Genet. Res., Camb. (2008), **90**, *pp.* 243–252. © 2008 Cambridge University Press
doi:10.1017/S0016672308009312 Printed in the United Kingdom

The *hobo*-related elements in the *melanogaster* species group

MAURO DE FREITAS ORTIZ[1] AND ELGION LUCIO SILVA LORETO[2]*
[1] *Curso de Ciências Biológicas, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil*
[2] *Departamento de Biologia, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil*

(*Received 1 November 2007 and in revised form 30 January 2008*)

Summary

The *hobo*-related sequences (*h*RSs) were considered as degenerate and inactive elements until recently, when one mobilizable copy was described. Using this sequence as the initial seed to search for homologous sequences in 12 available *Drosophila* genomes, in addition to searching for these sequences by PCR and Southern blot in nine other species, we found homologous sequences in every species of the *Drosophila melanogaster* species subgroup. Some evidence suggests that these non-autonomous sequences were kept mobilizable for at least 0·4 million years. Also, some very short sequences with miniature inverted-repeat transposable element (MITE) characteristics were found among these *h*RSs. These *h*RSs and their 'MITE-like' counterparts could provide a good example of the steps proposed in models that describe the MITEs origin.

1. Introduction

Transposable elements (TEs) are present in almost all species and, in many organisms, they contribute to a considerable portion of the genome. Nevertheless, the biological importance of TEs has not yet been adequately understood. Hypotheses about the roles played by TEs range from genomic parasites to symbiontic or mutualistic compounds (Kidwell & Lisch, 2001; Brookfield, 2005). Also, the TEs 'life cycle' in host genomes has likewise been a matter of discussion, and the way that TEs invade, are maintained, are controlled, are domesticated, or are even lost in genomes is not fully comprehended (Le Rouzic & Capy, 2005). TEs are extremely heterogenic in composition, molecular features and transpositional mechanisms. Class I elements are replicated by an RNA intermediary, and class II elements use DNA as a mediator for transposition. In both classes, there are autonomous elements that produce the necessary enzymes for transposition and non-autonomous elements that use the enzymes produced by autonomous elements (Capy *et al*., 1998).

* Corresponding author. Departamento de Biologia, CCNE, prédio 16A-Campus, Universidade Federal de Santa Maria, CEP 97105-900, Santa Maria, RS, Brazil. e-mail: elgion@base.ufsm.br

The *hobo* element is a class II TE, and belongs to the *hAT* superfamily, which is widely distributed in plants, animals and fungi (Calvi *et al*., 1991). While *hobo* itself is restricted to the *melanogaster* group of *Drosophila* (Daniels *et al*., 1990), some *hobo*-like elements have been found in several Diptera species, like *Musca domestica* (Atkinson *et al*., 1993), in some Lepidoptera species (DeVault & Narang, 1994; Borsatti *et al*., 2003) and in different tephritids (Handler & Gomez, 1996; Torti *et al*., 2005).

In *Drosophila*, *hobo* is found in three forms. The first form is the complete element, or canonical *hobo*, about 3 kb long, with 12 bp of terminal inverted repeats (TIRs) and a gene with the potential to encode a transposase enzyme. It is known that in *Drosophila melanogaster* the complete *hobo* element is active and capable of producing the hybrid dysgenesis syndrome (Blackman *et al*., 1989; Yannopoulos *et al*., 1987). The second form corresponds to defective elements. They exhibit sequences that are very similar to those of the canonical *hobo*; however, deletions of variable length in the internal portion of the element are found. Complete *hobo* elements and their deleted derivatives are present only in *D. melanogaster* and its sibling species, *Drosophila simulans* and *Drosophila mauritiana* (Anxolabehere *et al*., 1988). In *D. melanogaster* and *D. simulans*, these sequences are present in

some strains (called H), and absent in others (denominated E strains, for 'Empty'). The canonical *hobo* and its deleted derivatives are supposed to be recent acquisitions of the *D. melanogaster* genome (Anxolabehere *et al*., 1988; Boussy & Daniels, 1991; Simmons, 1992). Finally, the third form is described as a *hobo* relic or *hobo*-related sequence (*h*RS). In comparison with the canonical *hobo*, the characterized sequences have around 80% similarity, with multiple rearrangements, and they are not able to code for a functional transposase. The relics are present in all strains of the *melanogaster* subgroup species and the *montium* subgroup species (Daniels *et al*., 1990). The earliest analyses suggested that these sequences correspond to an ancient *hobo* element present in the *melanogaster* group ancestral. The sequences are supposed to be inactive (Lim, 1988; Daniels *et al*., 1990; Galindo *et al*., 2001).

Recently, we described a mobilizable *hobo* relic in *D. simulans*, isolated in a *de novo* mutation that occurred in a hypermutable strain (Torres *et al*., 2006). This *h*RS element, called *hobo*^{va}, is 1·2 kb long, defective, with roughly 82% similarity at DNA level with the canonical *hobo*. However, they have extremely conserved 200 bp in each subterminal region, which are significantly similar to the canonical *hobo*. The inner region of this element is almost completely composed of A and T arranged as imperfect microsatellites. It has also been suggested that this relic *hobo* could be mobilizable by the canonical element. Furthermore, the presence of sequences similar to *hobo*^{va} in *Drosophila sechellia* suggested that these relic *hobo* elements could have been kept mobilizable since the divergence time between *D. simulans* and *D. sechellia* (0·4 million years ago (MYA)).

In the present paper, we describe the presence of *hobo*^{va} homologous sequences (*hobo*^{vahs}) in various species of the *melanogaster* group and we discuss the possibilities of the origin and maintenance of these non-autonomous elements. Moreover, we have shown 'shrinking' events of some *hobo*^{vahs} sequences that could be the origin of some related miniature inverted-repeat TEs (MITEs).

