Genet. Res., Camb. (2011), **93**, *pp.* 381–385. © Cambridge University Press 2011
doi:10.1017/S0016672311000371

SHORT PAPER
msechBari, a new MITE-like element in *Drosophila sechellia* related to the *Bari* transposon

ELAINE SILVA DIAS AND CLAUDIA MARCIA APARECIDA CARARETO*
Department of Biology, UNESP – São Paulo State University, São José do Rio Preto, São Paulo, Brazil

(*Received 18 August 2011; revised 2 November 2011; accepted 8 November 2011*)

Summary

A few occurrences of miniature inverted-repeat transposable elements (MITEs) have been reported in species of the genus *Drosophila*. Here, we describe *msechBari*, a MITE-like element in *Drosophila sechellia*. The element is short, approximately 90 bp in length, AT-rich and occurs in association with, or close to, genes, characteristics that are typical for MITEs. The identification was performed *in silico* using the sequenced genome of *D. sechellia* and confirmed in a laboratory strain. This short element is related to the *Bari_DM* transposon of *Drosophila melanogaster*, having terminal inverted repeats (TIRs) of a similar length and a high identity with the full-length *Bari_DM* element. The estimated recent origin of the element and the homogeneity observed between copies found in the genome suggests that *msechBari* could be active in *D. sechellia*.

1. Introduction

Miniature inverted-repeat transposable elements (MITEs) are non-autonomous short repeats that mobilize within the host genome even without the potential to encode the protein (i.e. the transposase) responsible for their mobilization. The MITEs are, in general, derived from ancient, related autonomous elements, and their origin can occur through internal deletions in autonomous elements, where the only remaining are the terminal inverted repeats (TIRs) and, sometimes, portions between the TIRs and the coding region of the transposase. This origin supports the proposal of their mobilization *in trans* by a transposase encoded by a full-length element (Feschotte & Pritham, 2007). The autonomous transposons use the cellular machinery of the host cells for the protein synthesis necessary for their mobilization, whereas the MITEs use the machinery encoded by transposons for mobilization. In the 1980s, Orgel & Crick (1980) referred to 'selfish DNA' as the 'ultimate parasites' due to the relationship of parasitism between autonomous elements and the machinery of the cell. Recently, González & Petrov (2009) enlarged this idea to include the MITEs because of their dependency on autonomous elements for mobilization.

In general terms, MITE-like elements have been widely described in plants (Moreno-Vazquez *et al.*, 2005; Lin *et al.*, 2006; Guermonprez *et al.*, 2008) and specifically in grapevine (Benjak *et al.*, 2009), maize (Bureau & Wessler, 1992; Zerjal *et al.*, 2009), cereal grasses (Bureau & Wessler, 1994), *Arabidopsis* (Feschotte & Mouches, 2000), rice (Feschotte *et al.*, 2003; Jiang *et al.*, 2003; Nakazaki *et al.*, 2003; Shan *et al.*, 2005), *Medicago* (Grzebelus *et al.*, 2007, 2009), apple (Han & Korban, 2007), beet (Menzel *et al.*, 2006), barley (Lyons *et al.*, 2008; Petersen & Seberg, 2009), grasses (Park *et al.*, 2003), pearl millet (Remigereau *et al.*, 2006) and pome fruit trees (Wakasa *et al.*, 2003). Descriptions in other organisms, such as bacteria (Chen *et al.*, 2008), cyanobacteria (Zhou *et al.*, 2008), fungi (Xu *et al.*, 2010), silkworms (Han *et al.*, 2010), fish (de Boer *et al.*, 2007) and amphibians (Hikosaka *et al.*, 2011) are also found in the literature, but few occurrences have been reported in the *Drosophila* genus (Tudor *et al.*, 1992; Miller *et al.*, 2000; Ortiz *et al.*, 2010). Although numerous MITEs have been identified, the association with autonomous elements is often absent. Here, we describe an MITE-like element found in the genome of *Drosophila sechellia* that is associated with the *Bari* transposon described in *Drosophila melanogaster*. The

* Corresponding author: Department of Biology, UNESP – São Paulo State University, São José do Rio Preto, São Paulo, Brazil. E-mail: cararetо@ibilce.unesp.br

high similarity found with *Bari_DM* in both TIRs and internal regions suggests a close relationship with autonomous elements.

