 0 \cdot 042$ and $0 \cdot 18 \pm 0 \cdot 022$, respectively. Thus, the two insertions are more closely related to each other than to any other member of the family (Fig. 2), which is consistent with the hypothesis that
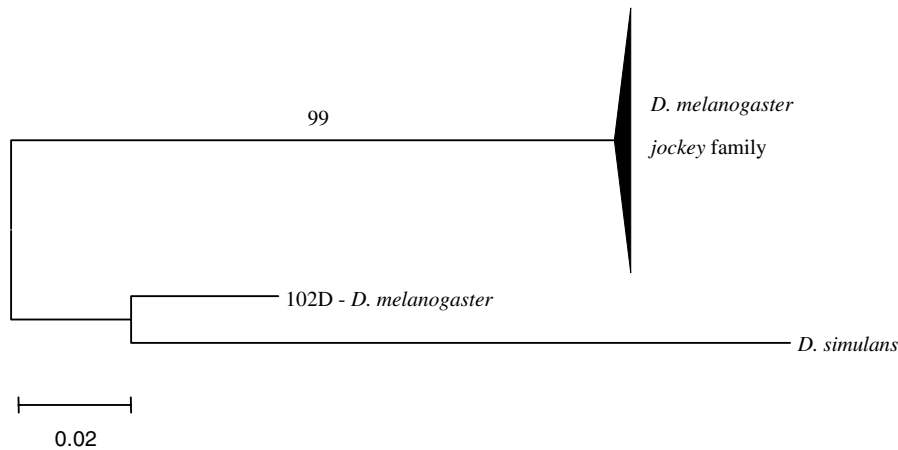
Fig. 2. Neighbor-joining tree representing the phylogenetic relations between the two *jockey* insertions found in the second intron of *eyeless* in *D. simulans* and *D. melanogaster* and the other 68 members of the TE family in *D. melanogaster*. The bar indicates genetic distance units (D_{XY}), and the number above the top branch indicates the bootstrap support value (1000 reps.).

they correspond to a unique insertion that occurred in the common ancestor of the two species.

It is also worth noting the low divergence among the 68 *D. melanogaster jockey* copies (mean pairwise distance $= 0.003 \pm 0.0009$). This had been reported before (Kaminker *et al.*, 2002; Lerat *et al.*, 2003) and could have different causes: (1) a high rate of turnover of copies in this species (i.e. individual elements are quickly lost from the population and replaced by new insertions at other sites, so that most extant elements at a given time have a very recent common ancestor, and therefore little time to diverge) (Brookfield, 1986), (2) a recent increase in the transpositional activity (Brookfield, 1986), and (3) an unusually high rate of gene conversion among members of this family (Charlesworth, 1986). Contrary to the last possibility, the four-allele test (Hudson & Kaplan, 1985) did not provide evidence for a particularly high rate of genetic exchange among copies of this family (data not shown).

Interestingly, in the absence of gene conversion, the lower divergence observed between the *jockey* element at 102D and the other members of the family, than between these and the *D. simulans* orthologue (Fig. 2), can only be explained by a higher rate of silent substitution in the *D. simulans* lineage, which contrasts with previous estimates of an overall higher rate of evolution at synonymous sites in *D. melanogaster* (Akashi, 1995, 1996; Takano, 1998). Assuming a uniform mutation rate at all sites in these species, this is consistent with a significant reduction in the strength of natural selection on codon usage in the *D. melanogaster* lineage (Akashi, 1995).

(iii) *Contributions of TEs to intronic and intergenic sequences*

Another aim of the survey was to investigate whether there are significant differences in the contribution of TE insertions to introns and intergenic regions, as a function of the recombination rate. In *Drosophila melanogaster* the vast majority of TEs are inserted into non-coding sequences, and the proportion of TE-derived DNA in intergenic regions is approximately twice as large as in introns (Bartolomé *et al.*, 2002). On chromosome 4, this ratio is maintained and TEs, apparently, do not contribute much to the build-up of introns, which are much longer than in other chromosomes (Bartolomé *et al.*, 2002). However, the results of our current survey suggest that this intron enlargement, common in non-recombining genomes (Comeron & Kreitman, 2000; Kurek *et al.*, 2000), could be driven by an accumulation of TEs that have been maintained at high frequencies for long periods, and whose insertions can be difficult to detect as such due to their stage of erosion. This is supported by the observed fixation of six TEs in introns in regions of suppressed crossing-over (Table 2), along with other cases in the *ankyrin* gene of *D. melanogaster* and *D. simulans* (Jensen *et al.*, 2002), and in the CG9025 and CG16799 genes of *D. miranda* (Bachtrog, 2003).

We also found that the average frequency of insertions of TEs in introns and in the intergenic fraction of the non-recombining genome was 0.86 ± 0.143 and 0.81 ± 0.126 respectively, suggesting that the genetic circumstances that allow for the fixation of elements in these regions seem to affect all non-coding DNA in a similar way.

In summary, all the results obtained in our survey are consistent with a considerable decrease in the nucleotide variation, in both introns and intergenic DNA, of the non-recombining fraction of the genome (Berry *et al.*, 1991; Jensen *et al.*, 2002). The TEs located in these regions are likely to be fixed and show signs of having been maintained in their location for long periods. The progressive accumulation of TEs in regions where crossing-over is suppressed is usually

the first step in the degeneration of chromosomes (Steinemann & Steinemann, 1992; Steinemann *et al.*, 1993), which suggests that chromosome 4 of *D. melanogaster* is undergoing the early stages of this degenerative and irreversible process.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at silent sites in Drosophila DNA. *Genetics* **139**, 1067–1076.