2. Material and methods

(i) *Fly stocks*

The PCR search for sequences homologous to *hobo*^{va} in genomic DNA was carried out in the followings species: *D. sechellia* (the Seichelles island, 1985; coll. J. David), *D. mauritiana* (the Mauritius island, 1988; coll. J. David), *Drosophila santomea* (São Tome, Parque Obo; coll. D. Lachaise), *D. melanogaster*, *Drosophila teissieri* (STO384.3 Uganda, Kibale Forest; coll. D. Lachaise), *Drosophila ananassae* (Florianópolis, Brazil, 2005; coll. M. Gottschakk), *Drosophila malerkotliana* (Florianópolis, Brazil, 2005; coll. M. Gottschakk), *Drosophila kikkawai* (Florianópolis, Brazil, 2005; coll. M. Gottschakk) and *D. simulans* (dpp strain, Eldorado, RS, Brazil, 1989). The source and collection date of stocks are given in parentheses.

(ii) *Genome search*

Initially, the search for sequences homologous to *hobo*^{va} (Torres *et al*., 2006) was carried out in the genomes of the following species: *D. ananassae*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila willistoni*, *Drosophila mojavensis*, *Drosophila virilis*, *Drosophila grimshawi*, *D. simulans*, *Drosophila yakuba*, *D. sechellia*, *D. melanogaster* and *Drosophila erecta*, recently available and analysed by Clark *et al*. (2007). The search was performed using the BLAT (Kent, 2002) tool available in the UCSC Genome Browser Database (Karolchik *et al*., 2003), with the assistance of the UCSC Table Browser data retrieval tool (Karolchik *et al*., 2004). All hits were analysed and 1 kb on top of both the sides of the hit was retrieved for subsequent alignments and analyses of these sequences. Searches were also performed using the FlyBase BLAST Service (http://flybase.bio.indiana.edu/blast/) and the NCBI Traces Archives using the Mega BLAST tool (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml) (Altschul *et al*., 1997) with the default parameters.

The initial sequences used as query were the *D. melanogaster* canonical *hobo* (M69216) and the *D. simulans hobo*^{va} (AY764286). Subsequently, all retrieved sequences were also used as query until no additional new sequences were obtained. The retrieved sequences were classified using the following criteria: (i) putatively mobilizable sequences (PMS) – in these sequences, TIRs and sometimes target sequence duplications (TSDs) were present; (ii) incomplete sequences – without one or both TIRs; and (iii) degenerate sequences, with similarity $<80\%$. The degenerate sequences were not analysed but can be made available on request.

The structural features that allow several *hobo*^{vahs} to be classified as PMS are the extremely conserved *hobo* TIRs (identical to canonical *hobo*) and also a well conserved 200 bp long component in each subterminal region of the element. These characteristics do not guarantee that these elements will be mobilizable, and it is only possible to show such a property for a specific sequence in an experimental way. Furthermore, some alterations in the TIRs and subterminal sequences can occur even when the element maintains itself mobilizable. In this perspective, our estimates are conservative and correspond only to the elements that showed characteristics suggesting that they are able to be mobilizable.

The genome assemblies used correspond to the final versions released (Clark *et al*., 2007). The contigs and assemblies names, the sequences coordinates and the length of the sequences used can be seen in Tables 1S and 3S (supplementary material). Also, the alignment of the complete dataset is available in the supplementary material.

(iii) *PCR amplification and sequencing*

The primers used to specifically amplify the *hobo*^{vahs} were: hva1s (forward), 5′-cataacggaagggtagagaag-3′; hva2as (reverse), 5′-cgtccacccgataaacactc-3′; Vanew1 (forward), 5′-caattttgwgtgcgggtgcy-3′; Vayak (reverse), 5′-gaactgcagcaagccaccgg-3′. These primers were designed using the sequences obtained in the genome search and they anneal, respectively, at positions 200–219, 1169–1188 and 50–70 using *hobo*^{va} as a reference sequence and Vayak anneal at nucleotide positions 1540–1560 using the sequence 6yak VA as a reference. This last sequence corresponds to the one obtained in the genomic search in the *D. yakuba* genome. Both reference sequences can be obtained in the supplementary material, in an alignment file (hobova_alignament.aln). The obtained amplicons correspond to a single band, with roughly 1 kb, while short elements have been observed in the cloned sequenced (see below). These primer sets, in different combinations, anneal to all the sequences retrieved from the genome search. PCR reactions were performed in 25 μl volumes using approximately 20 ng of template DNA, 20 pmol of each primer, 1·5 mM $MgCl_2$, 50 μM of each nucleotide and 1 unit of *Taq* DNA Polymerase (Invitrogen). After an initial denaturation step of 4 min at 95 °C, 35 cycles consisting of 40 s denaturation at 95 °C, 40 s annealing at 55 °C and 1 min extension at 72 °C were carried out. An additional 5 min extension step at 72 °C was performed after the last cycle. The PCR products were cloned into pCR-TOPO plasmid (Invitrogen). DNA sequencing was performed directly from the purified plasmids in a MegaBACE 500 automatic sequencer. The dideoxy chain-termination reaction was implemented using the DYEnamic ET kit (GE Healthcare). The sequences were then submitted to a 'confidence consensus' analysis using the Staden Package Gap 4 program (Staden, 1996). *D. santomea* sequences have been deposited in GenBank under the accession numbers DQ840031–DQ840035 and DQ823386, and *D. mauritiana* sequences under accession numbers DQ840036–DQ840038.

(iv) *Southern blot analyses*

Genomic DNA was obtained as described by Sassi *et al*. (2005). Approximately 7 μg of DNA samples were digested with *Eco*RI (Invitrogen), separated by electrophoresis on 1 % agarose gels and transferred to nylon membranes (HybondN+, Amersham Biosciences). The membranes were hybridized with probes corresponding to PCR fragments of *D. simulans hobo*^{va} or *D. santomea hobo*^{vahs}, amplified from plasmids used in the sequencing analyses described below. The divergence between these sequences is 24 %. To label and detect nucleic acids, an AlkPhos Direct Labeling and Detection System (Amersham Bioscience) kit was used according to the kit protocol.

(v) *Sequence analyses*

The following software was used in the sequence analyses: GENEDOC version 2.6.001 (Nicholas & Nicholas, 1997) for sequence editing and visualization; Einverted from the EMBOSS suite (http://emboss.sourceforge.net/) for TIR identification; Clustal W (Thompson *et al*., 1994) for sequence alignment; and MEGA version 3.1 (Kumar *et al*., 2001) for phylogenetic analysis. In the Maximum Parsimony analysis, the best tree was searched using close-neighbour interchange, with parameter values and random addition of sequences (ten replications) to produce the initial trees. In the Neighbour-Joining (NJ) method, the Kimura two-parameter model of nucleotide substitution (Kimura, 1980) was used to construct the distance matrices. In both analyses, bootstrap tests with 1000 replications were performed to assess the support value for each internal branch of the trees. The phylogenetic analysis was carried out with the junction of 1–200 nucleotides of the 5′ subterminal region and 1152–1220 nucleotides of the 3′ subterminal region (using the *hobo*^{va} sequence as a reference) because these are the more conserved regions, producing a more consistent alignment. The total length of the alignment corresponds to 290 bp and the gaps were included in the analysis.

3. Results

(i) *Search for homologous* hobo^{va} *by PCR and Southern blot*

Analyses by PCR have shown sequences homologous to *hobo*^{va}, as described by Torres *et al*. (2006), only in species of the *melanogaster* subgroup. As can be seen in Table 1, amplicons of *hobo*^{vahs} were obtained from *D. sechellia*, *D. mauritiana*, *D. simulans*, *D. melanogaster*, *D. santomea* and *D. teissieri*, which belong to the *melanogaster* subgroup, but no amplification was obtained from species of other subgroups of the *D. melanogaster* species group (*D. ananassae*, *D. malerkotliana* and *D. kikkawai*) (Clark *et al*., 2007). Southern blot analyses confirmed the PCR results. As can be seen in Fig. 1*A*, in which *hobo*^{va} of *D. simulans* was used as a probe, numerous hybridization bands were observed in *D. sechellia*,

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

Table 1. *PCR results with different primer combinations*

Primer combination	mau[a]	sim[b]	sec[c]	mel[d]	san[e]	tei[f]	ana[g]	mal[h]	kik[i]
Vanew1/hva2as	+	+	+	+	+	+	−	−	−
Vanew1/Vayak	+	+	+	+	+	+	−	−	−
hva1s/Vayak	+	+	+	+	+	+	−	−	−
hva1s/hva2as	+	+	−	+	−	−	−	−	−

[a]*D. mauritiana*, [b]*D. simulans*, [c]*D. sechellia*, [d]*D. melanogaster*, [e]*D. santomea*, [f]*D. teissieri*, [g]*D. ananassae*, [h]*D. malerkotliana*, [i]*D. kikkawai*.

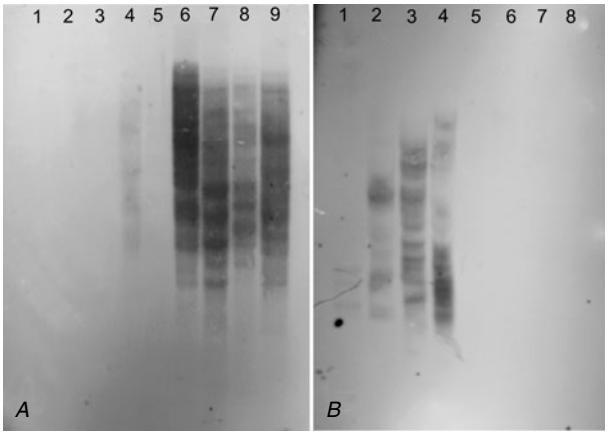

Fig. 1. Southern blot of genomic DNA digested with *Eco*RI. (*A*) Membranes were hybridized with a probe to *hobo*^{va} of *D. simulans*. 1, *D. kikkawai*; 2, *D. ananassae*; 3, *D. malerkotliana*; 4, *D. teissieri*; 5, *D. santomea*; 6, *D. melanogaster*; 7, *D. sechellia*; 8, *D. mauritiana*; 9, *D. simulans*. (*B*) Membranes were hybridized with a probe to *hobo*^{vahs} of *D. santomea*. 1, *D. simulans*; 2, *D. melanogaster*; 3, *D. teissieri*; 4, *D. santomea*; 5, *D. ananassae*; 6, *D. kikkawai*; 7, *D. malerkotliana*; 8, *D. mediostriata* (negative control).

D. mauritiana, *D. simulans* and *D. melanogaster*. A weak signal was seen in *D. teissieri* and no hybridization signal was observed outside the *melanogaster* subgroup such as *D. ananassae*, *D. malerkotliana* and *D. kikkawai*. When *hobo*^{vahs} of *D. santomea* was used as a probe (Fig. 1*B*), hybridization signals were seen in *D. teissieri* and *D. santomea*, while faint bands occurred in *D. melanogaster* and *D. simulans*.

Together, the PCR and Southern blot analyses show that the *hobo*^{vahs} are restricted to the *melanogaster* subgroup.

(ii) *Cloning and sequencing of* hobo^{vahs}

We have cloned and sequenced some elements for those species that have *hobo*^{vahs} but the genome sequences are not available. Three sequenced clones of *D. mauritiana hobo*^{vahs} were around 1·1 kb long and exhibited 90% general similarity to *hobo*^{va} of *D. simulans*. One clone showed a short *hobo*^{vahs} sequence with 251 bp.

The sequenced *D. santomea* clones, eight in total, deserve special attention due to their very short length (391 bp) and because they are almost identical in sequence. In the 5′ subterminal region of these elements, a 180 bp region exhibited 70% similarity to the *D. simulans hobo*^{vahs}, and in the 3′ end, the last 70 bp had 82% similarity. As in *hobo*^{va}, the middle region is AT-rich.

4. Genomic search

A search for homologous sequences in the 12 available *Drosophila* genomes, which represent diverse *Drosophila* groups, demonstrated the presence of *hobo*^{vahs} only in the *melanogaster* group (*D. melanogaster*, *D. simulans*, *D. yakuba*, *D. sechellia* and *D. erecta*).

The copy number of *hobo*^{vahs} varied highly among species. As shown in Table 2, 12 copies were found in the *D. melanogaster* genome. These copies were PMS, but only five (42%) showed the TSDs. The *hobo*^{vahs} copies described here do not correspond to those *hobo* elements previously annotated in the *D. melanogaster* genome (Kaminker *et al.*, 2002; Quesneville *et al.*, 2005). In *D. simulans*, a significantly higher copy number was found (147 copies), of which 55 copies were incomplete and 92 were PMS. There were 72 (78%) PMS copies in which we were able to find TSDs. In *D. yakuba*, 70 copies were found, of which 28 were incomplete sequences, along with 42 PMS. Among the PMS detected for *D. yakuba*, TSDs were observed in 37 (88%). In *D. sechellia*, 60 copies were found, with 53 being PMS and of which 73% possessed TSDs. In *D. erecta*, only one copy was found, and it was a PMS with TSD. For the genomes to which the chromosome assemblies are currently available, we were able to analyse the distribution of *hobo*^{vahs} copies in the chromosomes. As can be seen in Table 2, no preferential insertions were observed in the chromosome arms of *D. melanogaster*, *D. simulans* or *D. yakuba*.

The presence of 8 bp direct duplications of the insertion site (TSDs) typically characterizes *hobo*

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

Table 2. *Copy number and distribution of hobo^{vahs} sequences in the genomes*

	3L	3R	2R	2L	X	U	Random	Total
D. melanogaster								
PMS (TSDs)	2 (0)	4 (2)	1 (0)	2 (0)	1 (0)	0 (0)	1 (1)	11 (3)
D. simulans								
Incomplete sequences	5	4	4	8	9	16	9	55
PMS (TSDs)	17 (14)	23 (18)	21 (17)	9 (9)	12 (10)	3 (1)	7 (3)	92 (72)
Total	22	27	25	17	21	19	16	147
Drosophila yakuba								
Incomplete sequences	2	9	0	4	1	10	3	29
PMS (TSDs)	4 (4)	8 (6)	2 (2)	10 (10)	4 (5)	12 (9)	2 (1)	42 (37)
Total	6	17	2	14	5	22	5	71
D. sechellia[a]								
Incomplete sequences								7
PMS (TSDs)								53 (39)
Total								60
D. erecta[a]								
PMS (TSDs)								1 (1)

PMS = copy number of putatively mobilizable sequences.
(TSD) = copy number of *hobo^{vahs}* that showed target sequence duplication.
2L, 2R, 3L, 3R and X = chromosome arms; U and Random = chromosome position not identified.
[a] In the current genome assembly for these species, the chromosome assemblies are not available.

mobilization (McGinnis *et al*., 1983). The identification of TSDs in a significant number of copies (42–88 %) – together with high similarity between some copies – is suggestive of recent mobilization.

We have analysed the integration specificity of *hobo^{vahs}* elements through nucleotide frequency estimation in the TSDs. The TSDs observed in the different species are very similar. Nucleotides in positions 2 and 7 were the most information-rich. Thymidine was the most common nucleotide in position 2 and adenine the most abundant nucleotide in the seventh position. The consensus sequences observed were: *D. simulans* (GTNCGNAC), *D. sechellia* (GTNCNNAC), *D. yakuba* (GTNCNNAT) and *D. melanogaster* (GTNCNNAC) (Table 4 in the supplementary material).

(i) *Phylogenetic analysis*

For phylogenetic analysis, we used the PMS obtained in the genome search (200 sequences). Also, we used three partial sequences from *D. mauritiana* and eight from *D. santomea* (sequenced in this work).

The phylogenetic analysis showed the presence of two *hobo^{vahs}* clusters. As seen in Fig. 2, the cluster called 'A', which was statistically well supported, was formed only by sequences from *D. simulans* and *D. sechellia* and by two *D. melanogaster* sequences found in a polytomy. The divergence observed between the subclusters formed by *D. simulans* and *D. sechellia* sequences ranged from 0·0 to 18·7 % (3·7 % on average). When the *D. melanogaster* sequences were included, the divergences varied from 0·0 to 31·0 % (4·0 % on average). As can also be seen in Fig. 2, several *D. simulans* and *D. sechellia* sequences exhibited the presence of *Xho*I restriction sites in one or both extremities. Since the length of these sequences was normally 1·1 kb, the distance between the *Xho*I sites was around 0·7 kb, and these sequences correspond to 'deleted *hobo* sequences' described in the Southern blot analysis as defective canonical *hobo*, according to Boussy & Daniels (1991), Periquet *et al*. (1994) and Loreto *et al*. (1998). Cluster B showed a higher internal divergence, varying from 0·0 to 31·6 % (16·6 % on average). This cluster is represented mainly by sequences from *D. yakuba*, *D. santomea* and *D. erecta*. However, sequences from *D. mauritiana* and *D. melanogaster* are also present. The overall divergence observed in the *hobo^{vahs}* sequences from clusters A and B varied from 0·0 to 37·4 % with an average of 14·7 %.

5. Discussion

(i) hobo^{vahs} *are disseminated in the* D. melanogaster *subgroup*

*h*RSs or *hobo* relics were thought to be vestigial and inactive sequences of previous genome invasions by *hobo* elements in the *Drosophila* genome (Lim, 1988; Daniels *et al*., 1990). Nevertheless, Torres *et al*. (2006) have shown that one *h*RS, the *hobo^{va}*, is mobilizable and probably has been kept transpositionally active for 0·4 million years (MY), which corresponds to the divergence time between *D. simulans* and *D. sechellia*. This assumption was suggested since a similar

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

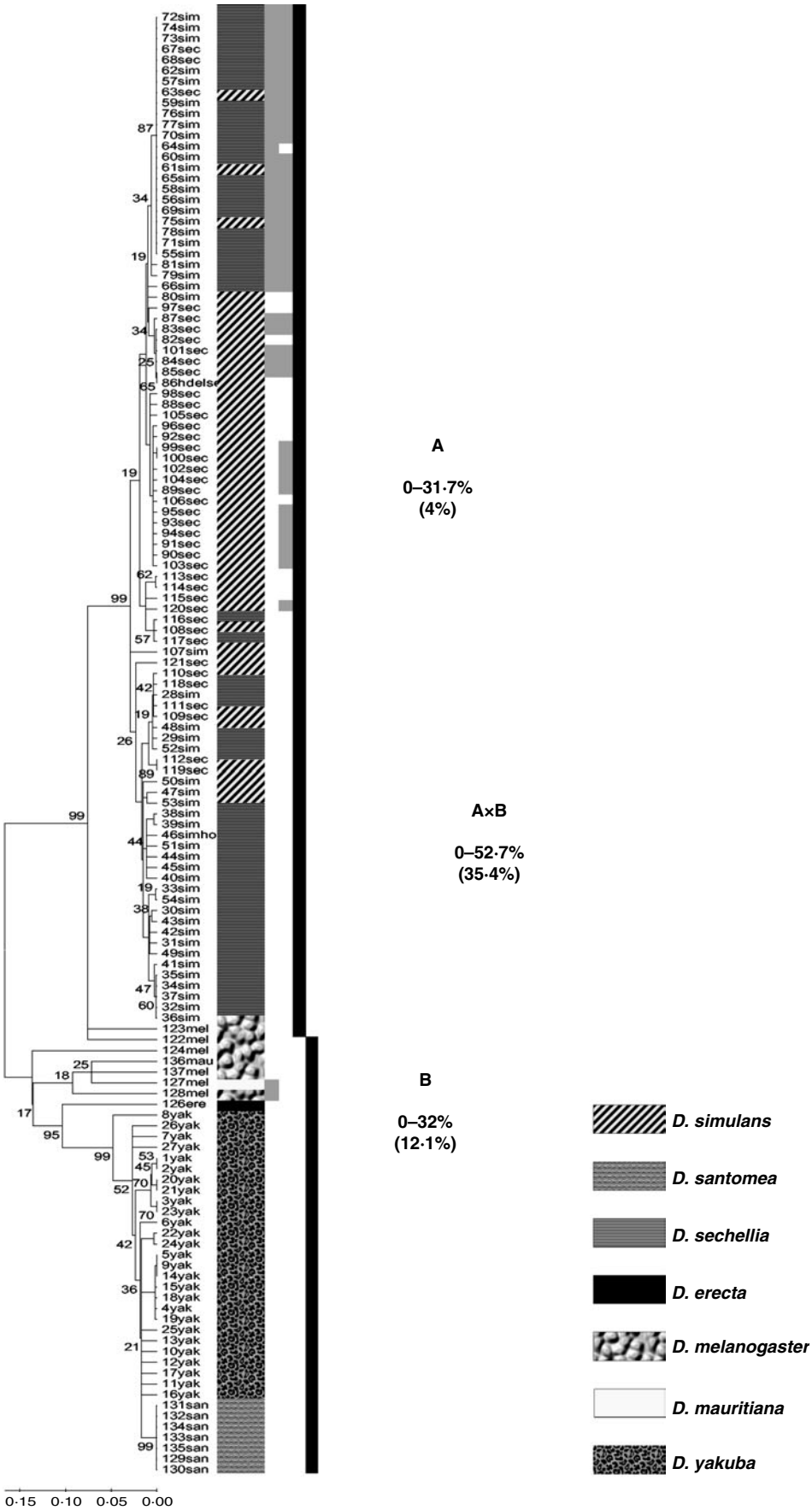

Fig. 2. For legend see opposite page.

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

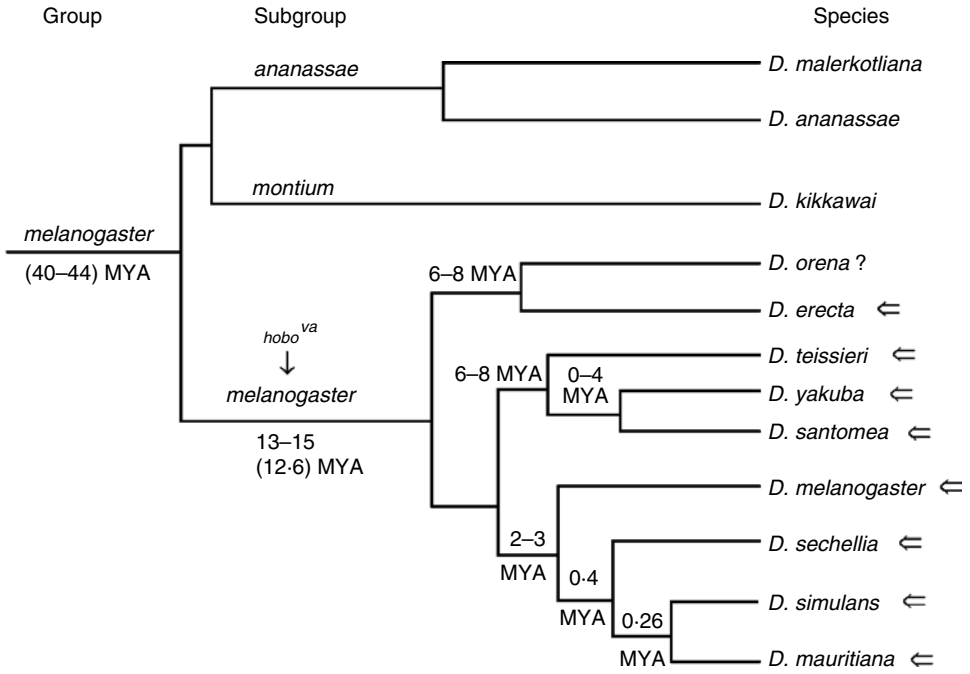

Fig. 3. Phylogenetic tree for the *D. melanogaster* species group with emphasis on the *D. melanogaster* subgroup. The divergence estimate in million years ago (MYA) are from Lachaise and Silvain (2004) and Tamura *et al.* (2004) (in parentheses). Arrows on the left indicate the presence of *hobo*^{vahs} and the question mark indicates that so far it has not been possible to establish whether the species possess the sequence.

sequence was also observed in this last species. Our results reinforce this supposition since several different *hobo*^{vahs} sequences (clusters A) are shared by *D. simulans* and *D. sechellia*, showing that these sequences are present in the ancestor of these species and, given their structural characteristics, are maintained active since then (Fig. 3).

The presence of sequences with the same *hobo*^{va} characteristics in every species of the *melanogaster* subgroup could be explained in two different ways: (i) *hobo*^{vahs} elements arose in the *melanogaster* subgroup ancestor, around 13–15 MYA; it was vertically transmitted and was kept mobilizable since then; (ii) it could be supposed that different *hobo*^{vahs} elements have originated independently, in different species, starting from diverse *hobo* elements. In this case, it would be interesting to understand why the same structural characteristics have arisen independently, in different times, in these elements. These possibilities are not mutually exclusive.

The fact that a significant portion of *hobo*^{vahs} described in this work shows high nucleotide similarity, alongside with the observation that part of them preserves intact TIRs and conserved TSDs, constitutes suggestive evidence that these sequences were kept mobilizable. Currently we are not able to discriminate the evolutionary time in which these sequences are maintained mobilizable. One possibility is 13–15 MY, if the element arose in the *melanogaster* group ancestor. However, the presence of very similar sequences in *D. simulans* and *D. sechellia* strongly suggests that these non-autonomous sequences were kept mobilizable at least for 0·4 MY.

The continued presence, over a prolonged evolutionary time, of mobilizable non-autonomous elements 'parasitizing' their TE master copies has rarely been reported and is intriguing. Analyses of the *D. melanogaster* genome have shown remarkable sequence homogeneity among copies of TEs (Bowen & McDonald, 2001; Kaminker *et al.*, 2002; Lerat *et al.*, 2003; Sanchez-Gracia *et al.*, 2005). Lerat *et al.* (2003) have proposed that this minute divergence may have resulted from a rapid turnover that eliminated TE copies as soon as they became inactive. The high similarity observed among the *hobo*^{vahs} copies – reinforced by scattered chromosome distribution over

Fig. 2. Phylogenetic analysis of *hobo*^{vahs} nucleotide sequences. Neighbour-Joining tree with Kimura two-parameter distances. Numbers above branches are percentage bootstrap values based on 1000 replications. The central rectangles with mark patterns correspond to species names identified in the figure. The dark line on the left identifies the different clusters. The percentages correspond to the nucleotide divergence within the cluster (minimum and maximum) and the average (in parentheses) or on the far right the divergence between the clusters. The grey squares on the side of the dark line represent the occurrence of *Xho*I restriction sites in one extremity or in both (respectively).

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

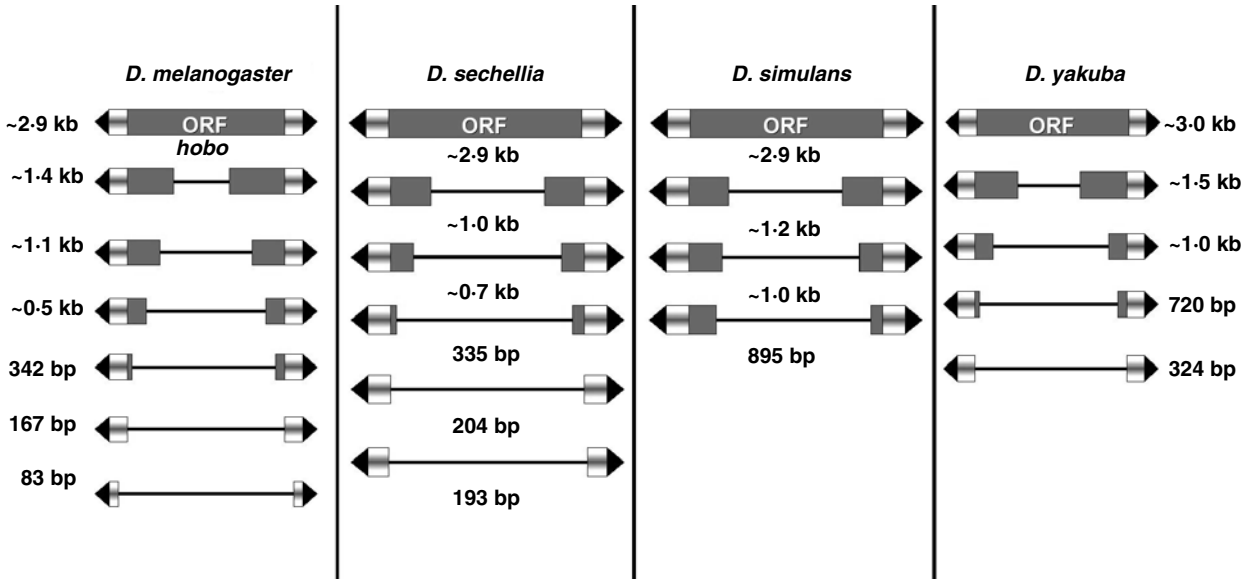

Fig. 4. Schematic representation of *hobo*^{vahs} MITEs. Arrowheads = TIR; white block = conserved 3′ and 5′ subterminal regions; grey block = the inner sequence. The complete *hobo* elements of about 3 kb produce elements with internal deletions while maintaining 3′ and 5′ subterminal regions. These elements exhibit sizes from 1·5 to 1·0 kb. The inner sequences are AT-rich. Finally, short elements are generated (700–83 bp) showing only the conserved subterminal regions and TIRs. These shorter sequences correspond to MITEs. The file with the alignment of these sequences can be found in the supplementary material.

all chromosomes arms – is in concordance with the Lerat *et al.* (2003) hypothesis of high TE turnover. From this perspective, the *hobo*^{vahs} relic sequences could be kept in the genomes of these *Drosophila* species exactly because they are kept mobilizable, avoiding losses in the turnover process.

It is notable that the number of PMS found in the analysed genomes is higher than the not mobilizable ones. As our analyses were performed in final versions of the genome assemblies released, probably the *hobo* sequences described here reflect very well the *h*RSs present in the euchromatic regions of these genomes. However, it is possible that degenerated copies of *h*RSs and PNM *hobo*^{vahs} copies can be more abundant in the heterochromatic regions that are under-represented in the available versions of genome assemblies (Clark *et al.*, 2007).