2. Materials and methods

Searches for *Bari_DM* elements in the genomes of species of *Drosophila* (unpublished data) resulted in the identification of an ~90 bp sequence, with TIRs and no coding sequence, in the *D. sechellia* genome. After this observation, the sequence of the TIRs of the *Bari_DM* element of *D. melanogaster* (X67681) was used to search the genome of *D. sechellia* (release 1.3, June 2009) (Drosophila 12 Genomes Consortium, 2007) using the BLASTn software (Altschul *et al.*, 1990). Analyses aimed at identifying the target site duplications (TSDs) and estimations of the gene density in the adjacent regions of the MITEs were also performed extracting the 10 kb 5′ and 3′ flanking regions of each insertion. The ability to form secondary structure was analysed using Mfold (Zuker, 2003) (available at http://mfold.rna.albany.edu/).

To confirm that these MITEs were not a sequencing artefact, their occurrence was searched in a *D. sechellia* strain maintained in our laboratory. Genomic DNA was extracted from 50 individuals according to a previously described protocol (Jowett, 1986). The amplification, cloning and sequencing were performed using specific primers based on the consensus sequence of the MITE identified in the *D. sechellia* genome (Forward, 5′-MYRGTCATGGTCAAAATTATTTTCACAA-3′ and Reverse, 5′-ACAGAGGTGGTCAAAAGTATTTTCACWW-3′). PCR amplification was performed using 0·3125 unit of Taq polymerase (Invitrogen), 200 ng genomic DNA, 1 mM of $MgCl_2$, 1 × buffer, 0·08 mM of dNTPs and 0·4 mM of primers for a final volume of 25 μl. The PCR conditions were as follows: initial denaturation (94 °C, 120 s), followed by 30 cycles of denaturation (94 °C, 15 s), annealing (59 °C, 10 s) and extension (72 °C, 20 s). The PCR products were purified (DNA GFX DNA & Gel Band, GE) and cloned (TOPO TA Cloning kit, Invitrogen) according to the specifications of the manufacturers. Eight clones were selected for extraction of the plasmid by a phenol/chloroform protocol and sequenced using universal primers, M13F and M13R, resulting in four sequences with good quality.

The evolutionary relationships between the sequences were reconstructed using the software Network with the Median Joining algorithm (Bandelt *et al.*, 1999) and the default parameters, using the nucleotide sequences extracted from the *D. sechellia* genome. The age of these insertions was estimated using the following molecular clock equation ($r = k/2T$), where r is the neutral synonymous substitution rate of the *Drosophila* genus ($r = 0{\cdot}011$/site/Myr) (Tamura *et al.*, 2004) and k is the divergence rate (Kimura 2-parameter distance) (Kimura, 1980). The consensus sequence was reconstructed using the software, DAMBE (Xia & Xie, 2001), and the distances were calculated using MEGA version 5 (Tamura *et al.*, 2011).

3. Results and discussion

In general, MITEs are smaller than 600 bp in length, have conserved TIRs, a target site preference, no coding potential and are AT-rich (Feschotte *et al.*, 2002). We found 49 MITE-like sequences in the sequenced genome of *D. sechellia* (see Supplementary Table S1 available at http://journals.cambridge.org/GRH) that presented lengths between 65 and 89 bp, TIRs of 28 bp and AT contents of approximately 66%. Approximately 63% of these sequences are flanked by AT dinucleotides, which are typical TSDs of the MITE family *Stowaway* (Feschotte *et al.*, 2002). Both consensus sequences showed potential to form secondary structure (see Supplementary Figure S1 available at http://journals.cambridge.org/GRH), ability present in MITEs. Additionally, as other MITEs (Zerjal *et al.*, 2009; Han *et al.*, 2010), these sequences are preferentially associated with gene regions (62% of the insertions were localized within genes or harboured genes in their 10 kb flanking regions).

The MITE-like sequences described here (Fig. 1 and Supplementary Table S1) show a high similarity with the *Bari_DM* transposon described in *D. melanogaster*, but they are significantly smaller (65–89 bp) than this autonomous element (1728 bp). Two types of sequences were found, with their TIRs 100 and 89% similar to the *Bari_DM*, and both shared three internal regions of 100% identity to *Bari_DM* and between them. Thus, we concluded that the sequences described in the *D. sechellia* genome are derivatives of the *Bari* element, hereafter termed *msechBari* elements.

