EUROPEAN JOURNAL OF ENTOMOLOGY
ISSN (online): 1802-8829
http://www.eje.cz
*Eur. J. Entomol.* **114**: 379–390, 2017
doi: 10.14411/eje.2017.049
ORIGINAL ARTICLE

# An overview of *irritans-mariner* transposons in two *Mayetiola* species (Diptera: Cecidomyiidae)

**Wiem BEN AMARA[1], Salma DJEBBI[1], Dhia BOUKTILA[1,2], Mohamed MAKNI[1], