Genet. Res., Camb. (2005), **85**, *pp.* 195–203. *With 2 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007548 Printed in the United Kingdom

Fixation of transposable elements in the *Drosophila melanogaster* genome

XULIO MASIDE[1]*, STAVROULA ASSIMACOPOULOS[2] AND BRIAN CHARLESWORTH[1]
[1] *Institute of Evolutionary Biology, School of Biological Sciences, Edinburgh University, King's Buildings, Edinburgh EH9 3JT, UK*
[2] *Department of Neurobiology, Pharmacology, and Physiology, University of Chicago, Chicago, Illinois, USA*

(*Received 14 January 2005 and in revised form 31 March 2005*)

Summary

We have investigated at the molecular level four cases in which *D. melanogaster* middle repetitive DNA probes consistently hybridized to a particular band on chromosomes sampled from a *D. melanogaster* natural population. Two corresponded to true fixations of a *roo* and a *Stalker* element, and the others were artefacts of the *in situ* hybridization technique caused by the presence of genomic DNA flanking the transposable elements (TEs) in the probes. The two fixed elements are located in the β-heterochromatin (20A and 80B, respectively) and are embedded in large clusters of other elements, many of which may also be fixed. We also found evidence that this accumulation is an ongoing process. These results support the hypothesis that TEs accumulate in the non-recombining part of the genome. Their implications for the effects of TEs on determining the chromatin structure of the host genomes are discussed in the light of recent evidence for the role of TE-derived small interfering-RNAs as *cis*-acting determinants of heterochromatin formation.

1. Introduction

The selfish DNA theory (Doolitle & Sapienza, 1980; Orgel & Crick, 1980) proposes that genomes are made up of two components: *specific*, which corresponds to all DNA sequences beneficial to the organism, and *non-specific*, with little or no functional significance for the host, whose only purpose is its own perpetuation within the genome. Given that it does not play a role for the host, the abundance of this non-specific DNA is expected to be determined by a balance between its ability to spread and its elimination due to the deleterious fitness effects associated with its presence. For many researchers, transposable elements epitomize the essence of the selfish DNA (Doolitle & Sapienza, 1980), and several aspects of their genome distribution and population properties are consistent with this view: (i) surveys of the genomes of various organisms, including *Arabidopis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Tetraodon nigroviridis*, indicate that TEs tend to be more abundant in regions where they are likely to be less deleterious and/or where natural selection is expected to be less effective in removing them (Bartolomé *et al.*, 2002; Kaminker *et al.*, 2002; Rizzon *et al.*, 2002; Wright *et al.*, 2003; Fischer *et al.*, 2004), and (ii) population data reveal that individual insertions usually segregate at low frequencies in *Arabidopsis* (Wright *et al.*, 2001) and *Drosophila* (Charlesworth *et al.*, 1992; Hoogland & Biémont, 1996; Petrov *et al.*, 2003). Fits of these data to models of the population dynamics of TEs are consistent with natural selection acting as the main force opposing element spread (Charlesworth *et al.*, 1994*b*).

However, the selfish DNA theory does not exclude the possibility that 'the host genome occasionally finds some use for particular selfish DNA sequences' (Orgel & Crick, 1980). This possibility has attracted much interest, and issues such as the potential benefits of the TEs to their hosts or their role in the evolution of genomes have been the subject of much debate (McDonald, 2000; Kidwell & Lisch, 2001; Kazazian, 2004). Many properties of TEs could make them serve

* Corresponding author.
Present address: Grupo de Xenética Evolutiva, Unidade de Medicina Molecular-Complexo Hospitalario Universitario de Compostela, Choupana s/n, Edificio Consultas planta -2, 15706 Santiago de Compostela, Galiza, Spain. Tel.: +34 981 951491. Fax: +34 981 951473. e-mail: xmaside@usc.es

as a substrate for the evolution of their hosts, and there is a growing catalogue of cases where TE-derived sequences provide or modify a host function and have been assimilated by the host genome (Franchini *et al*., 2004, and citations therein).

It is difficult to assess the significance of this contribution of TEs to the evolutionary process of their hosts by listing isolated cases from many different organisms. One approach would be to estimate the fraction of all TEs in a genome that confer a selective advantage on their carriers. From a population genetics perspective, these insertions are expected to segregate at increasingly high frequencies within the population in which they originated, driven by natural selection, until they reach fixation and thus become stable functional components of the host genome.

In their population survey of *D. melanogaster*, Charlesworth *et al*. (1992) found that five of their middle-repetitive DNA probes consistently hybridized to certain genomic locations in all the chromosomes in the sample. This was interpreted as suggesting that these elements could be fixed at these positions in natural populations. A detailed investigation of one case showed that it involved a fragment of an *S-element*, fixed at 87C in the *D. melanogaster* genome, just upstream of the promoter region of *Hsp70Bb*, a member of the *Hsp70* gene family. Patterns of nucleotide sequence variation in this region suggest that this insertion is under purifying selection, and that it may be of adaptive value for the host (Maside *et al*., 2002). The subsequent finding of three other sites where *S-element* fragments are fixed strengthens this conclusion (Bartolomé & Maside, 2004).