These two types of *msechBari*, which essentially differ by three nucleotides in their TIRs, were grouped into two well-defined clusters in a network tree; thus, they can be considered to be two MITE subfamilies (Fig. 2). The network suggests the existence of a master sequence that would have given rise to the two groups of sequences. Evolution under the master gene model is characterized, in graphic reconstructions of evolutionary relationships, by a star topology, where the central sequence gives rise to the derived sequences (Cordaux *et al.*, 2004). The length of the branches is related to the elapsed time since the origin of each sequence: short branches suggest a recent origin, and long branches indicate an old origin.

The two subfamilies derived from the two master sequences, *msechBari1* and *msechBari2*, have

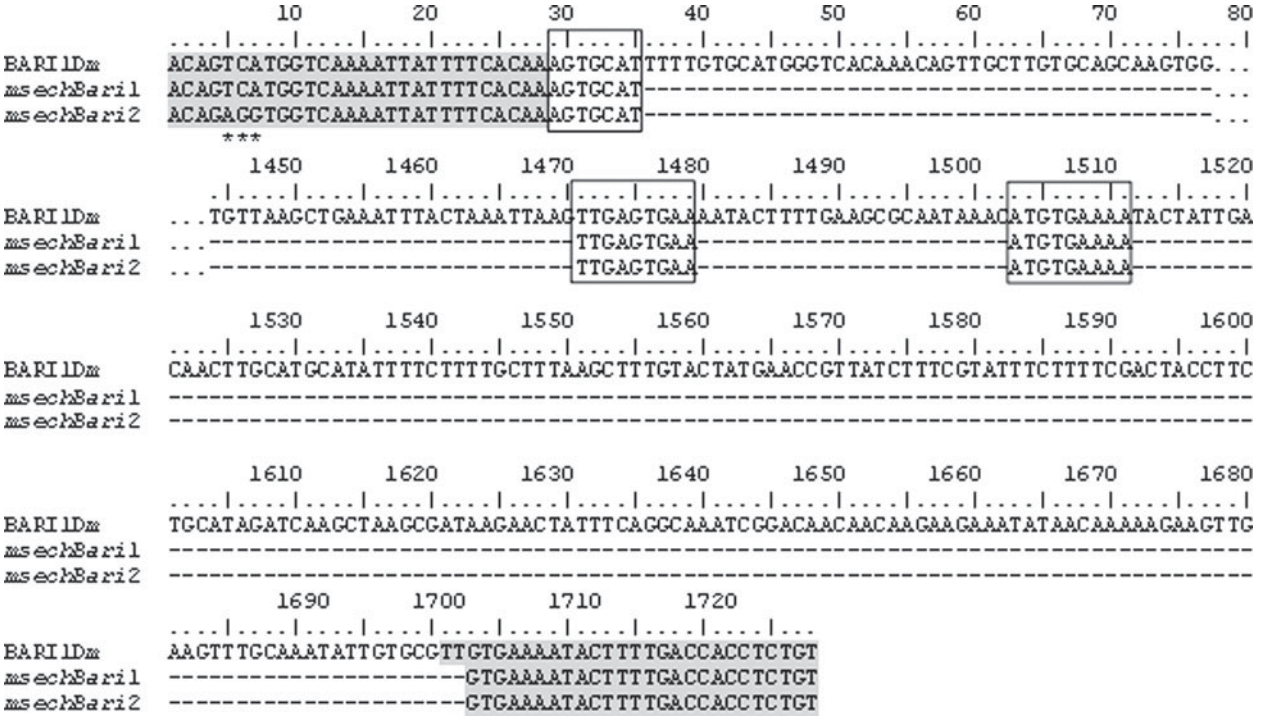

Fig. 1. Alignment of the MITE-like sequences identified in the sequenced genome of *D. sechellia* with the consensus sequence of the transposon, *Bari_DM*, as described in *D. melanogaster*. The shaded region corresponds to the TIRs, with the three diagnostic substitutions of the two MITE subfamilies highlighted with asterisks; the boxes indicate the remaining non-coding regions found in the MITEs and the dotted region corresponds to the not shown nucleotides 77–1444 present only in the transposon *Bari_DM*.