Akashi, H. (1996). Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**, 1297–1307.

Andolfatto, P. (2001). Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution* **18**, 279–290.

Bachtrog, D. (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. *Nature Genetics* **34**, 215–219.

Bachtrog, D. & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323–326.

Bartolomé, C., Maside, X. & Charlesworth, B. (2002). On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Molecular Biology and Evolution* **19**, 926–937.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Betancourt, A. J., Presgraves, D. C. & Swanson, W. J. (2002). A test for faster X evolution in *Drosophila*. *Molecular Biology and Evolution* **19**, 1816–1819.

Biémont, C., Lemeunier, F., Garcia Guerreiro, M. P., Brookfield, J. F., Gautier, C., Aulard, S. & Pasyukova, E. G. (1994). Population dynamics of the *copia*, *mdg1*, *mdg3*, *gypsy*, and *P* transposable elements in a natural population of *Drosophila melanogaster*. *Genetical Research* **63**, 197–212.

Blumenstiel, J. P., Hartl, D. L. & Lozovsky, E. R. (2002). Patterns of insertion and deletion in contrasting chromatin domains. *Molecular Biology and Evolution* **19**, 2211–2225.

Brookfield, J. F. (1986). A model for DNA sequence evolution within transposable element families. *Genetics* **112**, 393–407.

Brookfield, J. F. (1996). Models for the spread of non-autonomous selfish transposable elements when transpo-

sition and fitness are coupled. *Genetical Research* **67**, 199–210.

Carmena, M. & Gonzalez, C. (1995). Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in *Drosophila melanogaster*. *Chromosoma* **103**, 676–684.

Carr, M., Soloway, J. R., Robinson, T. E. & Brookfield, J. F. (2001). An investigation of the cause of low variability on the fourth chromosome of *Drosophila melanogaster*. *Molecular Biology and Evolution* **18**, 2260–2269.

Charlesworth, B. (1986). Genetic divergence between transposable elements. *Genetical Research* **48**, 111–118.

Charlesworth, B. & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetical Research* **42**, 1–27.

Charlesworth, B., Lapid, A. & Canada, D. (1992*a*). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* **60**, 103–114.

Charlesworth, B., Lapid, A. & Canada, D. (1992*b*). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. *Genetical Research* **60**, 115–130.

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

Comeron, J. M. & Kreitman, M. (2000). The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* **156**, 1175–1190.

David, J. R. & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics* **4**, 106–111.

Doolitle, W. F. & Sapienza, D. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

Duret, L., Marais, G. & Biémont, C. (2000). Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* **156**, 1661–1669.

Hill, W. G. & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

Hochman, B. (1976). The fourth chromosome of *Drosophila melanogaster*. In *The Genetics and Biology of Drosophila* (ed. M. Ashburner & E. Novitski), pp. 903–928. London and New York: Academic Press.

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

Jensen, M. A., Charlesworth, B. & Kreitman, M. (2002). Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* **160**, 493–507.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., *et al.* (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* **3**, RESEARCH 0084-0084.

Keightley, P. D. & Johnson, T. (2004). MCALIGN: stochastic alignment of noncoding DNA sequences based on an evolutionary model of sequence evolution. *Genome Research*, in press.

Kumar, A., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). *MEGA2*: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, Arizona, USA.

Kurek, R., Reugels, A. M., Lammermann, U. & Bunemann, H. (2000). Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of *Drosophila* sp. *Genetica* **109**, 113–123.

Langley, C. H., Montgomery, E., Hudson, R., Kaplan, N. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **52**, 223–235.

Lerat, E., Rizzon, C. & Biemont, C. (2003). Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. *Genome Research* **13**, 1889–1896.

Maside, X., Bartolomé, C. & Charlesworth, B. (2002). *S-element* insertions are associated with the evolution of the *Hsp70* genes in *Drosophila melanogaster*. *Current Biology* **12**, 1686–1691.

Maside, X., Bartolomé, C. & Charlesworth, B. (2003). Inferences on the evolutionary history of the *S-element* family of *Drosophila melanogaster*. *Molecular Biology and Evolution* **20**, 1183–1187.

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

Merriman, P. J., Grimes, C. D., Ambroziak, J., Hackett, D. A., Skinner, P. & Simmons, M. J. (1995). *S* elements: a family of *Tc1*-like transposons in the genome of *Drosophila melanogaster*. *Genetics* **141**, 1425–1438.

Muller, H. (1964). The relation of recombinational to mutational advance. *Mutational Research* **1**, 2–9.

Orgel, L. E. & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

Petrov, D. A., Aminetzach, Y. T., Davis, J. C., Bensasson, D. & Hirsh, A. E. (2003). Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Molecular Biology and Evolution* **20**, 880–892.

Rizzon, C., Marais, G., Gouy, M. & Biemont, C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Research* **12**, 400–407.

Steinemann, M. & Steinemann, S. (1992). Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *Proceedings of the National Academy of Sciences of the USA* **89**, 7591–7595.

Steinemann, M., Steinemann, S. & Lottspeich, F. (1993). How Y chromosomes become genetically inert. *Proceedings of the National Academy of Sciences of the USA* **90**, 5737–5741.

Takano, T. S. (1998). Rate variation of DNA sequence evolution in the *Drosophila* lineages. *Genetics* **149**, 959–970.

Wang, W., Thornton, K., Berry, A. & Long, M. (2002). Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* **295**, 134–137.

Wright, S. I., Agrawal, N. & Bureau, T. E. (2003). Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research* **13**, 1897–1903.