Here, we report an analysis of the other four putative TE fixations described by Charlesworth *et al*. (1992). We found that two were artefacts caused by the nature of the probes, and the other two corresponded to true fixations of a *roo* and a *Stalker* element at the bases of the X chromosome and 3L, respectively. These fixed insertions are embedded in large clusters of other elements, most of which may also be fixed. The implications of these findings for the evolutionary role of TEs in the host genomes are discussed in the light of recent data on the role of TEs in the molecular structure of chromatin.

2. Materials and methods

(i) *Fly stocks*

We used samples of two *D. melanogaster* natural populations, from Beltsville (Maryland, USA) and Zimbabwe. The first one consisted of 7 isogenic lines – IS2, IS3, IS4, IS5, IS9, IS24, IS25 – each of them constructed from a set of X, Y, second and third chromosomes extracted from a natural population at Beltsville (Charlesworth *et al*., 1994*a*). The second consisted of 24 isofemale lines originated from gravid females collected from three locations in Zimbabwe (Victoria Falls Hotel grounds (Z), Harare city (ZH) and Sengwa Wildlife Reserve (ZS)): Z139, Z144, Z149, Z155, Z164, Z168, Z184, Z185, Z189, Z190, ZH12, ZH16, ZH23, ZH26, ZH27, ZH32, ZH33, ZH40, ZH42, ZS2, ZS11, ZS24, ZS49, ZS56, described in Begun & Aquadro (1993). As PCR controls, we used the stock y^1; cn^1 bw^1 sp^1 (stock no. 2057 from Bloomington Stock Center), which was the source of the DNA used to obtain the *D. melanogaster* genome sequence by the DGP (Adams *et al*., 2000), and Oregon-R.

All flies were reared on Lewis' cornmeal–yeast–sucrose–agar medium and maintained at 25 °C.

(ii) *Probes*

cDm2158, *cDm2181*, *cDm4006* and '*roo*' are middle repetitive genomic sequences of *D. melanogaster* cloned into pBR322. They were obtained from a cell line Kc_0, which was derived from an F1 cross between an *se* and an *e D. melanogaster* lines (Finnegan *et al*., 1978; table 7·1 of Ashburner, 1989). Restriction maps of these clones and the identities of the element families they contain can be found in Charlesworth *et al*. (1994*a*).

(iii) *DNA cloning*

DNA restriction fragments were separated by agarose gel electrophoresis, purified using QIAquick Gel Purification Kit (QIAgen) and subcloned into pBluescript K/S following the manufactures' instructions and standard DNA cloning methods.

DNA sequencing of the subclones was performed using the M13 phage forward and reverse primers and the Dyenamic Sequencing Kit (Amersham Biosciences); products were run on an ABI 377 automated sequencing machine and sequence readouts were checked manually for accurate base calling and assembled using Sequencher (Gene Codes Corporation).

(iv) *DNA sequence analyses*

Nucleotide sequences were edited with Bioedit v.4.8.9 (Hall, 1999). Homology analyses between problem sequences and (i) the *D. melanogaster* genome (release 3·2) and all *D. melanogaster* repetitive sequences entries in GenBank and FlyBase were done using the BLAST (Altschul *et al*., 1990) portals at NCBI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) and FlyBase (http://flybase.net/blast/) and (ii) the Repbase Update library of all prototypic *D. melanogaster* repetitive sequences (http://www.girinst.org/

https://doi.org/10.1017/S0016672305007548 Published online by Cambridge University Press

Repbase_Update.html; Jurka, 2000), using Censor (Jurka *et al*., 1996) from its online server at the Genetic Information Research Institute website (www.girinst.org/Censor_Server.html). Descriptions of all TEs referred to in this study can be found in FlyBase (http://flybase.bio.indiana.edu/transposons/lk/melanogaster-transposon. html) and Repbase Update.

Population genetic parameters were estimated with DnaSP v·4 (Rozas *et al*., 1999) and Mega v2·1 (Kumar *et al*., 2001). The exploration of the *Drosophila* genome annotation was done with the aid of Apollo (v1.4.4) (Lewis *et al*., 2002; www.fruitfly.org/annot/apollo/).

(v) In situ *hybridization*

Preparation of the slides, hybridization and detection by the diaminobenzidine/peroxidase method were performed with minor modifications of previously described protocols (Maside *et al*., 2000).

(vi) *PCR estimates of the population frequency of TE insertions*

Direct estimates of the allelic frequencies of TE insertions were obtained by PCR amplification with primers at both sides of each insertion (Petrov *et al*., 2003). Given the large size of the *roo* elements, instead of both primers being at either side of the insertion, we designed one of the primers to be inside the element. In these cases, the presence of the element is indicated by a PCR product of the expected size, and its absence by the lack of a product. Template genomic DNAs were extracted from single adult males.

The primer pairs used were F1: GGCACTTGAAATTCTGAGTC with L1: CCTGCTTCGTTTCTTAGTGT for *Stalker*; F2: ATTGCGTTTACTGGTTGG with L2: AAGATTTAGATTTCCGTCATA, for *roo*[1] and F3: AGTTGTGGTATTCGGCTTTCTT with L3: AGTCTTCTCATTTGGGATTTTA and F4: GTCGTTCAAGGTAATAGG with L4: ACAGTGGTCAAAGTAAGA for *roo*[2]. All primers produced amplicons of the expected sizes when tested with control genomic DNA from y^1; cn^1 bw^1 sp^1.

3. Results

The middle-repetitive DNA clones used as *in situ* hybridization probes in the population survey by Charlesworth *et al*. (1992) have been characterized by restriction mapping. Each was found to contain a copy of an element of the family in question, usually flanked by smaller fragments of genomic DNA (Charlesworth *et al*., 1994*a*). Therefore, the consistent hybridization of one of these probes to a particular band in all chromosomes from a population sample may have different causes: (i) the probe contains a fragment of genomic DNA homologous to the band in question along with the TE fragment, (ii) there is more than one element of the same family inserted into different positions in the same chromosome band and segregating in the population, (iii) a single copy of this TE inserted at that particular site, and the insertion is at high frequency or fixed in the population.

To test alternative (i), we isolated restriction fragments internal to and flanking the elements in the clones, and used them as probes for *in situ* hybridization to chromosomes from the Beltsville population. Once the presence of an element at a certain chromosomal site was ascertained, we were able to distinguish between (ii) and (iii) by searching for that particular insertion in the DGP sequence, followed by an estimate of its frequency by means of its specific PCR amplification, in two population samples: one from the same Beltsville (ML, USA) population as that used in the original study (Charlesworth *et al*., 1994*a*), and a second one from Zimbabwe (see Section 2).

(i) cDm4006 (297) *at 99E*

Three restriction fragments from *cDm4006* were subcloned and analysed separately: a 2·5 kb *Hind*III fragment internal to the *297* element and two fragments from the flanking sequence – a 2·0 kb *Eco*RI–*Pst*I and a 2·3 kb *Pst*I–*Bam*HI (Fig. 1*a*). When used as probes for *in situ* hybridization to chromosomes from the Beltsville population, and the Oregon R and y^1; cn^1 bw^1 sp^1 stocks, the former revealed no signal at 99E1 but hybridized to multiple other sites, and the two latter hybridized at 99E1–E2 only. The nucleotide sequence of the *Eco*RI–*Pst*I fragment was obtained and a BLAST search against the DGP sequence revealed that it is homologous to a genomic sequence from 99E1–E2 (AN: AE003772). Furthermore, the pattern of restriction sites in this genomic region precisely matches that of the 7·5 kb sequence flanking the *297* element in *cDm4006*. This result, together with the *in situ* data, strongly suggests that the DNA flanking *297* in *cDm4006* is responsible for the hybridization at 99E1, and that *297* is not present in this position in the genome sequenced by DGP (y^1; cn^1 bw^1 sp^1), nor in the Beltsville population. It is thus likely that this *297* insertion was present at this location only in the genome of the cell line used to obtain *cDm4006* (see Section 2). Given that TEs are notoriously unstable in cell culture (Potter *et al*., 1979), a search for this particular insertion in the cell line used to obtain the probe was not carried out.

(ii) cDm2181 (mdg1) *at 60B*

Two restriction fragments from *cDm2181* were used as probes for *in situ* hybridization on chromosomes