Fig. 2. Network of the MITE-like sequences identified in the sequenced genome of *D. sechellia*. The size of the circles corresponds to sequence frequency; the size of the branches is proportional to the number of mutations occurred, as indicated by numbers above branches. Black circles correspond to the sequences of *msechBari1* subfamily and the grey circles to the *msechBari2* subfamily.

short evolutionary distances within the group, $0{\cdot}00341 \pm 0{\cdot}00036$ and $0{\cdot}00279 \pm 0{\cdot}0004$, respectively; however, when a comparison was made between the subfamilies, the distance was larger, $0{\cdot}05020 \pm 0{\cdot}00029$. The short distances between the sequences within a subfamily, the short branches and the absence of reticulation in the network suggest a recent burst of transposition of these elements in the genome of the strain that was sequenced. Accordingly, the groupings of sequences in the network, represented by large circles, indicate that the sequences are identical; therefore, these sequences are very recent and have not had sufficient time to diverge. Similar events have been reported for other transposable elements (Yang *et al*., 2006; de Boer *et al*., 2007; Marzo *et al*., 2008; Konovalov *et al*., 2010; Lerat, 2010) and MITE-like sequences in different organisms (Jiang *et al*., 2003; Chen *et al*., 2008; Zhou *et al*., 2008; Han *et al*., 2010; Hikosaka *et al*., 2011). This recent origin is also supported by the average time of origin of the insertions of each subfamily, 155 000 years (*msechBari1*) and 127 000 years (*msechBari2*). We confirmed the presence of *msechBari* in a laboratory strain (Fig. 1). The sequences found were similar to those in the *D. sechellia* sequenced genome. They had the internal region conserved, but the 2 bp of the 5′ TIRs were variable (see Supplementary Figure S2 available at http://journals.cambridge.org/GRH). This variation, if real, could indicate inactivity of these MITES. However, as we obtained only four sequences, it is possible that

https://doi.org/10.1017/S0016672311000371 Published online by Cambridge University Press

the two pairs of variable bases are sequencing artefacts.

For the mobilization of a transposon, such as *Bari*, the transposase proteins recognize and bind to specific sites in the TIRs to promote transposition. For some MITEs found in plants, the mobilization of transposons that do not have coding capacity has been suggested to occur via transposases *in trans* from elements that are distantly related. For example, up to approximately 20 000 insertions of rice MITE-like elements of the *Stowaway* family, which exhibit TIRs similar to other mariner-like elements, have been reported. However, these elements are not homologous to any other autonomous elements that have been described in rice; thus, it has been proposed that these elements be mobilized by a transposase encoded by other distantly related autonomous elements (Feschotte, 2008). Therefore, the recent transposition of the *msechBari* could have resulted from the presence of a transposase from an active *Bari_DM* transposon in *D. sechellia* or from other *Bari*-like elements in the *D. sechellia* genome that can recognize the TIRs. Only two full-length *Bari* copies, with both intact TIRs, were found in the *D. sechellia* genome, but both have many stop codons in their transposase coding sequences (see Supplementary Figure S3 available at http://journals.cambridge.org/GRH), indicating that the *Bari* element in *D. sechellia* is inactive. However, both copies exhibit a low diversity, when compared with the consensus sequence of *Bari_DM*, suggesting that this inactivity is recent. Therefore, this autonomous element is potentially responsible for the *msechBari* mobilization in the recent past; however, the mobilization by another distantly related autonomous element cannot be disregarded.

Funding for this project was provided by the Brazilian agencies, FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant 2010/10731-4 to C. M. A. C. and fellowship 2008/07629-3 to E. S. D.), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) (Grant 304880/2009-4 to C. M. A. C.) and FUNDUNESP (Fundação para o Desenvolvimento da UNESP) (Grant 670/10). We thank Jean David, PhD for providing the strain used in this study.

4. Supplementary material

The online data are available at http://journals.cambridge.org/GRH

References

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.

Bandelt, H., Forster, P. & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Benjak, A., Boue, S., Forneck, A. & Casacuberta, J. M. (2009). Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). *Genome Biology and Evolution* **1**, 75–84.

Bureau, T. E. & Wessler, S. R. (1992). *Tourist* – a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4**, 1283–1294.

Bureau, T. E. & Wessler, S. R. (1994). Mobile inverted-repeat elements of the *tourist* family are associated with the genes of many cereal grasses. *Proceedings of the National Academy of Sciences USA* **91**, 1411–1415.

Chen, Y., Zhou, F. F., Li, G. J. & Xu, Y. (2008). A recently active miniature inverted-repeat transposable element, *Chunjie*, inserted into an operon without disturbing the operon structure in *Geobacter uraniireducens* Rf4. *Genetics* **179**, 2291–2297.

Cordaux, R., Hedges, D. & Batzer, M. (2004). Retrotransposition of Alu elements: how many sources? *Trends in Genetics* **20**, 464–467.

de Boer, J. G., Yazawa, R., Davidson, W. S. & Koop, B. F. (2007). Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* **8**, 422. doi:10.1186/1471-2164-8-422.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218.

Feschotte, C. (2008). Opinion – transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9**, 397–405.

Feschotte, C. & Mouches, C. (2000). Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Molecular Biology and Evolution* **17**, 730–737.

Feschotte, C. & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41**, 331–368.

Feschotte, C., Swamy, L. & Wessler, S. R. (2003). Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *stowaway* miniature inverted repeat transposable elements (MITEs). *Genetics* **163**, 747–758.

Feschotte, C., Zhang, X. & Wessler, S. R. (2002). Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In *Mobile DNA II* (ed. N. Craig, R. Craigie, M. Gellert & A. Lambowitz), pp. 1093–1110. Washington, DC: ASM Press.

González, J. & Petrov, D. (2009). MITEs – the ultimate parasites. *Science* **325**, 1352–1353.

Grzebelus, D., Gladysz, M., Macko-Podgorni, A., Gambin, T., Golis, B., Rakoczy, R. & Gambin, A. (2009). Population dynamics of miniature inverted-repeat transposable elements (MITEs) in *Medicago truncatula*. *Gene* **448**, 214–220.

Grzebelus, D., Lasota, S., Gambin, T., Kucherov, G. & Gambin, A. (2007). Diversity and structure of PIF/Harbinger-like elements in the genome of *Medicago trunculata*. *BMC Genomics* **8**, 409. doi:10.1186/1471_2164-8-409.

Guermonprez, H., Loot, C. & Casacuberta, J. M. (2008). Different strategies to persist: the *pogo*-like *lemi*1 transposon produces miniature inverted-repeat transposable elements or typical defective elements in different plant genomes. *Genetics* **180**, 83–92.

Han, M. J., Shen, Y. H., Gao, Y. H., Chen, L. Y., Xiang, Z. H. & Zhang, Z. (2010). Burst expansion, distribution and diversification of MITEs in the silkworm genome. *BMC Genomics* **11**, 520. doi: 10.1186/1471-2164-11-520.

Han, Y. & Korban, S. S. (2007). *Spring*: A novel family of miniature inverted-repeat transposable elements is associated with genes in apple. *Genomics* **90**, 195–200.

Hikosaka, A., Nishimura, K., Hikosaka-Katayama, T. & Kawahara, A. (2011). Recent transposition activity of *Xenopus* T2 family miniature inverted-repeat transposable elements. *Molecular Genetics and Genomics* **285**, 219–224.

Jiang, N., Bao, Z. R., Zhang, X. Y., Hirochika, H., Eddy, S. R., McCouch, S. R. & Wessler, S. R. (2003). An active DNA transposon family in rice. *Nature* **421**, 163–167.

Jowett, T. (1986). Preparation of nucleic acids. In *Drosophila: A Practical Approach* (ed. D. B. Roberts), pp. 275–285. Oxford: IRL Press.

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.

Konovalov, F. A., Goncharov, N. P., Goryunova, S., Shaturova, A., Proshlyakova, T. & Kudryavtsev, A. (2010). Molecular markers based on LTR retrotransposons *BARE-1* and *Jeli* uncover different strata of evolutionary relationships in diploid wheats. *Molecular Genetics and Genomics* **283**, 551–563.

Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity* **104**, 520–533.

Lin, X. Y., Long, L. K., Shan, X. H., Zhang, S. Y., Shen, S. & Liu, B. (2006). *In planta* mobilization of *mPing* and its putative autonomous element *Pong* in rice by hydrostatic pressurization. *Journal of Experimental Botany* **57**, 2313–2323.

Lyons, M., Cardle, L., Rostoks, N., Waugh, R. & Flavell, A. J. (2008). Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. *Molecular Genetics and Genomics* **280**, 275–285.

Marzo, M., Puig, M. & Ruiz, A. (2008). The *Foldback*-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. *Proceedings of the National Academy of Sciences USA* **105**, 2957–2962.

Menzel, G., Dechyeva, D., Keller, H., Lange, C., Himmelbauer, H. & Schmidt, T. (2006). Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. *Chromosome Research* **14**, 831–844.

Miller, W. J., Nagel, A., Bachmann, J. & Bachmann, L. (2000). Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Molecular Biology and Evolution* **17**, 1597–1609.

Moreno-Vazquez, S., Ning, J. C. & Meyers, B. C. (2005). *hATpin*, a family of MITE-like *hAT* mobile elements conserved in diverse plant species that forms highly stable secondary structures. *Plant Molecular Biology* **58**, 869–886.

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H. & Tanisaka, T. (2003). Mobilization of a transposon in the rice genome. *Nature* **421**, 170–172.

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

Ortiz, M. D., Lorenzatto, K. R., Correa, B. R. S. & Loreto, E. L. S. (2010). *hAT* transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica* **138**, 649–655.

Park, K. C., Jeong, C. S., Song, M. T. & Kim, N. S. (2003). A new MITE family, Pangrangja, in Gramineae species. *Molecules and Cells* **15**, 373–380.

Petersen, G. & Seberg, O. (2009). Stowaway MITEs in Hordeum (Poaceae): evolutionary history, ancestral elements and classification. *Cladistics* **25**, 198–208.

Remigereau, M. S., Robin, O., Siljak-Yakovlev, S., Sarr, A., Robert, T. & Langin, T. (2006). *Tuareg*, a novel miniature-inverted repeat family of pearl millet (*Pennisetum glaucum*) related to the *PIF* superfamily of maize. *Genetica* **128**, 205–216.

Shan, X. H., Liu, Z. L., Dong, Z. Y., Wang, Y. M., Chen, Y., Lin, X. Y., Long, L. K., Han, F. P., Dong, Y. S. & Liu, B. (2005). Mobilization of the active MITE transposons *mping* and *pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.). *Molecular Biology and Evolution* **22**, 976–990.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.

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.

Tudor, M., Lobocka, M., Goodell, M., Pettitt, J. & O'Hare, K. (1992). The *pogo* transposable element family of *Drosophila melanogaster*. In *Molecular and General Genetics*, Vol. 232, pp. 126–134. Berlin: Springer.

Wakasa, Y., Ishikawa, R., Niizeki, M., Harada, T., Jin, S., Senda, M. & Akada, S. (2003). *Majin*: a miniature DNA element associated with the genomes of pome fruit trees. *Hortscience* **38**, 17–20.

Xia, X. & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92**, 371–373.

Xu, J. S., Wang, M., Zhang, X. Y., Tang, F. H., Pan, G. Q. & Zhou, Z. Y. (2010). Identification of *NbME* MITE families: Potential molecular markers in the microsporidia *Nosema bombycis*. *Journal of Invertebrate Pathology* **103**, 48–52.

Yang, H. P., Hung, T. L., You, T. L. & Yang, T. H. (2006). Genomewide comparative analysis of the highly abundant transposable element *DINE-1* suggests a recent transpositional burst in *Drosophila yakuba*. *Genetics* **173**, 189–196.

Zerjal, T., Joets, J., Alix, K., Grandbastien, M. & Tenaillon, M. (2009). Contrasting evolutionary patterns and target specificities among three *Tourist*-like MITE families in the maize genome. *Plant Molecular Biology* **71**, 99–114.

Zhou, F., Tran, T. & Xu, Y. (2008). *Nezha*, a novel active miniature inverted-repeat transposable element in cyanobacteria. *Biochemical and Biophysical Research Communications* **365**, 790–794.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415.