



HAL
open science

Genome-wide characterization of Mariner Like transposons and their derived MITEs in the Whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae)

Marwa Zidi, Françoise Denis, Khouloud Klai, Benoît Chénais, Aurore Caruso, Salma Djebbi, Maha Mezghani, Nathalie Casse

► To cite this version:

Marwa Zidi, Françoise Denis, Khouloud Klai, Benoît Chénais, Aurore Caruso, et al.. Genome-wide characterization of Mariner Like transposons and their derived MITEs in the Whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *G3*, 2021, 11 (12), pp.jkab287. 10.1093/g3journal/jkab287. hal-03347514

HAL Id: hal-03347514



<https://univ-lemans.hal.science/hal-03347514v1>

Submitted on 4 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genome-wide characterization of Mariner-like transposons and their derived MITEs in the Whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae)

Marwa Zidi ^{1,2}, Françoise Denis,^{2,3} Khoulood Klai,^{1,2} Benoît Chénais ², Aurore Caruso,² Salma Djebbi,¹ Maha Mezghani,^{1,*†} and Nathalie Casse^{2,*†}

¹Laboratory of Biochemistry and Biotechnology (LR01ES05), Faculty of Sciences of Tunis, University of Tunis El Manar, 2092 Tunis, Tunisia

²Biologie des Organismes, Stress, Santé, Environnement, Le Mans Université, F-72085 Le Mans, France and

³Laboratoire BOREA MNHN, CNRS FRE 2030, SU, IRD 207, UCN, UA, 75231 Paris, France

*Corresponding authors: Laboratory Biologie des Organismes, Stress, Santé, Environnement, Department of Biology, Le Mans University, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09 France. Email: ncasse@univ-lemans.fr (N.C.), Laboratory of Biochemistry and Biotechnology (LR01ES05), Department of Biology, Faculty of Sciences of Tunis, University of Tunis El Manar, Campus Universitaire Farhat Hached, B.P. n° 94 - ROMMANA, Tunis 1068, Tunisie. Email: maha.mezghani@fst.utm.tn (M.M.)

†Equivalent authors.

Abstract

The whitefly, *Bemisia tabaci* is a hemipteran pest of vegetable crops vectoring a broad category of viruses. Currently, this insect pest showed a high adaptability and resistance to almost all the chemical compounds commonly used for its control. In many cases, transposable elements (TEs) contributed to the evolution of host genomic plasticity. This study focuses on the annotation of *Mariner*-like elements (MLEs) and their derived Miniature Inverted repeat Transposable Elements (MITEs) in the genome of *B. tabaci*. Two full-length MLEs belonging to *mauritiana* and *irritans* subfamilies were detected and named *Btmar1.1* and *Btmar2.1*, respectively. Additionally, 548 defective MLE sequences clustering mainly into 19 different *Mariner* lineages of *mauritiana* and *irritans* subfamilies were identified. Each subfamily showed a significant variation in MLE copy number and size. Furthermore, 71 MITEs were identified as MLEs derivatives that could be mobilized via the potentially active transposases encoded by *Btmar1.1* and *Btmar2.1*. The vast majority of sequences detected in the whitefly genome present unusual terminal inverted repeats (TIRs) of up to 400 bp in length. However, some exceptions are sequences without TIRs. This feature of the MLEs and their derived MITEs in *B. tabaci* genome that distinguishes them from all the other MLEs so far described in insects, which have TIRs size ranging from 20 to 40 bp. Overall, our study provides an overview of MLEs, especially those with large TIRs, and their related MITEs, as well as diversity of their families, which will provide a better understanding of the evolution and adaptation of the whitefly genome.

Keywords: *Bemisia tabaci*; large terminal inverted repeats; *Mariner*-like elements; miniature inverted repeats transposable elements

Introduction

Transposable elements (TEs) are repetitive mobile DNA sequences that can transpose from a position to another in a host genome. These elements are classified into two classes based on their transposition mechanisms (Finnegan 1989). Class I TEs, known as “copy and paste” elements, are called retrotransposons since they transpose via a reverse transcriptase that can generate cDNA from RNA intermediates. Class II TEs, also known as DNA transposons, primarily use a transposase enzyme for their excision and insertion; although some elements of this class, such as *Helitron* and *Maverick*, encode other proteins required for their transposition (Mat Razali et al. 2019). A further classification for eukaryotic TEs was proposed by Wicker et al. (2007) based on mechanistic and enzymatic differences. This classification organized the TEs in a hierarchical manner into subclasses, orders, superfamilies, and families. Since then, the TEs classification has been a challenge and several classification systems have been

proposed (Curcio and Derbyshire 2003; Piégu et al. 2015; Hoen et al. 2015). The Tc1/*Mariner* superfamily is a major category of DNA transposons and is considered as one of the most widespread TE superfamilies among eukaryotes. The elements of Tc1/*Mariner* superfamily are classified into at least eight families according to the number of amino acid residues present between the second aspartic acid (D) and the third aspartic acid (D) or glutamic acid (E) of the transposase catalytic domain, namely: *Mariner* (DD34D; Jacobson et al. 1986), *maT* (DD37D; Zhang et al. 2016), *VS* (DD41D; Gomulski et al. 2001), *pogo* (DDxD; Shao and Tu 2001; Dupeyron et al. 2020), Tc1-like elements (TLE; DD34E; Emmons et al. 1983; Sang et al. 2019), TR (DD35E; Zong et al. 2020), IC (DD36E; Sang et al. 2019), and TRT (DD37E; Zhang et al. 2016).

The members of the *Mariner* family (DD34D) are called *Mariner*-like elements (MLEs) and constitute a large family that is widespread in all organisms. MLEs are distributed into five major subfamilies *mauritiana*, *cecropia*, *mellifera*, *elegans*, and *irritans*,

Received: May 19, 2021. Accepted: July 28, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