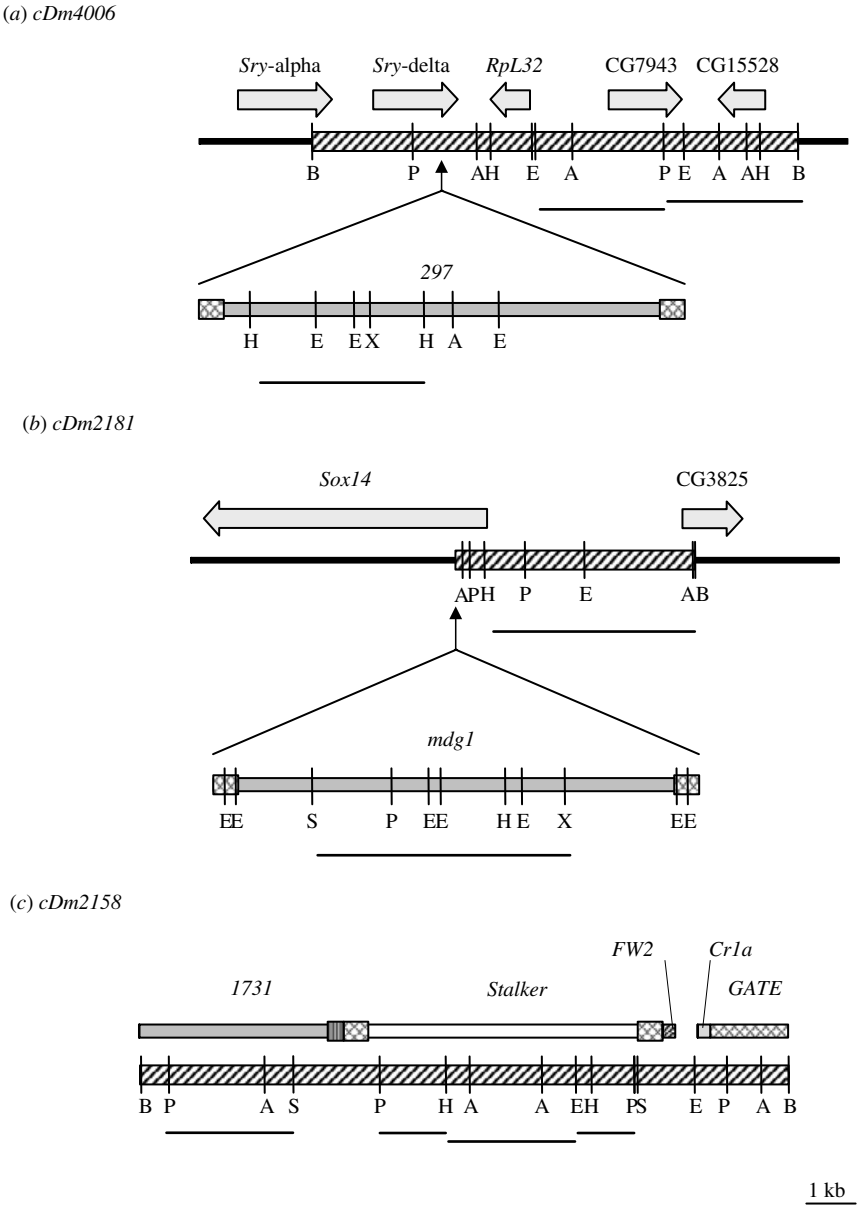

Fig. 1. Schematic representation of the genomic regions cloned in *cDm4006*, *cDm2181* and *cDm2158*. Dashed boxes are the segments included in the clones and the black lines represent flanking genomic DNA. Above are genetic features identified in the genomic sequences: genes (arrows) and TEs (boxes; thicker boxes are the LTRs). Below are TE insertions found in the clones. Restriction sites: A, *Ava* I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; X, *Xba*I. Black lines represent the restriction fragments used as probes for *in situ* hybridization.

from the Beltsville population (Fig. 1*b*). A 4·9 kb *Sal*I–*Xba*I fragment internal to *mdg-1* hybridized at all the sites previously scored as insertions of *mdg1* in the population using the whole of *cDm2181* as a probe (Charlesworth *et al*., 1992), but not at 60B. A 3·5 kb *Hin*dIII–*Ava*I fragment from the region flanking *mdg1* in *cDm2181* hybridized at 60B only, suggesting that it is the flanking DNA, rather than *mdg1*, which is responsible for the consistent hybridization of *cDm2181* at this site. A BLAST analysis showed that the DNA flanking *mdg1* corresponds to a unique sequence from the chromosomal region 60A1–60B8 in DGP (AN: AE003462). It comprises the 5′ ends of two genes located in opposite directions (*Sox14* and CG3825) and the intergenic sequence between them (Fig. 1*b*). In the DGP sequence, there is no *mdg1* insert at this position. It is thus reasonable to postulate that this *mdg1* insertion was present in the individual whose genomic DNA was used to produce *cDm2181*, but not in the y^1; cn^1 bw^1 sp^1 stock nor in the Beltsville population. As explained above, a search of this *mdg1* insertion in the genome of the cell line *cDm2181* was obtained from is not recommended.

https://doi.org/10.1017/S0016672305007548 Published online by Cambridge University Press

(iii) cDm2158 (1731) *at 80B*

The region flanking *1731* in *cDm2158* was digested with a cocktail of different restriction enzymes, and fragments covering the whole region were subcloned independently and their nucleotide sequences obtained. In addition to the *1731* element, this genomic sequence includes a full copy of *Stalker* and fragments of three other elements: *FW2*, *Cr1a* and *GATE* (Fig. 1*c*).

A probe for the 2·6 kb *Pst* I–*Sal* I restriction fragment, which is internal to *1731*, hybridized to a single chromosomal site and the centromeric region, in nuclei from all Beltsville isogenic lines and the y^1; cn^1 bw^1 sp^1 stock. These sites differed among genetic backgrounds but never occurred at 80B. Three other restriction fragments from the region corresponding to the *Stalker* element (*Pst*I–*Hin*dIII, *Hin*dIII–*Eco*RI and *Eco*RI–*Pst*I; Fig. 1*c*) hybridized at 80B, as well as at several other euchromatic sites and the centromeric region in the same lines. This clearly suggests that it is *Stalker* and not *1731* that is responsible for the consistent hybridization of *cDm2158* at 80B in the Beltsville population.

A survey of the DGP sequence failed to detect the particular association of *Stalker* and *1731* found in *cDm2158*, which suggests that this probe may come from the heterochromatin or that it is specific to the strain used to make the construct. However, there is a *Stalker* element at the base of 3L (BACR34K23). It is made up of three fragments: TE20246, TE20252 and TE20256 (Kaminker *et al*., 2002). The first one corresponds to 3553 bp of the internal region of the element and the others are two fragments of the long terminal repeats (LTRs) (240 and 284 bp long, respectively). A homology analysis with a library of all *D. melanogaster* TEs (see Section 2) revealed that this *Stalker* insert is the only one in the 80A–F chromosome region of the DGP, and that it is embedded in a large cluster of over 200 fragments of other TEs that extends over the 400 kb at the proximal end of the 3L sequence.

PCR analysis (see Section 2) indicated that this particular *Stalker* element is also present in all the lines from American and African samples, which suggests that it is fixed at this location in the *D. melanogaster* genome.

(iv) roo *at 20A*

A 2·3 kb *Sal*I–*Hin*dIII probe from the *roo*, internal to the *roo* fragment, hybridizes at 20A on chromosomes of all isogenic lines from the Beltsville population, as well as of y^1; cn^1 bw^1 sp^1. All *roo* copies annotated in the cytological region 20A–B in the DGP sequence are localized in a cluster of elements that extends over a region of approximately 30 kb (between positions

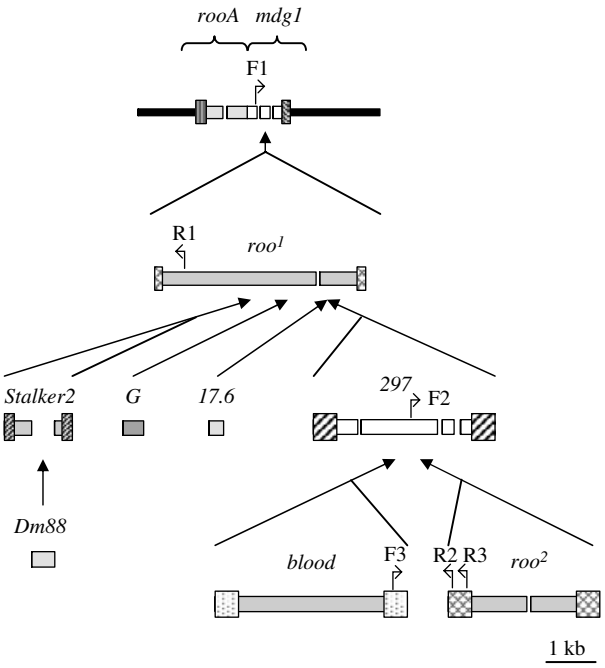

Fig. 2. Structure of a cluster of TEs found in chromosome band 20A in the *D. melanogaster* genome. TEs are depicted as boxes (LTRs as thicker boxes). Blank spaces represent deletions in the TEs as compared with the canonical sequences. Priming sites of PCR primers are shown as small arrows. Roo^1 and roo^2 are two insertions of the *roo* family.

X:21195105 and X:21226616; AE003122). Fig. 2 depicts a reconstruction of the organization of this cluster based on the order and genetic content of each fragment, as has been performed in maize (SanMiguel *et al*., 1998). We identified 24 traces that correspond to 10 elements from 9 different families. The oldest insertions are two incomplete *rooA* and *mdg1* elements. The latter is interrupted by one *roo* (roo^1), which has itself been the target for the insertion of the remaining seven elements.

The proposed reconstruction of the cluster is backed up by estimates of the ages of the insertions based on the relative levels of nucleotide divergence between (i) the elements found here and representatives of full-length active copies of each family (or canonical sequences; Kaminker *et al*., 2002) and (ii) the two LTRs of each element. In Table 1 it can be seen that the presumably older insertions (*rooA*, *mdg1*, roo^1) have indeed diverged from the canonical sequences more than relatively younger ones (*297*, *blood* and roo^2). It should be noticed that *rooA* elements are, in fact, the probable ancestors of the extant *roo* family, and were active in the *D. melanogaster* genome 2·5 million years ago (Kapitonov & Jurka, 2002). On the other hand, the large divergence estimates for the insertions identified as *Stalker2*, *Dm88* and *G* are not consistent with their having inserted into a relatively much younger copy of *roo* (roo^1;

https://doi.org/10.1017/S0016672305007548 Published online by Cambridge University Press

Table 1. *Genetic distance between the TEs found in a cluster at 20A and their respective canonical sequences*

TE	*d*	SE
rooA[a]	0·05	0·007
mdg1	0·11	0·008
roo[1]	0·03	0·002
Stalker2	0·25	0·022
Dm88	0·54	0·053
G	0·34	0·023
17·6	0·01	0·003
297	0·02	0·002
blood	0·00	0·000
roo[2]	0·00	0·000

d, pairwise genetic distance was estimated using the method of Tamura & Nei; SE, standard error calculated by bootstrap (500 reps).
[a] *rooA* elements are ancestors of the extant *roo* family (see text).

Table 1 and Fig. 2). There are two possible explanations for this: either these elements belong to other active TE families for which no canonical sequences have been identified yet (the above assignments correspond to the TE families in FlyBase and Repbase Update with which the sequences in question show the highest homology; see Section 2), or they have been inserted for a long period at some other location and have recently been integrated into *roo*[1] by gene conversion or a non-autonomous transposition-like process.

Although we cannot fully explain their origin, two pieces of evidence suggest that these insertions occurred independently and contradict the gene conversion hypothesis. First, *Stalker2* and *G* are flanked by two short repeated motifs (ATCCT and ACCACA, respectively). These repeats are thought to result from the duplication of the target site during the integration process (Berg & Howe, 1989). Second, there is no evidence for gene conversion in the *roo*[1] fragment between *Stalker2* and *G*. If they originated from older insertions into another locus, one would expect that the *roo* fragment between them would also be of the same age. However, this fragment shares the same level of divergence with the *roo* canonical sequence as the other *roo*[1] fragments ($d=0{\cdot}03$), which suggests that all derived from a single insertion.

Because of the large deletions they have experienced, comparisons of the divergence among LTRs of the same copy was only possible for those of *297*, *blood* and *roo*[2]. Consistent with their recent origin, the LTRs of these elements are all intact and exact copies of each other.

PCR-based estimates suggest that *roo*[1] is present in all individuals from Beltsville (7/7), and in 88% of the African sample (21/24). *roo*[2] could only be amplified from genomic DNA of the *y*[1]*; cn*[1] *bw*[1] *sp*[1] stock that was used as a control. This suggests that *roo*[1] is responsible for the hybridization of *roo* at 20A, and that it is fixed or at high frequency in the Beltsville population and African samples. The variation detected in the African sample does not exclude the presence of *roo*[1] in the individuals where it failed to amplify. The absence of a PCR product of the expected size could result from a mutation in one of the priming sites, a change of the length of the amplicon due to a deletion, or to the insertion of another element between the priming sites.

4. Discussion

(i) *A small fraction of elements in the* D. melanogaster *euchromatin are fixed*

Taken together with the population survey results reported by Charlesworth *et al.* (1992), and the previously reported fixation of an *S-element* at 87C (Maside *et al.*, 2002), these results indicate that a small fraction (3/2741) of the euchromatic insertions detected by *in situ* hybridization are fixed in the Beltsville natural population of *D. melanogaster*. These represent insertions at over 1500 independent polytene chromosome locations. In contrast, a much higher frequency of fixed elements has been found in recent studies using PCR amplification methods (Petrov *et al.*, 2003; Bartolomé & Maside, 2004). Of course, the latter methods rely on detection of TEs at sites where they are known to be located in the DGP sequence, and hence are much more biased towards high-frequency elements than the *in situ* surveys that used multiple haploid genomes to identify insertion sites. In addition, the greater average age of TEs in the non-recombining chromatin is reflected in a higher degree of fragmentation and degeneration of their nucleotide sequences (by accumulation of nucleotide substitutions, insertion and deletions) (Bartolomé & Maside, 2004). This may have diminished the resolution of the *in situ* hybridization technique, introducing a bias in the earlier studies towards the over-representation of more recent insertions, which are less likely to have had time to reach high frequencies or fixation.

(ii) *TEs accumulate in the non-recombining part of the* Drosophila *genome*

Roo and *Stalker* are fixed at 20A and 80B. These cytological positions roughly coincide with the distal boundaries of the regions of suppressed meiotic crossing over at the bases of the X and 2R. Interestingly, both elements are embedded in large clusters of other TEs (Sections 3.iii and 3.iv).

https://doi.org/10.1017/S0016672305007548 Published online by Cambridge University Press

Assuming that these clusters originated by the progressive accumulation of individual insertions (Fig. 2), our findings imply that, conservatively, at least all insertions in each cluster that are older than the ones studied here should also be fixed at these loci. Furthermore, the presence of recent TE insertions in some of the strains, such as *roo*^{2} and *blood* at 20A, which were found only in y^{1}; cn^{1} bw^{1} sp^{1} (see Section 3), suggests that the buildup of TEs is an ongoing process.

Given that these *roo* and *Stalker* insertions are far away from any known gene, and are surrounded and interrupted by other elements, it is difficult to conceive that they have any adaptive value to their hosts. These results illustrate the large scale of the accumulation of TEs in the *D. melanogaster* non-recombining genome outside the α-heterochromatin, and are in good agreement with previous evidence (Bartolomé & Maside, 2004). However, clusters of TEs can be found in the recombining euchromatin. In fact, the re-annotation of the DGP sequence revealed that the cluster of α and β repeated elements found close to the fixed *S-element* insertion at 87C (as well as in the pericentromeric heterochromatin) in *D. melanogaster* (Leigh Brown & Ish-Horowicz, 1981; Maside *et al*., 2002) correspond to tandemly duplicated fragments of two recently described LTR-retrotransposons: *Dm88* and *invader1* (C. Bergman, personal communication).

An analysis of the organization of the repetitive DNA in this cluster by means of Censor (see Section 1), suggests that the most parsimonious explanation is that it originated through a series of up to 17 tandem duplications of an initial repeat unit, which consisted of two incomplete copies of *Dm88* and *invader1*: a 1192 bp fragment of *Dm88* (two incomplete LTRs flanking a fragment of the internal coding region) and a complete *invader1* LTR (423 bp). At some point during this process a *1360* and a *micropia* element inserted into the cluster. The latter was included into a repeat unit and twice partially duplicated along with it. The average sequence divergence between repeat units is around 1% ($d = 0{\cdot}01 \pm 0{\cdot}001$), which suggests that the cluster has a recent origin, probably after the split between *D. melanogaster* and *D. simulans* (Leigh Brown & Ish-Horowicz, 1981). It could thus be hypothesized that the initial repeat unit became fixed along with the *S-element* insertion and the new organization of the *Hsp70* family in the *D. melanogaster* lineage (Bettencourt & Feder, 2001; Maside *et al*., 2002), and that it was later duplicated in tandem to attain its present state. We cannot exclude the existence of variation on the number of repeats of this cluster. The origin of this cluster is clearly different from that of those from the non-recombining heterochromatin described above, which supports the view that the fixation of TEs in the recombining portion of the *Drosophila* genome is unlikely, due to the efficacy of selection against them (Charlesworth *et al*., 1994*b*; Bartolomé & Maside, 2004).