In order to be kept mobilizable for such a long time, a non-autonomous element necessarily requires a transposase source. As for the transposase source for *hobo*^{vahs}, the canonical *hobo* is the most likely supplier. The consensus sequences of TSDs observed for *hobo*^{vahs} in different species correspond to what has been described for the *D. melanogaster hobo* element (Saville *et al.*, 1999). However, similar consensus sequences were also observed for other elements of the *hAT* superfamily (Guimond *et al.*, 2003). Furthermore, the *hobo* element has been cross-mobilized by other transposases, such as the *Hermes* element (Sundararajan *et al.*, 1999), or else by unidentified transposases from different tephritid species (Handler & Gomez, 1996). Thus, even though other sources of transposases available to *hobo*^{vahs} cannot be discarded at this moment, we suggest that the most probable source is indeed the *hobo* element. Still, the canonical *hobo* is thought to be a recent acquisition by *D. melanogaster* and *D. simulans* genomes through horizontal transfer (Daniels *et al.*, 1990; Periquet *et al.*, 1990, 1994; Simmons, 1992) and, for this reason, the canonical *hobo* could not be the transposase source available throughout the whole evolution of *hobo*^{vahs}.

(ii) hobo^{vahs} *and a* hobo*-related MITE origin*

Some *hobo*^{vahs} sequences showed a remarkably short length, for example, 83 bp in *D. melanogaster*, 324 bp in *D. yakuba*, 193 bp in *D. sechellia*, 391 bp in *D. santomea* and 251 bp in *D. mauritiana*, while the shorter sequences observed in *D. simulans* and *D. erecta* were about 800 bp. The short sequences exhibit characteristics that are typical of MITEs. The distinctive marks of this TE group are: (i) the short length, typically ranging from 80 to 500 bp in size (but they sometimes reach lengths of up to 1·6 kb); (ii) the presence of TIRs; (iii) high copy number; and (iv) an internal AT-rich region (Feschotte *et al.*, 2002). While it is outstanding that some MITEs can be found in an extraordinarily high copy number (30 000–40 000 copies), this is not an invariable characteristic and many MITEs occur at a lower number (as low as 20 copies) (Feschotte *et al.*, 2002).

The origin of MITEs is not fully understood. Solo TIRs, which by recombination became close to each

https://doi.org/10.1017/S0016672308009312 Published online by Cambridge University Press

other, could be the origin of some MITEs. However, Feschotte *et al*. (2002) have proposed a model in which (i) autonomous transposons suffered internal deletions and became non-autonomous, and (ii) some copies of non-autonomous transposons underwent a 'shrink' and a rapid amplification of copy number. Some studies have been carried out and support this model. Jiang *et al*. (2004) illustrated it with TEs of the rice genome, showing different cases in which the origin of some MITEs is related to their 'cousin' autonomous elements. For example, the MITE *mPing* is 430 bp long with subterminal sequences (252 bp at the 5′ end and 178 bp at the 3′ end) and with TIRs identical to the autonomous transposon *Ping*. Also, Saito *et al*. (2005) have shown that the wheat MITE *Hikkoshi* exhibits subterminal regions and identified TIRs of *Hikkoshi*-like transposons in rice.

Quesneville *et al*. (2006) described the origin of MITEs related to *P* elements (*P*MITE). Ten different *P*MITE families were found in the *Anopheles gambiae* genome. These MITEs present conserved ~100 bp fragments in the 5′ and 3′ subterminal regions that permit identification of the *P* element family that gave rise to each MITE family. *A. gambiae* has nine different *P* families and six of them have given rise to MITEs. As in *P*MITEs described by Quesneville *et al*. (2006), the shorter *hobo*^{vahs} described here maintain 5′ and 3′ subterminal regions conserved in relation to the *hobo* element.

By examining the *hobo*^{vahs} sequences, representative candidates for each phase of MITE origin, according to the model proposed by Feschotte *et al*. (2002), can be identified. As shown in Fig. 4, examples of each MITE origin phase can be found in *D. melanogaster*, *D. sechellia*, *D. simulans* and *D. yakuba* and are depicted in a schematic form. In the process of MITE origin suggested here, the starting point could be complete and autonomous elements, like the canonical *hobo* or elements *hobo*-like of previous genomic invasions. In the next step, some of the autonomous elements are converted into non-autonomous elements, which maintain the conserved 5′ and 3′ subterminal regions but undergo divergence in the inner region, which becomes AT-rich. These relic elements showed a variation in length from 1·5 to 0·7 kb (the typical *hobo*^{vahs} described here). Finally, in *D. melanogaster*, *D. sechellia* and *D. yakuba*, there are very short elements (700–83 bp) showing conserved extremities and TIRs with a typical MITE structure. For these reasons, we propose that the short *hobo*^{vahs} could be classified as MITEs and that they offer a well-documented example of the origin of a '*hobo*-related' MITE.

We thank the *Drosophila* genome projects for making sequences freely available to the scientific community; two anonymous referees for constructive comments; Dr Jean David; Daniel Lachaise (CNRS; France) and Marco Gottschakk (UFRGS, Brazil) for the *Drosophila* strains; Lenira Sepel and Nina Roth Mota for help in preparing the manuscript. This work was supported by grants from CNPq, FAPERGS (number 0411965) and Probic-Fapergs.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Anxolabehere, D., Kidwell, M. & Periquet, G. (1988). Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Molecular Biology and Evolution* **5**, 252–269.

Atkinson, P. W., Warren, W. D. & O'Brochta, D. A. (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises the *Ac* element of maize. *Proceedings of the National Academy of Sciences of the USA* **90**, 9693–9697.

Blackman, R. K., Koehler, M. M., Grimaila, R. & Gelbart, W. M. (1989). Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO Journal* **8**, 211–217.

Borsatti, F., Azzoni, P. & Mandrioli, M. (2003). Identification of a new *hobo* element in the cabbage moth, *Mamestra brassicae* (Lepidoptera). *Hereditas* **139**, 151–155.

Boussy, I. A. & Daniels, S. B. (1991). *hobo* transposable elements in *Drosophila melanogaster* and *D. simulans*. *Genetical Research* **58**, 27–34.

Bowen, N. J. & McDonald, J. F. (2001). *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Research* **11**, 1527–1540.

Brookfield, J. F. Y. (2005). The ecology of the genome – mobile DNA elements and their hosts. *Nature Reviews Genetics* **6**, 128–136.

Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. (1991). Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell* **66**, 465–471.

Capy, P., Bazin, C., Higuet, D. & Langin, T. (1998). *Dynamics and Evolution of Transposable Elements*. Austin, TX: Landes Bioscience.

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T., Kellis, M., Gelbart, W. *et al*. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218.

Daniels, S. B., Chovnick, A. & Boussy, I. A. (1990). Distribution of *hobo* transposable elements in the genus *Drosophila*. *Molecular Biology and Evolution* **7**, 589–606.

Devault, J. D. & Narang, S. K. (1994). Transposable elements in *Lepidoptera*: *hobo*-like transposons in *Heliothis virescensand* and *Helicoverpa zea*. *Biochemical and Biophysical Research Communications* **203**, 169–175.

Feschotte, C., Zhang, X. & Wessler, S. R. (2002). Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. In *Mobile DNA II* (ed. N. L. Craig, R. Craigie, M. Gellert & A. M. Lambowitz), pp. 1147–1158. Washington, DC: ASM Press.

Galindo, M. I., Bigot, Y., Sánchez, M. D., Periquet, G. & Pascual, L. (2001). Sequences homologous to the *hobo* transposable element in E strains of *Drosophila melanogaster*. *Molecular Biology and Evolution* **18**, 1532–1539.

Guimond, N., Bideshi, D. K., Pinkerton, A. C., Atkinson, P. W. & O'Brochta, D. A. (2003). Patterns of *Hermes* transposition in *Drosophila melanogaster*. *Molecular Genetics and Genomics* **268**, 779–790.

Handler, A. M. & Gomez, S. P. (1996). The *hobo* transposable element excises and has related elements in tephritid species. *Genetics* **143**, 1339–1347.

Jiang, N., Feschotte, C., Zhang, X. & Wessler, S. R. (2004). Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology* **7**, 115–119.

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., Ashburner, M. & Celniker, S. E. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* **3**, research0084·1–84·20.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y.T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., Weber, R. J., Haussler, D. & Kent, W. J. (2003). The UCSC Genome Browser Database. *Nucleic Acids Research* **31**, 51–54.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D. & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research* **32** (Suppl. 1): D493–D496.

Kent, W. J. (2002). BLAT – the BLAST-like alignment tool. *Genome Research* **12**, 656–664.

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

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.

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

Lachaise, D. & Silvain, J.-F. (2004). How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster–D. simulans* palaeogeographic riddle. *Genetica* **120**, 17–39.

Le Rouzic, A. & Capy, P. (2005). The first steps of transposable elements invasion: parasitic strategy *versus* genetic drift. *Genetics* **169**, 1033–1043.

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

Lim, J. K. (1988). Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **85**, 9153–9157.

Loreto, E. L. S., Silva, B., Zaha, A. & Valente, V. L. S. (1998). Distribution of transposable elements in neotropical species of *Drosophila*. *Genetica* **7**, 1–13.

McGinnis, W., Shermoen, A. W. & Beckendorf, S. K. (1983). A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75–84.

Nicholas, K. B. & Nicholas Jr, H. B. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Periquet, G., Hamelin, M. H., Kalmes, R. & Eeken, J. (1990). *hobo* elements and their deletion-derivative sequences in *Drosophila melanogaster* and its sibling species *D. simulans*, *D. mauritiana* and *D. sechellia*. *Genetics Selection Evolution* **22**, 393–402.

Periquet, G., Lemeunier, F., Bigot, Y., Hamelin, M. H., Bazin, C., Ladevèze, V., Eeken, J., Galindo, M. I., Pascual, L. & Boussy, I. (1994). The evolutionary genetics of the *hobo* transposable element in the *Drosophila melanogaster* complex. *Genetica* **93**, 79–90.

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M. & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Computational Biology* **1**, e22.

Quesneville, H., Nouaud, D. & Anxolabéhère, D. (2006). *P* elements and MITE relatives in the whole genome sequence of *Anopheles gambiae*. *BMC Genomics* **7**, 214.

Saito, M., Yonemaru, J., Ishikawa, G. & Nakamura, T. (2005). A candidate autonomous version of the wheat MITE *Hikkoshi* is present in the rice genome. *Molecular Genetics and Genomics* **273**, 404–414.

Sanchez-Gracia, A., Maside, X. & Charlesworth, B. (2005). High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends in Genetics* **21**, 200–203.

Sassi, A. K., Herédia, F., Loreto, E. L. S., Valente, V. L. S. & Rohde, C. (2005). Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. *Genetics and Molecular Biology* **28**, 734–739.

Saville, K. J., Warren, W. D., Atkinson, P. W. & O'Brochta, D. A. (1999). Integration specificity of the *hobo* element of *Drosophila melanogaster* is dependent on sequences flanking the integration site. *Genetica* **105**, 133–147.

Simmons, G. M. (1992). Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Molecular Biology and Evolution* **9**, 1050–1060.

Staden, R. (1996). The Staden sequence analysis package. *Molecular Biotechnology* **5**, 233–241.

Sundararajan, P., Atkinson, P. W. & O'Brochta, D. A. (1999). Transposable element interactions in insects: cross mobilization of *hobo* and *Hermes*. *Insect Molecular Biology* **8**, 359–368.

Tamura, K., Subramanian, S. & Kumar, S. (2004). Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Molecular Biology and Evolution* **21**, 36–44.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Torres, F. P., Fonte, L. F. M., Valente, V. L. S. & Loreto, E. L. S. (2006). Mobilization of a *hobo*-related sequence in the genome of *Drosophila simulans*. *Genetica* **126**, 101–110.

Torti, C., Gomulski, L. M., Bonizzoni, M., Murelli, V., Moralli, D., Guglielmino, C. R., Raimondi, E., Crisafulli, D., Capy, P., Gasperi, G. & Malacrida, R. (2005). *Cchobo*, a *hobo*-related sequence in *Ceratitis capitata*. *Genetica* **123**, 313–325.

Yannopoulos, G., Stamatis, N., Monastrioti, M., Hatzopoulos, P. & Louis, C. (1987). *Hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell* **49**, 487–495.