depending on their sequence similarities and phylogenetic relationships (Robertson, 2002; Bigot et al. 2005). Besides, eight minor subfamilies with a limited distribution were also described, among which four have been well-described, namely *vertumnana*, *lineata*, *mosellana*, and *drosophila*, while the other four are still not well-defined (Robertson 2002; Rouault et al. 2009; Filée et al. 2015; Grace and Carr 2020; Dupeyron et al. 2020). MLEs have a simple structure consisting of a transposase open reading frame (ORF) flanked by untranslated regions (UTRs), two terminal inverted repeats (TIRs), and TA target site duplications (TSDs; Plasterk et al. 1999; Leroy et al. 2003). The TIRs are generally sized from 20 to 40 bp, with the exception of the *Mcmar1* element identified in the nematod *Meloidogyne chitwoodi*, which has large TIRs (LTIRs) sized up to 355 bp. The transposase ORF encodes a 350 amino acid polypeptide containing two major domains: (1) The N-terminal DNA-binding domain, which contains the helix-turn-helix (HTH) motif responsible for the binding of the transposase to the TIRs, and the nuclear localization sequence (NLS), which is involved in the translocation of the transposase in the nucleus (Brillet et al. 2007). This domain holds also the first signature motif of the transposase, i.e., WVPREL (Augé-Gouillou et al. 2005). (2) The C-terminal-catalytic domain contains the DD34D motif, meaning that the last two aspartic residues (D) are separated by 34 amino acids. The three residues are anchored to three conserved motifs, i.e., TGDE, HDNA, and YSPDLAP(x)D. This catalytic triad DD34D is required for the cleavage and the integration of the TE into the target site (Brillet et al. 2007; Yuan and Wessler 2011). Only a few *Mariner* elements are naturally active, namely *mos1* (*mauritaniana* subfamily) detected in *Drosophila mauritiana*, *Famar1* (*mellifera* subfamily) found in *Forficula auricularia* and *Mboumar 9* (*mauritaniana* subfamily) discovered in the ant *Messor bouvieri* (Jacobson et al. 1986; Barry et al. 2004; Munoz-Lopez and Garcia-Perez 2010; Sanllorente et al. 2020). Most of the described MLEs are inactive because of the natural selection and the accumulation of mutations. Mutations can impact all parts of the elements even the ORF, making them defective and unable to encode the protein needed to process their own transposition. A fraction of these defective nonautonomous DNA transposons that kept the TIRs, TSDs, and are capable of mobilization could be considered as MITEs (Feschotte et al. 2003). MITEs have the same general features of MLEs and therefore exhibit TIRs flanked by TSDs. However, they do not have an ORF coding for a functional transposase, so they rely for their transposition on the transposase of autonomous class II elements (Yang et al. 2006).

TEs are major factors of genomic innovations. They confer selective advantages to the host by inducing genome sequences variation due to their movement capacity that leads to frequent DNA cleavage and rearrangement, or by being recurrent source of new functional genes and molecular regulators through a process called TEs molecular domestication (Schrader and Schmitz 2019; Jangam, et al. 2017; Drezén et al. 2017). For example, it was well-documented that TEs played a role in the adaptation of *D. melanogaster* to temperate climates (Casola et al. 2007; González et al. 2008). Another example in *D. melanogaster* was given by the insertion of a long terminal repeat of a retrotransposon into the 5' end of the *Cyp6g1* gene inducing its overexpression and causing the resistance of *D. melanogaster* to a wide range of insecticide classes (Chénais et al. 2012). The cotton pest *Heliothis virescens* showed high levels of resistance to the Bt toxin Cry1Ac linked to the insertion of TEs into a cadherin-superfamily gene (Gahan et al. 2001). In *Helicoverpa armigera* nine TE insertions belonging to RTE, R2, CACTA, *Mariner*, and *hAT* superfamilies hosted in exons and introns of cytochrome P450 (CyP450), glutathione S-

transferase (GST), and ATP-binding cassette (ABC) transporter genes were described by Klai et al. (2020).

The hemipteran species *Bemisia tabaci* (Chen et al. 2016) is an economically important agriculture pest, which is responsible of the transmission of more than 300 plant viruses (Navas-Castillo et al. 2011) and infests more than 1000 plant species including Asteraceae, Fabaceae, and Solanaceae families (Abd-Rabou and Simmons 2010). Furthermore, *B. tabaci* displays a high adaptability and resistance to almost all the pesticide used for its control. It resists to the organophosphates, carbamates, pyrethroids, and also to the neonicotinoids and buprofezin (Horowitz et al. 2020). TEs have been described as an important source of genetic modification to acquire insecticide resistance and provide a selective advantage through their insertion sites. For example, the insertion of a MITE 0.2-kb upstream of the P450 gene *CYP9M10* has been correlated to pyrethroid resistance in *Culex quinquefasciatus* (Itokawa et al. 2010). Despite the involvement of TE in genome evolution and adaptation in many insects leading to their resistance to a wide range of insecticides, there is still no study describing TE families in the genus *Bemisia*.

The aim of this study was to detect *Mariner*-like transposons in *B. tabaci* genome, including their derived MITEs, and to characterize complete potentially active MLEs sequences using bioinformatic tools.

Materials and methods

Supporting data

The MEAM1/B *B. tabaci* genome of 615 Mb available in the NCBI database (assembly ASM185493v1) was used to identify MLEs and their derived MITEs. This genome has been sequenced by Chen et al. (2016) using both Illumina short reads and PacBio long-read approaches, then assembled into 19,751 scaffolds with 3,232,964 kbp N50 length.

Identification of MLEs and their derived MITEs

The search for MLEs in *B. tabaci* genome was carried out according to a similarity-based method as described by Bouallègue et al. (2017) and Xie et al. (2018). A total of 57 nucleotide and 19 protein reference sequences from the five major *Mariner* subfamilies (*irritans*, *mauritaniana*, *mellifera*, *elegans*, and *cecropia*) were used as queries (Supplementary Tables S1 and S2). The nucleotide and protein queries were used to perform BLASTN and tBLASTN searches, respectively (Altschul et al. 1990). A first filtration step was conducted to exclude sequences with identity lower than 30%. The remaining sequences were extended to 1000 extra base pairs of their flanking regions in order to get full-length MLE sequences with TSDs, TIRs, UTRs, and ORF. The TIRs were manually detected by aligning the sequences with their reverse complement using *multAlin* algorithm (Corpet, 1988). In order to detect the one-sided TIR sequences, we aligned them with those holding the two-side TIRs. The identified MLE sequences were next used as nucleotide queries to retrieve more sequences using BLASTN algorithm. The obtained sequences were submitted to a second filtration step to exclude redundant copies and those smaller than 500 bp (Wallau et al. 2014; Berthelie et al. 2018). The remaining sequences were clustered using CD-HIT with a threshold of 80% identity and a coverage of 60%. Then, their ORFs were identified using BLASTX algorithm and SIXPACK program implemented in the EMBOSS portal (https://www.ebi.ac.uk/Tools/st/emboss_sixpack). Additional searches of the HTH DNA-binding domain and NLS were performed using NPS@: Network Protein Sequence Analysis and SeqNLS (Dodd and Egan 1990; Lin

and Hu 2013), respectively. The search for the catalytic domain represented by the three aspartic acid separated by 34 amino acid anchored to the TGDE, HDNA, and YSPDLA(X)D motifs was carried out manually. Finally, a rearmost filtration step was carried out to discard copies with no catalytic domain.

To retrieve MITEs derived from MLEs, the MEAM/B genome was first submitted to MITE Tracker (Crescente et al. 2018) and the obtained putative MITE sequences were then used as queries to perform a BLASTN (e -value $< 10^{-10}$; Lu et al. 2012) against the previously identified MLEs.

MLEs clusterization

The MLE sequences were compared to each other and clustered based on a sequence identity higher than 80% and a minimum alignment coverage of 60% to their longest sequence. The obtained lineages were visualized by Cytoscape 3.7.2 software then used to perform *maximum-likelihood* analysis by mean of MEGAX software (Kumar et al. 2018). The assignation of the MLEs lineages into their subfamilies was performed based on their comparison to the NCBI database using BLASTX algorithm, then,

a maximum-likelihood phylogenetic analysis was conducted using PhyML 3.0 software package (Figure 1), based on the alignment of the reconstructed amino acid sequences from each lineage with known reference sequences (Supplementary Table S3), the bootstrap replicates are ranging from 0 to 1 (Guindon et al. 2010). These sequences were aligned using CLUSTAL 2.1 Multiple Sequence Alignments software (Chenna et al. 2003).

Transcriptome data analysis

The transcriptome data analysis was conducted in order to know whether the complete MLE sequences are expressed considering the protocol described by Pertea et al. (2016). In this study, we used six sequence read archive (SRA)-runs of the Whitefly *B. tabaci* MEAM1 transcriptome data under the bioproject accession PRJNA312467. We studied the expression of the two complete MLEs on whiteflies during the first 3 days of acquisition of Tomato yellow leaf curl virus (TYLCV).

The SRA data were first downloaded using NCBI SRA Toolkit, these SRA were mapped using HISAT2 (<http://ccb.jhu.edu/software/hisat2> or <http://github.com/infp/hisat2>) to the Scaffolds

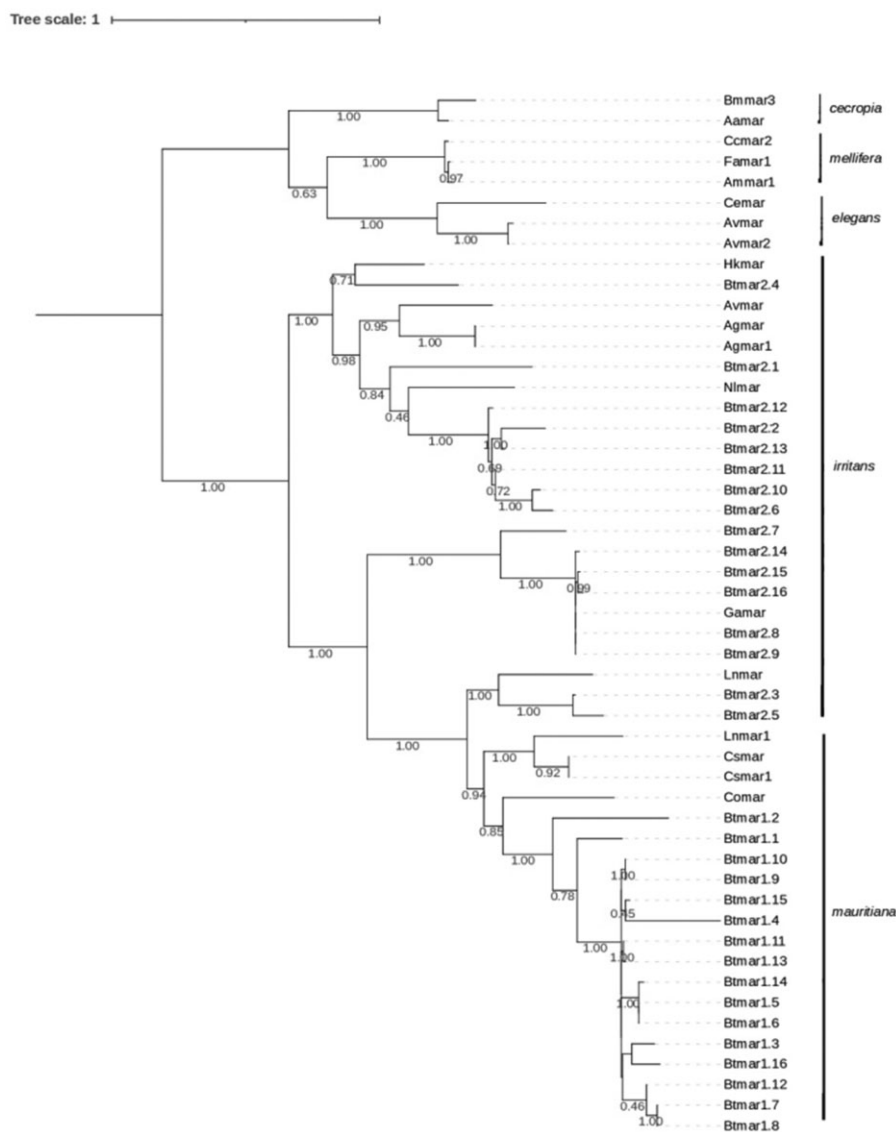


Figure 1 Phylogenetic relationships based on the reconstructed amino acid sequences from each lineage of the identified MLEs and the known reference sequences of mariner elements. The tree was inferred using the maximum-likelihood method with a bootstrap ranging from 0 to 1.

Table 1 Characteristics of *mariner*-like transposons of the *mauritiana* subfamily in the whitefly genome

	Names	Number of related sequences	Sequences length (bp)	Transposase motifs	TIR length (bp)	TSD
Single copy	Btmar1.1 ^a	1	1,573	TGDE-HDNA-YSPDLAPAD	5':180 3':180	5': TA 3': TA
	Btmar1.2	1	1,325	TGDE-HDNA-YSPDLAPAD	5':217 3':217	5': TA 3': TA
	Btmar1.3	1	1,225	YSPDLVPCD	5':287 3':282	—
	Btmar1.4	1	1,118	HHHA-PLPLIHLTLLLV	5':308 3':296	—
	Btmar1.5	1	654	HDNA-YSPDLAPCD	5':343 3':23	—
	Btmar1.6	1	1,188	HDNA-YSPEPDLAPCD	5':446 3':448	5': - 3': TA
Lineages	Btmar1.7	89	631–986	FGDN	5':125 3':125	5': TA 3': TA
	Btmar1.8	76	757–978	FGDN	5':125 3':125	5': TA 3': TA
	Btmar1.9	34	535–657	DLAPCD	5': 130 3': -	5': TA 3': -
	Btmar1.10	33	528–703	DLAPCD	5': - 3':438	5': - 3': TA
	Btmar1.11	23	731–1193	HDNA-YSPDLAPCD	5':439 3':440	5': TA 3': -
	Btmar1.12	17	642–966	FGDN	5':125 3':125	5': TA 3': TA
	Btmar1.13	7	793–1,177	HDNA-YSPNLAPCD	5':444 3':227	5': TA 3': -
	Btmar1.14	4	554–700	HDNA-YLSD	5':122 3': -	5': TA 3': -
	Btmar1.15	3	1,216–1,194	HDNA-YSPDLAPCD	5':459 3':459	5': TA 3': TA
	Btmar1.16	2	1,586–1,705	TGDE-HDNA-CSPDLAPCD	—	—

^a Full-length copy.

where the complete MLEs were identified. The alignments are submitted to StringTie (<http://ccb.jhu.edu/software/stringtie> or <https://github.com/gpertea/stringtie>) to assemble and quantify the transcripts of each sample. Assembled transcripts were merged together by a special StringTie module, which creates a uniform set of transcripts for all samples. In order to compare the genes and transcripts with the reference annotation and reports statistics, we conducted an optional step using gffcompare (<http://ccb.jhu.edu/software/stringtie/gff.shtml> or <http://github.com/gpertea/gffcompare>). Finally, StringTie estimate transcript abundances and create table counts for Ballgown, which is an R package, used to calculate the normalized FPKM (fragments per kilobase of exon per million reads mapped) values and genes with FPKM > 1 (Macko-Podgómi et al. 2021).

Results

Presence of MLEs in the genome of *B. tabaci*

The search for MLEs in the MEAM1/B genome led to the identification of 550 sequences. The total genomic mass of all the MLEs is 0.6 Mb, or about 0.1% of the *B. tabaci* genome size of 615 Mb. The NCBI database investigation using BLASTX algorithm showed that 294 of the identified elements shared the best amino acid identity, greater than 44%, with *mauritiana* elements while the 256 remaining elements shared the best amino acid identity, greater than 52%, with *irritans* elements.

To confirm these latter results, a *maximum-likelihood* phylogenetic analysis of the reconstructed amino acid sequences from each lineage was performed (Figure 1). Repeatedly, *B. tabaci* *Mariner* elements were found in *mauritiana* and *irritans* subfamilies

indicating significant polymorphism among the different lineages of each subfamily.

Mauritiana sequences analysis

The *mauritiana* MLEs clusterization resulted in 6 single copies, i.e., *Btmar1.1* to *Btmar1.6*, and 10 different lineages, i.e., *Btmar1.7* to *Btmar1.16* (Table 1). Clustering analysis using the Cytoscape tool showed that the majority of the *mauritiana* elements were clustered into the lineages *Btmar1.7* (89) and *Btmar1.8* (76) (Table 1). These elements are sized from 631 to 986 bp. All these copies are defective due to mutations in their ORFs, which code for a truncated transposase that have only one mutated catalytic motif. Most of the elements in the *Btmar1.7* and *Btmar1.8* lineages have LTIRs of 125 bp in length and flanked by TA TSDs.

The other *mauritiana* copies are also defective. However, most of their transposase coding sequences retain the second and third aspartic acid of the DD34D catalytic motif. Some elements have both LTIRs retained, ranging in size from 125 to 459 bp. Whereas other elements are truncated copies with only one terminal sequence, ranging in size from 122 to 438 bp. The remainder is fragments not associated with any recognizable LTIRs, suggesting they are old and degenerated. These LTIRs could be flanked or not by TA TSDs. All the *mauritiana* sequences described are truncated with the notable exception for *Btmar1.1*, which is a full-length MLE (Table 1).

Feature analysis of *Btmar1.1*

The *mauritiana* element *Btmar1.1* is placed on Scaffold 26,953, spanning from positions 130,789 to 132,365. This element revealed an ORF sized up to 1077 bp and encoding a putative



Figure 2 Sequence features of the full-length *mauritiana* element *Btmar1.1* (1573 bp) and *Btmar2.1* (1807 bp). The pink and blue arrows at the extremities refer to the TIRs of *Btmar1.1* and *Btmar2.1*, respectively. TA TSDs are written with bold characters located toward the outer end of each TIRs. The transposase ORF is represented in the figure by a yellow box for *Btmar1.1* and green box for *Btmar2.1*, respectively. The three signature motifs of the catalytic domain motifs DD34D are given within the ORF. The bipartite NLS is shown as a blue line, the HTH motif is defined by a blue diamond for the two elements.

protein of 358 amino acids. The comparison of this protein sequence to those in the database revealed identities with various MLE transposases, and the best identity was greater than 53% with the *mauritiana* element from *Cryptotermes secundus* (Sequence ID: PNF20350.1). As shown in Figure 2, the *Btmar1.1* putative transposase displays the three catalytic motifs TGDE, HDNA, YSPDLAPAD in the C-terminal domain and a slightly different WVPREL motif (LVPRLL in *Btmar1.1*), a possible NLS and an HTH motif in the N-terminal domain. The *Btmar1.1* transposase ORF is flanked by 5' and 3' UTRs, nearly perfect 180-bp LTIRs, which show 11 substitutions distributed all along the LTIRs, and TA TSDs.

Irritans sequences analysis

The *irritans* MLEs clusterization resulted in seven single copies, i.e., *Btmar2.1* to *Btmar2.7*, and nine different lineages, i.e., *Btmar2.8* to *Btmar2.16* (Table 2). The majority of the *irritans* elements were clustered into the lineages *Btmar 2.8* (120 elements) and *Btmar 2.9* (87 elements) (Table 2). These elements are ranging from 605 to 990 bp in size. All these copies are defective due to mutations in their ORFs, which encode a truncated transposase that has only the two last catalytic motifs QDNA and YSPDLAPCD. Most of the elements in the *Btmar2.8* and *Btmar2.9* lineages have LTIRs of 210 bp in size and flanked by TA TSDs. The remaining *irritans* copies are also truncated and usually have either LTIRs on both sides with a size ranging from 200 to 237 bp or mutated TIR on one side with a size ranging from 27 to 197 bp. The following are exceptions, *Btmar 2.7*, which has no TIRs, *Btmar2.3* and *Btmar2.4*, which have typical TIRs of 11 and 28 bp, respectively, but mutated, and especially *Btmar2.1*, which is a full-length MLE (Table 2).

Feature analysis of *Btmar2.1*

The *irritans* element *Btmar2.1* is located on Scaffold 831, spanning from positions 3,236,801 to 3,238,607. The comparison of *Btmar1.1* to *Btmar2.1* did not show any significant similarity at the nucleic acid level while it revealed a highly conserved regions corresponding to the three catalytic motifs. *Btmar2.1* displays an ORF up to 1092 bp in size, encoding a putative 363 amino acid protein. The comparison of this protein sequence to those in the database revealed identities with various MLE transposases and the best identity was greater than 48% with the *irritans* element from the coleopteran insect *Anoplophora glabripennis* (Sequence ID: XP_023310047.1). As shown in Figure 2, the *Btmar2.1* putative transposase displays the three catalytic motifs TGDE, HDNA, and YSPDLAPSD in the C-terminal domain and a possible NLS, an HTH motif, and the first signature motif WVPREL in the N-terminal domain. The *Btmar2.1* transposase ORF is flanked by 5' and 3' UTRs, nearly perfect large 200-bp TIRs, and TA TSDs. The *Btmar2.1* LTIRs exhibit fewer substitutions than those of *Btmar1.1*, i.e., only four at positions 18, 179, 193, and 198, and a single deletion of one nucleotide at the 36th position of the 3' LTIR.

MITEs deriving from MLE: detection and sequence analysis

The search for MITEs in the MEAM1/B genome led to the identification of 8024 MITE families. Among these families, 1030 MITE families belong to the *Tc1/mariner* superfamily from which we identified 71 MITE families related to MLE sequences. These 71 MITE sequences showed 12,605 copies, TA TSD patterns, and TIRs. BLASTN (e -value $< 10^{-10}$) sequences analysis revealed that a total of 35 MITEs families derived from *mauritiana* elements and 36 MITEs derived from *irritans* elements. The MITE sequences showed different groups based on their TIR sequences (Figure 3). Most of the MITEs derived from *mauritiana* MLEs originated from mutated MLEs sequences belonging to *Btmar1.7*, *Btmar1.8*, *Btmar1.12*, and *Btmar1.15* lineages. These elements are ranging from 128 to 790 bp in size with TIRs ranging from 4 to 314 bp in size. However, most of the MITEs derived from *irritans* MLEs originated from the full-length element *Btmar2.1*. These elements are ranging from 179 to 796 bp in size with TIRs of 8 to 282 bp.

Sequences alignments revealed that in all the detected MITEs the first 10 base pairs in TIRs are conserved with the exception of 18 copies that have only 4–8 bp conserved in TIRs, while 20 elements have conserved LTIRs with a size exceeding 100 bp and up to 314 bp. All the described MITEs exhibited mutations such as deletions and substitutions targeting the middle of the sequences and extended to the extremities in highly mutated MITEs, which led to the generation of the MITEs with imperfectly conserved TIRs of about 4–20 bp.

Transcriptome analysis of *Btmar1.1* and *Btmar2.1*

Btmar1.1 is inserted in the 9th intronic region of the cytosolic carboxypeptidase2-like gene (LOC109039198), which codes for the CCP2 enzyme involved in the deglutamylation reaction on tubulin, leading to $\Delta 2$ -tubulin (Tort et al. 2014). We tested the expression of this gene in order to see whether *Btmar1.1* is expressed.

The TYLCV transcriptome analysis showed that the LOC109039198 gene has an FPKM higher than one for the dix SRA-runs (Supplementary Table S4), which means that this gene is expressed. Accordingly, the LOC109039198 gene, and then *Btmar1.1*, is also expressed.

The FPKM for *Btmar2.1* is also higher than one in the same transcriptome conditions (Supplementary Table S4), which means that *Btmar2.1* is expressed.

Discussion

In this study, we have shown that the MEAM/B genome exhibit a total of 550 MLEs belonging to both *mauritiana* (*Btmar1.1* to *Btmar1.16*) and *irritans* (*Btmar2.1* to *Btmar2.16*) *Mariner* subfamilies. The majority of *Mariner* sequences are mainly represented by four lineages (*Btmar1.7*, *Btmar1.8*, *Btmar2.8*, and *Btmar2.9*) comprising 67% of the identified *Mariner* elements. Although high copy number lineages have been described in some species such as the Tricladida *Girardia*

Table 2 Characteristics of *mariner*-like transposons of the *irritans* subfamily in the whitefly genome

	Names	Number of related sequences	Sequences length (bp)	Transposase motifs	TIR length (bp)	TSD
Single copy	Btmar2.1^a	1	1,807	TGDE-HDNA-YSPDLAPSD	5':200 3':199	5':TA 3':TA
	<i>Btmar2.2</i>	1	534	QDNA-YSPVLPLSD	—	—
	<i>Btmar2.3</i>	1	1,252	TGDE-HDNA-YSPDLAPCD	5':11 3':11	—
	<i>Btmar2.4</i>	1	1,359	TGDK-HDNA-YSPDLAPSD	5':28 3':28	—
	<i>Btmar2.5</i>	1	515	PDLALCD	5': - 3':197	5': - 3':TA
	<i>Btmar2.6</i>	1	606	YSPMMTHRD	5': - 3':27	5': - 3':TA
	<i>Btmar2.7</i>	1	605	QDNA-YSPDLAPCD	—	—
Lineages	<i>Btmar2.8</i>	120	605–990	QDNA-YSPDLAPCD	5':210 3':210	5':TA 3':TA
	<i>Btmar2.9</i>	87	628–983	QDNA-YSPDLAPCD	5': 210 3': 210	5': TA 3':TA
	<i>Btmar2.10</i>	15	702–1,017	QDNA-SYSPNLAPCD	5': 229 3':225	5': TA 3':TA
	<i>Btmar2.11</i>	13	737–1,018	QDNA-YSPDLTPSD	5':237 3':237	5':TA 3':TA
	<i>Btmar2.12</i>	5	743–996	QDNA-YSPDLPSD	5':233 3':235	5':TA 3':TA
	<i>Btmar2.13</i>	3	986–1,000	QDNA-YSPDLPPSD	5':233 3':233	5':TA 3':TA
	<i>Btmar2.14</i>	2	544–573	QDNA-YSPDLAPCD	5': - 3': 197	5': - 3': TA
	<i>Btmar2.15</i>	2	527–559	YSADLAPCN	5': 197 3': -	5':TA 3': -
	<i>Btmar2.16</i>	2	528–534	NSADLAPCD	5': - 3':197	5': - 3':TA

^a Full-length copy.

tigrina, the Hemiptera *Rhodnius prolixus* (García-Fernández et al. 1995; Filée et al. 2015; Fernández-Medina et al. 2016), and the *Drosophila* genus genomes (Wallau et al. 2014), MLEs are usually described as a low copy number family of TEs and most of the other genomes display less than 50 copies per genome, hence, the *B. tabaci* genome seems to be rather permissive for MLEs expansion.

As previously described, the typical size of MLEs is 1300 bp (Robertson and Martos 1997), which is consistent with the size of most of the identified MLEs described in this study. Among the 550 identified elements, only 4 MLEs sequences have an unusual length, which exceeded the known size by 277–507 bp. However, analysis of the TIRs has shown that their length is extended (up to 459 bp) in most sequences. This particularity is usually observed among the closely related TLE transposons (Robertson, 1995; Shao and Tu 2001; Ruvolo et al. 1992). Although TLE have variable TIRs size, ranging from 24 to 756 bp, most of them are over 50 bp in length (Ruvolo et al. 1992; Jehle et al. 1998). In addition, MLEs with LTIRs have been described previously in the genome of the phytoparasitic nematode *M. chitwoodi*, and these elements, namely *Mcmar1-1* and *Mcmar1-2*, present perfectly conserved LTIRs of 355 bp (Leroy et al. 2003).

In this study, most of the MLEs sequences are degenerated copies with in-frame stop codons and frameshift, or indel mutations. As reported previously, these mutations are sufficient to abolish the catalytic activity or prevent the synthesis of the putative transposase and make the MLEs inactive or nonautonomous (Lohe et al. 1995). Importantly, our results have also highlighted two complete copies (from TSD to TSD) with putative conserved transposases, which make them probably capable of their own transposition and the transposition of close or unrelated nonautonomous elements via trans-complementation (Lohe et al. 1995).

As previously reported, MITEs can be derived from autonomous or nonautonomous elements due to mechanisms such as abortive gap repair (de Ortiz et al. 2010; Fattash et al. 2013). However, the loss or the inactivation of their original autonomous elements does not make them imperatively inactive, a few MITE sequences can be cross-mobilized using nearly noncognate autonomous elements machinery (Fattash et al. 2013). For example, the loss of the original autonomous element Ping did not stop the activity of the MITE element mPing and this element can be cross-mobilized via the transposase of other elements such as Pong (Yang et al. 2006). Moreover, a *mariner* derived Stowaway-like MITE element, named *MiHsmar1* has been mobilized using the reconstructed *Hsmar1* transposase (Miskey et al. 2007). For most of the identified MITEs-MLE sequences, we have been able to retrieve relatively intact TIR sequences and TA TSDs needed for the transposition. The TIRs presented high similarities with their related *mariner* subfamily. These findings support that in *B. tabaci* genome, MITEs can be mobilized through the transposase of *mariner* lineages that share almost identical TIR patterns.

MITEs have been widely described in plant genomes as small elements shorter than 800 bp and with a high copy number reaching thousands of copies (Guermónprez et al. 2008; Crescente et al. 2018). MITEs have also been characterized in insects. For example, the genome of *Drosophila sechellia* harbors only 46 MITE sequences (Dias and Carareto 2011). However, in the genome of *Aedes aegypti* the MITE copy numbers range from 2100 to 3000 per haploid genome (Tu 1997). In a more recent study, Han et al. (2016) identified 6012 MITE families from the genome of 98 insect species and constructed the first insect MITE database (iMITEdb). In this database, the number of the MITE families belonging to the *Tc1/mariner* superfamily is 1698 families from 90 insect species genomes, among which 192

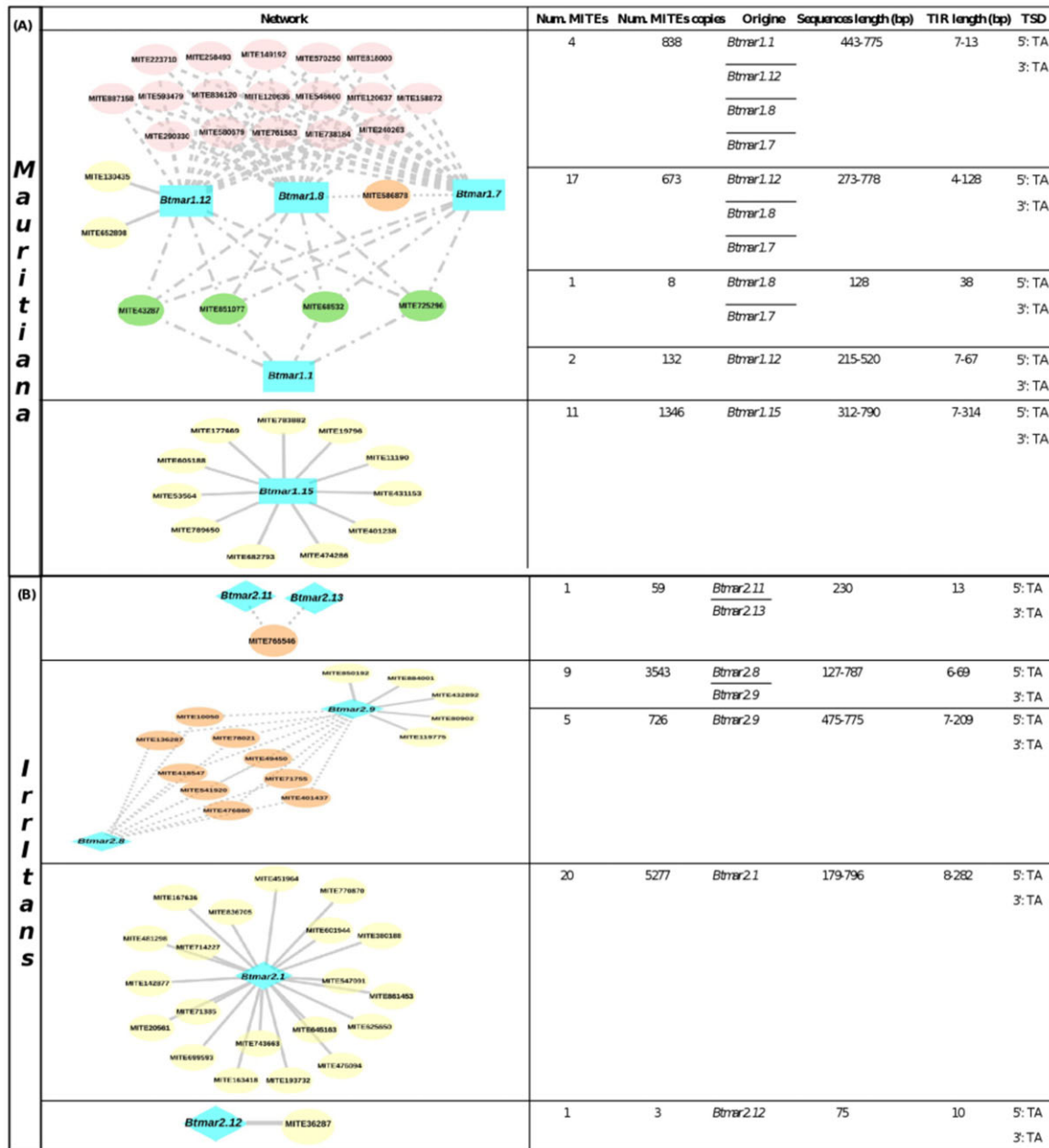


Figure 3 Characteristics of MLE-derived MITEs in the whitefly genome. (A) MITEs derived from the *mauritiana* subfamily. (B) MITEs derived from the *irritans* subfamily. Light blue rectangles refer to *mauritiana* MLEs while light blue diamonds refer to *irritans* MLEs. MITEs are represented by ellipses. Green ellipses mean that MITEs have four possible counterparts but they are derived from one of them. Pink ellipses mean that MITEs have three possible MLEs origins, orange and yellow ellipses indicate that MITEs have two or one possible MLEs origins.

families from 8 Hemiptera genomes. Based on the data obtained in this study, the relative abundance of MITEs in the *B. tabaci* genome is high compared to other insect genomes since the number of the identified MITE related to MLE sequences exceeds the total MITE copies number described in the majority of the insects. Although, most of the identified MITEs did not originate from autonomous elements, MITEs related to full-length TEs have been described in some species such as *D. sechellia* (Dias and Carareto 2011). Here, we also highlight four copies of MITEs related to the full-length element *mauritiana* and 20 related to the *irritans* full-length element.

Conclusion

This work represents the first *in silico* characterization and classification of MLEs and their related MITEs in the genome of *B. tabaci*. The results highlight the presence of LTIR MLEs and their related MITEs, as well as the presence of two full-length MLEs that may be involved in the transposition of nonautonomous elements. These results provide new data for further analysis to understand the involvement of TEs in the adaptive evolution of the whitefly.

Data availability

Supplementary materials are available at GSA figshare portal: <https://doi.org/10.25387/g3.15001164>. Supplementary Tables S1 and S2 contain nucleic acid and transposase MLEs queries for BLASTN and tBLASTN searches, respectively. Supplementary Table S3 contains phylogenetic tree references characteristics. Supplementary Table S4 contains SRA data analysis of the expression of the two complete MLEs in *Bemisia tabaci*. All identified MLE and MITE sequences are available in text files S1 and S2, respectively. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Funding

This work was financially supported by the « PHC Utique » program of the French Ministry of Foreign Affairs and Ministry of Higher Education, Research and Innovation and the Tunisian Ministry of higher education and scientific research in the CMCU project number « 19G0901 ».

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

- Abd-Rabou S, Simmons AM. 2010. Survey of reproductive host plants of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Egypt, including new host records. *Entomol News*. 121:456–465. doi:10.3157/021.121.0507.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410. doi:10.1016/S0022-2836(05)80360-2.
- Augé-Gouillou C, Brillet B, Germon S, Hamelin MH, Bigot Y. 2005. Mariner Mos1 transposase dimerizes prior to ITR binding. *J Mol Biol*. 351:117–130. doi:10.1016/j.jmb.2005.05.019.
- Barry EG, Witherspoon DJ, Lampe DJ. 2004. A bacterial genetic screen identifies functional coding sequences of the insect mariner transposable element famar1 amplified from the genome of the earwig, *Forficula auricularia*. *Genetics*. 166:823–833. doi:10.1534/genetics.166.2.823.
- Berthelie J, Casse N, Daccord N, Jamilloux V, Saint-Jean B, et al. 2018. A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. *BMC Genomics*. 19:1–14. doi:10.1186/s12864-018-4763-1.
- Bigot Y, Brillet B, Augé-Gouillou C. 2005. Conservation of palindromic and mirror motifs within inverted terminal repeats of mariner-like elements. *J Mol Biol*. 351:108–116. doi:10.1016/j.jmb.2005.05.006.
- Bouallègue M, Filée J, Kharrat I, Mezghani-Khemakhem M, Rouault JD, et al. 2017. Diversity and evolution of mariner-like elements in aphid genomes. *BMC Genomics*. 18:1–12. doi:10.1186/s12864-017-3856-6.
- Brillet B, Benjamin B, Bigot Y, Yves B, Augé-Gouillou C, et al. 2007. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica*. 130:105–120. doi:10.1007/s10709-006-0025-2.
- Casola C, Lawing AM, Betrán E, Feschotte C. 2007. PIF-like transposons are common in *Drosophila* and have been repeatedly domesticated to generate new host genes. *Mol Biol Evol*. 24:1872–1888. doi:10.1093/molbev/msm116.
- Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, et al. 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol*. 14:1–15. doi:10.1186/s12915-016-0321-y.
- Chénais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*. 509:7–15. doi:10.1016/j.gene.2012.07.042.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 31:3497–3500. doi:10.1093/nar/gkg500.
- Corpet F. 1988. Multiples sequence alignment with hierarchical clustering. *Nucleic Acids Research*. 16(22):10881–10890. doi:10.1093/nar/16.22.10881
- Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics*. 19:10. doi:10.1186/s12859-018-2376-y.
- Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from MU to kangaroo. *Nat Rev Mol Cell Biol*. 4:865–877. doi:10.1038/nrm1241.
- de Ortiz MF, Lorenzatto KR, Corrêa BRS, Loreto ELS. 2010. hAT transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica*. 138:649–655. doi:10.1007/s10709-010-9439-y.
- Dias ES, Carareto CMA. 2011. MsechBari, a new MITE-like element in *Drosophila sechellia* related to the Bari transposon. *Genet Res (Camb)*. 93:381–385. doi:10.1017/S0016672311000371.
- Dodd IB, Egan JB. 1990. Detection of helix-turn-helix DNA-binding motifs. *Nucleic Acids Res*. 18:5019–5026.
- Drezen JM, Gauthier J, Josse T, Bézier A, Herniou E, et al. 2017. Foreign DNA acquisition by invertebrate genomes. *J Invertebr Pathol*. 147:157–168. doi:10.1016/j.jip.2016.09.004.
- Dupeyron M, Baril T, Bass C, Hayward A. 2020. Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored diversity of pogo-like elements. *Mob DNA*. 11:21. doi:10.1186/s13100-020-00212-0.
- Emmons SW, Yesner L, Ruan KS, Katzenberg D. 1983. Evidence for a transposon in *Caenorhabditis elegans*. *Cell*. 32:55–65. doi:10.1016/0092-8674(83)90496-8.
- Fattash I, Rooke R, Wong A, Hui C, Luu T, et al. 2013. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. *Genome*. 486:475–486.
- Fernández-Medina RD, Granzotto A, Ribeiro JM, Carareto CMA. 2016. Transposition burst of mariner-like elements in the sequenced genome of *Rhodnius prolixus*. *Insect Biochem Mol Biol*. 69:14–24. doi:10.1016/j.ibmb.2015.09.003.
- Feschotte C, Swamy L, Wessler SR. 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.
- Filée J, Rouault JD, Harry M, Hua-Van A. 2015. Mariner transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. *BMC Genomics*. 16:1061. doi:10.1186/s12864-015-2060-9.
- Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet*. 5:103–107.
- Gahan LJ, Gould F, Heckel DG. 2001. Identification of a gene associated with Bt resistance in *Heliothis virescens*. *Science*. 293:857–860. doi:10.1126/science.1060949.
- García-Fernández J, Bayascas-Ramírez JR, Marfany G, Munoz-Marmol AM, Casali A, et al. 1995. High copy number of highly similar mariner-like transposons in planarian (platyhelminth): evidence for a trans-phyla horizontal transfer. *Mol Biol Evol*. 12:421–431. doi:10.1093/oxfordjournals.molbev.a040217.

- Gomulski LM, Torti C, Bonizzoni M, Moralli D, Raimondi E, et al. 2001. A new basal subfamily of mariner elements in *Ceratitis rosa* and other tephritid flies. *J Mol Evol.* 53:597–606. doi:10.1007/s002390010246.
- González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol.* 6:2109–2129. doi:10.1371/journal.pbio.0060251.
- Grace CA, Carr M. 2020. The evolutionary history of mariner elements in stalk-eyed flies reveals the horizontal transfer of transposons from insects into the genome of the cnidarian *Hydra vulgaris*. *PLoS One.* 15:1–24. doi:10.1371/journal.pone.0235984.
- Guermónprez H, Loot C, Casacuberta JM. 2008. Different strategies to persist: the pogo-like Lemi1 transposon produces miniature inverted-repeat transposable elements or typical defective elements in different plant genomes. *Genetics.* 180:83–92. doi:10.1534/genetics.108.089615.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Han MJ, Zhou QZ, Zhang HH, Tong X, Lu C, et al. 2016. IMITEdb: the genome-wide landscape of miniature inverted-repeat transposable elements in insects. *Database.* 2016:1–7. doi:10.1093/database/baw148.
- Hoen DR, Hickey G, Bourque G, Casacuberta J, Cordaux R, et al. 2015. A call for benchmarking transposable element annotation methods. *Mobile DNA.* 6:1. doi:10.1186/s13100-015-0044-6.
- Horowitz AR, Ghanim M, Roditakis E, Nauen R, Ishaaya I. 2020. Insecticide resistance and its management in *Bemisia tabaci* species. *J Pest Sci.* 93:893–910. doi:10.1007/s10340-020-01210-0.
- Itokawa K, Komagata O, Kasai S, Okamura Y, Masada M, et al. 2010. Genomic structures of Cyp9m10 in pyrethroid resistant and susceptible strains of *Culex quinquefasciatus*. *Insect Biochem Mol Biol.* 40:631–640. doi:10.1016/j.ibmb.2010.06.001.
- Jacobson JW, Medhora MM, Hartl DL. 1986. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc Natl Acad Sci U S A.* 83:8684–8688. doi:10.1073/pnas.83.22.8684.
- Jangam D, Feschotte C, Betrán E. 2017. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* 33:817–831. doi:10.1016/j.tig.2017.07.011.
- Jehle JA, Nickel A, Vlak JM, Backhaus H. 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella granulovirus*. *J Mol Evol.* 46:215–224. doi:10.1007/PL00006296.
- Klai K, Chénaïs B, Zidi M, Djebbi S, Caruso A, et al. 2020. Screening of *Helicoverpa armigera* mobilome revealed transposable element insertions in insecticide resistance genes. *Insects.* 11:879. doi:10.3390/insects11120879.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 35:1547–1549. doi:10.1093/molbev/msy096.
- Leroy HE, Castagnone-Sereno P, Renault S, Augé-Gouillou C, Bigot Y, et al. 2003. Characterization of Mcmar1, a mariner-like element with large inverted terminal repeats (ITRs) from the phytoparasitic nematode *Meloidogyne chitwoodi*. *Gene.* 304:35–41. doi:10.1016/S0378-1119(02)01144-7.
- Lin J-R, Hu J. 2013. SeqNLS: nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. *PLoS One.* 8:e76864. doi:10.1371/journal.pone.0076864.
- Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol Biol Evol.* 12:62–72. doi:10.1093/oxfordjournals.molbev.a040191.
- Lu C, Chen J, Zhang Y, Hu Q, Su W, et al. 2012. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol Biol Evol.* 29:1005–1017. doi:10.1093/molbev/msr282.
- Macko-Podgórní A, Machaj G, Grzebelus D. 2021. A global landscape of miniature inverted-repeat transposable elements in the carrot genome. *Genes.* 12:859. doi:10.3390/genes12060859.
- Mat Razali N, Cheah B, Nadarajah K. 2019. Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. *Int J Mol Sci.* 20:3597. doi:10.3390/ijms20143597.
- Miskey C, Papp B, Mátés L, Sinzelle L, Keller H, et al. 2007. The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol Cell Biol.* 27:4589–4600. doi:10.1128/mcb.02027-06.
- Munoz-Lopez M, Garcia-Perez J. 2010. DNA transposons: nature and applications in genomics. *Curr Genomics.* 11:115–128. doi:10.2174/138920210790886871.
- Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. 2011. Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol.* 49:219–248. doi:10.1146/annurev-phyto-072910-095235.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 11:1650–1667. doi:10.1038/nprot.2016.095.
- Piégu B, Bire S, Arensburg P, Bigot Y. 2015. A survey of transposable element classification systems—a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol Phylogenet Evol.* 86:90–109. doi:10.1016/j.ympev.2015.03.009.
- Plasterk RHA, Izsvák Z, Ivics Z. 1999. Resident aliens the Tc1/mariner superfamily of transposable elements. *Trends Genet.* 15:326–332. doi:10.1016/S0168-9525(99)01777-1.
- Robertson HM. 2002. Evolution of DNA transposons in eukaryotes. *Mobile DNA.* II:1093–1110. doi:10.1128/9781555817954.ch48.
- Robertson HM, Martos R. 1997. Molecular evolution of the second ancient human mariner transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome lineage. *Gene.* 205:219–228. doi:10.1016/S0378-1119(97)00471-X.
- Robertson M. 1995. Mini-review the Tel-mariner superfamily transposons in animals. *Science.* 41:99–105.
- Rouault JD, Casse N, Chénaïs B, Hua-Van A, Filée J, et al. 2009. Automatic classification within families of transposable elements: application to the mariner Family. *Gene.* 448:227–232. doi:10.1016/j.gene.2009.08.009.
- Ruvolo V, Hill JE, Levitt A. 1992. The Tc2 transposon of *Caenorhabditis elegans* has the structure of a self-regulated element. *DNA Cell Biol.* 11:111–122. doi:10.1089/dna.1992.11.111.
- Sang Y, Gao B, Diaby M, Zong W, Chen C, et al. 2019. Incomer, a DD36E family of Tc1/mariner transposons newly discovered in animals. *Mob DNA.* 10:1–12. doi:10.1186/s13100-019-0188-x.
- Sanlloriente O, Vela J, Mora P, Ruiz-Mena A, Torres MI, et al. 2020. Complex evolutionary history of Mboumar, a mariner element widely represented in ant genomes. *Sci Rep.* 10:2610. doi:10.1038/s41598-020-59422-4.
- Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. *Mol Ecol.* 28:1537–1549. doi:10.1111/mec.14794.
- Shao H, Tu Z. 2001. Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics.* 159:1103–1115.

- Tort O, Tanco S, Rocha C, Bièche I, Seixas C, et al. 2014. The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids. *Mol Biol Cell*. 25:3017–3027. doi: 10.1091/mbc.E14-06-1072.
- Tu Z. 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A*. 94:7475–7480.
- Wallau GL, Capy P, Loreto E, Hua-Van A. 2014. Genomic landscape and evolutionary dynamics of mariner transposable elements within the *Drosophila* genus. *BMC Genomics*. 15:727. doi: 10.1186/1471-2164-15-727.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. 2007. Reply: a unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nat Rev Genet*. 10: 276. doi:10.1038/nrg2165-c4.
- Xie LQ, Wang PL, Jiang SH, Zhang Z, Zhang HH. 2018. Genome-wide identification and evolution of TC1/Mariner in the silkworm (*Bombyx mori*) genome. *Genes Genomics*. 40:485–495. doi: 10.1007/s13258-018-0648-6.
- Yang G, Weil CF, Wessler SR. 2006. A rice Tc1/mariner-like element transposes in yeast. *Plant Cell*. 18:2469–2478. doi:10.1105/tpc.106.045906.
- Yuan YW, Wessler SR. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A*. 108:7884–7889. doi:10.1073/pnas.1104208108.
- Zhang HH, Li GY, Xiong XM, Han MJ, Zhang XG, et al. 2016. Y2016 TRT, a vertebrate and protozoan tc1-like transposon: current activity and horizontal transfer. *Genome Biol Evol*. 8:2994–3005. doi:10.1093/gbe/evw213.
- Zong W, Gao B, Diaby M, Shen D, Wang S, et al. 2020. Traveler, a new DD35E family of Tc1/Mariner transposons, invaded vertebrates very recently. *Genome Biol Evol*. 12:66–76. doi:10.1093/gbe/evaa034.

Communicating editor: J. Blumenstiel