(iii) *TE accumulation and chromatin structure*

Early cytogenetic studies described three chromatin domains in *Drosophila* polytene chromosomes: the highly polytenized and clearly banded euchromatin; the tightly packed, non-polytenized and intensively stained α-heterochromatin; and the polytenized but poorly banded β-heterochromatin, in the boundaries between the two (Ashburner, 1989).

The α-heterochromatin contributes to one-third of the *D. melanogaster* genome. It is localized around the centromeres of the X chromosome and the autosomes and makes up the whole of the Y chromosome. It consists of tandem arrays of short AT-rich satellite repeats, interspersed with large amounts of TE-derived DNA, which contribute to a large fraction of its total DNA (Kapitonov & Jurka, 2003). The α-heterochromatin has much lower gene density than the euchromatin, its densely packed chromatin structure is associated with gene silencing, and it harbours chromosome transmission functions, such as centromere formation and meiotic pairing (reviewed in Sullivan *et al*., 2001).

In most organisms, heterochromatin formation involves a number of physical and biochemical changes of the DNA and associated histones (i.e. histone hypoacetylation, histone H3-Lys9 methylation and cytosine methylation, as well as association with specific proteins such as HP1 and HP2) that result in a tightly condensed chromatin structure that severely limits the access of the transcription machinery to the DNA causing gene silencing (Maison & Almouzni, 2004). Recent studies in animal and plant model organisms suggest that these modifications are triggered by the RNA interference machinery which, in turn, specifically targets small interfering RNAs produced by repetitive sequences such as tandem repeats or TEs (Lippman *et al*., 2004). This evidence establishes a direct association between the accumulation of TEs and heterochromatinization. In agreement with this, insertions of the *1360* transposon family have been invoked as *cis*-acting determinants for heterochromatin formation along the *D. melanogaster* chromosome four (Sun *et al*., 2004).

(iv) *Evolutionary implications of the accumulation of TEs*

The observed accumulation of TEs around the centromeres is common to various organisms such as *Arabidopsis thaliana*, *D. melanogaster* and *Tetraodon nigrovidis* (Bartolomé *et al*., 2002; Kaminker *et al*.,

2002; Rizzon *et al*., 2002; Kapitonov & Jurka, 2003; Wright *et al*., 2003; Fischer *et al*., 2004). In *D. melanogaster*, this has been explained on the basis of suppression of meiotic recombination in these regions. Population genetics theory suggests that a lack of recombination permits a buildup of TEs in non-recombining regions because: (i) TEs are there less likely to cause deleterious chromosomal arrangements via ectopic exchange (Langley *et al*., 1988; Montgomery *et al*., 1991) and (ii) genetic interference between linked sites subject to selection (Hill & Robertson, 1966) reduces the efficacy of natural selection on TE insertions (Charlesworth *et al*., 1994*b*).

RNAi-mediated gene silencing and heterochromatin formation may have evolved in response to the accumulation of TEs on genomic regions where natural selection is unable to prevent their spread (Agrawal *et al*., 2003), reducing their ability to transpose and cause damage elsewhere in the genome. In this scenario, the β-heterochromatin in *D. melanogaster* would represent that part of the euchromatin that is in the initial stages of heterochromatinization, as a consequence of the suppression of recombination and concomitant buildup in TE numbers. Similarly, this process could contribute to the genetic degeneration and heterochromatinization of Y chromosomes. Evolutionary models suggest that the suppression of crossing over between proto-X and proto-Y chromosomes near sex-determining regions is favoured by natural selection, since genetic linkage between a newly evolved sex-determining factor and other loci with antagonistic fitness effects reduces the fraction of gametes with the wrong allele combinations (Bull, 1983; Rice, 1987). Accordingly, recent studies on proto-Y chromosomes in a variety of organisms including the papaya plant (*Carica papaya*) and vertebrates such as the threespined stickleback (*Gasterosteus aculeatus*) and the medaka (*Oryzias latipes*) fishes, reveal reduced crossing over and an enrichment in repetitive sequences and transposons around the sex-determining regions (reviewed by Charlesworth, 2004). An important difference between these two processes is that the genetic losses associated with the degeneration of a proto-Y chromosome can be compensated for by increased expression of the genes on the proto-X. In the case of the β-heterochromatin, essential genes need to escape silencing before complete heterochromatinization, either by moving into the recombining euchromatin, or by preservation of their expression in a heterochromatic environment (Weiler & Wakimoto, 1995).

We thank C. Bergman for discussions and C. Bartolomé and two anonymous reviewers for helpful comments on earlier versions of the manuscript. B.C. was supported by the Royal Society and X.M. by a Marie Curie TMR-Fellowship.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., *et al*. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K. & Mukherjee, S. K. (2003). RNA interference: biology, mechanism, and applications. *Microbiology and Molecular Biology Reviews* **67**, 657–685.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Ashburner, M. (1989). *Drosophila: A Laboratory Handbook*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Bartolomé, C. & Maside, X. (2004). The lack of recombination drives the fixation of transposable elements on the fourth chromosome of *Drosophila melanogaster*. *Genetical Research* **83**, 91–100.

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.

Begun, D. J. & Aquadro, C. F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.

Berg, D. E. & Howe, M. M. (eds.) (1989). *Mobile DNA*. Washington, D.C.: American Society for Microbiology.

Bettencourt, B. R. & Feder, M. E. (2001). *Hsp70* duplication in the *Drosophila melanogaster* species group: how and when did two become five? *Molecular Biology and Evolution* **18**, 1272–1282.

Bull, J. J. (1983). *Evolution of Sex Determining Mechanisms*. Menlo Park, CA: Benjamin Cummings.

Charlesworth, B. (2004). Sex determination: primitive Y chromosomes in fish. *Current Biology* **14**, R745–R747.

Charlesworth, B., Lapid, A. & Canada, D. (1992). 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., Jarne, P. & Assimacopoulos, S. (1994*a*). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin. *Genetical Research* **64**, 183–197.

Charlesworth, B., Sniegowski, P. D. & Stephan, W. (1994*b*). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220.

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

Finnegan, D., Rubin, G., Young, M. & Hogness, D. (1978). Repeated gene families in *Drosophila melanogaster*. *Cold Spring Harbor Symposia on Quantitative Biology* **42**, 1053–1063.

Fischer, C., Bouneau, L., Coutanceau, J. P., Weissenbach, J., Volff, J. N. & Ozouf-Costaz, C. (2004). Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. *Gene* **336**, 175–183.

Franchini, L. F., Ganko, E. W. & McDonald, J. F. (2004). Retrotransposon–gene associations are widespread among *D. melanogaster* populations. *Molecular Biology and Evolution* **21**, 1323–1331.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for

Windows 95/98/NT. *Nucleic Acids Symposia Seriès* **41**, 95–98.

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

Hoogland, C. & Biémont, C. (1996). Chromosomal distribution of transposable elements in *Drosophila melanogaster*: test of the ectopic recombination model for maintenance of insertion site number. *Genetics* **144**, 197–204.

Jurka, J. (2000). Repbase Update: a database and an electronic journal of repetitive elements. *Trends in Genetics* **16**, 418–420.

Jurka, J., Klonowski, P., Dagman, V. & Pelton, P. (1996). CENSOR – a program for identification and elimination of repetitive elements from DNA sequences. *Computers & Chemistry* **20**, 119–121.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., *et al*. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* **3**, research0084. 0081-0020.

Kapitonov, V. V. & Jurka, J. (2002). ROOA, an ancient family of endogenous retrovirus ancestral to the modern ROO family. *Repbase Reports* **2**, 9.

Kapitonov, V. V. & Jurka, J. (2003). Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences of the USA* **100**, 6569–6574.

Kazazian, H. H., Jr (2004). Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632.

Kidwell, M. G. & Lisch, D. R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, 1–24.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

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.

Leigh Brown, A. & Ish-Horowicz, D. (1981). Evolution of the 87A and 87C heat-shock loci in *Drosophila*. *Nature* **290**, 677–682.

Lewis, S., Searle, S., Harris, N., Gibson, M., Iyer, V., *et al*. (2002). Apollo: a sequence annotation editor. *Genome Biology* **3**, research0082.0081–0082.0014.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., *et al*. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476.

Maison, C. & Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nature Reviews Molecular Cell Biology* **5**, 296–304.

Maside, X., Assimacopoulos, S. & Charlesworth, B. (2000). Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. *Genetical Research* **75**, 275–284.

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.

McDonald, J. F. (ed.) (2000). *Transposable Elements and Genome Evolution*. Dordrecht: Kluwer Academic.

Montgomery, E. A., Huang, S. M., Langley, C. H. & Judd, B. H. (1991). Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* **129**, 1085–1098.

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.

Potter, S. S., Brorein, W. J., Jr, Dunsmuir, P. & Rubin, G. M. (1979). Transposition of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. *Cell* **17**, 415–427.

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914.

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.

Rozas, J., Sánchez-Delbarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**, 43–45.

Sullivan, B. A., Blower, M. D. & Karpen, G. H. (2001). Determining centromere identity: cyclical stories and forking paths. *Nature Review Genetics* **2**, 584–596.

Sun, F.-L., Haynes, K., Simpson, C. L., Lee, S. D., Collins, L., Wuller, J., Eissenberg, J. C. & Elgin, S. C. R. (2004). *cis*-Acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome four. *Molecular and Cellular Biology* **24**, 8210–8220.

Weiler, K. S. & Wakimoto, B. T. (1995). Heterochromatin and gene expression in *Drosophila*. *Annual Review of Genetics* **29**, 577–605.

Wright, S. I., Le, Q. H., Schoen, D. J. & Bureau, T. E. (2001). Population dynamics of an Ac-like transposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* **158**, 1279–1288.

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.