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

**MAURO DE FREITAS ORTIZ** AND **ELGION LUCIO SILVA LORETO**

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 symbiotic 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).

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; Boussey & Daniels, 1991; Simmons, 1992). Finally, the third form is described as a *hobo* relic or *hobo*-related sequence (hRS). 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 hRS element, called *hobo* "a, 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* "a 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* "a homologous sequences (*hobo* "a*) 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* "a* sequences that could be the origin of some related miniature inverted-repeat TEs (MITEs).

2. Material and methods

(i) Fly stocks


(ii) Genome search

Initially, the search for sequences homologous to *hobo* "a* (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* "a* (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* "a* 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<sub>va</sub>ths were: hva1s (forward), 5′-cataacggaaggtagaagaag-3′; hva2as (reverse), 5′-cgtccacgtaaacaect-3′; Vanew1 (forward), 5′-caattttgwgctcggtgy-3′; Vayak (reverse), 5′-gaactgcaagacgacgg-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<sup>va</sup> 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_alignment.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 MgCl₂, 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 EcoRI (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<sup>va</sup> or D. santomea hobo<sup>va</sup>, 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 closest-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′ sub-terminal region and 1152–1220 nucleotides of the 3′ sub-terminal region (using the hobo<sup>va</sup> 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<sup>va</sup> by PCR and Southern blot

Analyses by PCR have shown sequences homologous to hobo<sup>va</sup>, as described by Torres et al. (2006), only in species of the melanogaster subgroup. As can be seen in Table 1, amplicons of hobo<sup>va</sup>ths 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. 1A, in which hobo<sup>va</sup> of D. simulans was used as a probe, numerous hybridization bands were observed in D. sechellia,
Table 1. PCR results with different primer combinations

<table>
<thead>
<tr>
<th>Primer combination</th>
<th>mau&lt;sup&gt;a&lt;/sup&gt;</th>
<th>sim&lt;sup&gt;b&lt;/sup&gt;</th>
<th>sec&lt;sup&gt;c&lt;/sup&gt;</th>
<th>mel&lt;sup&gt;d&lt;/sup&gt;</th>
<th>san&lt;sup&gt;e&lt;/sup&gt;</th>
<th>tei&lt;sup&gt;f&lt;/sup&gt;</th>
<th>ana&lt;sup&gt;g&lt;/sup&gt;</th>
<th>mal&lt;sup&gt;h&lt;/sup&gt;</th>
<th>kik&lt;sup&gt;i&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanew1/hva2as</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Vanew1/Vayak</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>hva1s/Vayak</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>hva1s/hva2as</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<sup>a</sup>D. mauritiana, <sup>b</sup>D. simulans, <sup>c</sup>D. sechellia, <sup>d</sup>D. melanogaster, <sup>e</sup>D. santomea, <sup>f</sup>D. teissieri, <sup>g</sup>D. ananassae, <sup>h</sup>D. malerkotliana, <sup>i</sup>D. kikkawai.

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<sup>vals</sup> of D. santomea was used as a probe (Fig. 1B), 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<sup>vals</sup> are restricted to the melanogaster subgroup.

(ii) Cloning and sequencing of hobo<sup>vals</sup>

We have cloned and sequenced some elements for those species that have hobo<sup>vals</sup> but the genome sequences are not available. Three sequenced clones of D. mauritiana hobo<sup>vals</sup> were around 1·1 kb long and exhibited 90% general similarity to hobo<sup>va</sup> of D. simulans. One clone showed a short hobo<sup>vals</sup> 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<sup>vals</sup>, and in the 3’ end, the last 70 bp had 82% similarity. As in hobo<sup>va</sup>, the middle region is AT-rich.

4. Genomic search

A search for homologous sequences in the 12 available Drosophila genomes, which represent diverse Drosophilidae groups, demonstrated the presence of hobo<sup>vals</sup> only in the melanogaster group (D. melanogaster, D. simulans, D. yakuba, D. sechellia and D. erecta).

The copy number of hobo<sup>vals</sup> 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<sup>vals</sup> 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<sup>vals</sup> 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...
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 hobovahs 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 hobovahs 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 XhoI restriction sites in one or both extremities. Since the length of these sequences was normally 1·1 kb, the distance between the XhoI 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 Boussey & 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 hobovahs sequences from clusters A and B varied from 0·0 to 37·4% with an average of 14·7%.

5. Discussion

(i) **hobo**<sup>chas</sup> are disseminated in the *D. melanogaster* subgroup

hRSs 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 hRS, the hobo<sup>wa</sup>, 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
Fig. 2. For legend see opposite page.
sequence was also observed in this last species. Our results reinforce this supposition since several different *hobo*	extsuperscript{vals} 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*	extsuperscript{va} characteristics in every species of the *melanogaster* subgroup could be explained in two different ways: (i) *hobo*	extsuperscript{vals} 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*	extsuperscript{vals} 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*	extsuperscript{vals} 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*	extsuperscript{vals} copies – reinforced by scattered chromosome distribution over

---

**Fig. 2.** Phylogenetic analysis of *hobo*	extsuperscript{vals} 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).

**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*	extsuperscript{vals} and the question mark indicates that so far it has not been possible to establish whether the species possess the sequence.
all chromosomes arms – is in concordance with the Lerat et al. (2003) hypothesis of high TE turnover. From this perspective, the hobovahs 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 hRSs present in the euchromatic regions of these genomes. However, it is possible that degenerated copies of hRSs and PNM hobo\textsuperscript{vals} 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\textsuperscript{vals}, the canonical hobo is the most likely supplier. The consensus sequences of TSDs observed for hobo\textsuperscript{vals} 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\textsuperscript{vals} 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\textsuperscript{vals}.

(ii) hobo\textsuperscript{vals} and a hobo-related MITE origin

Some hobo\textsuperscript{vals} 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).

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

---

**Fig. 4.** Schematic representation of hobo\textsuperscript{vals} 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.
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 Hikoshi-like transposons in rice.

Quesneville et al. (2006) described the origin of MITEs related to P elements (PMITE). Ten different PMITE 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 PMITEs described by Quesneville et al. (2006), the shorter hobo rats described here maintain 5’ and 3’ subterminal regions conserved in relation to the hob element.

By examining the hobo rats 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 rats 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 rats 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 Gottschalk (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


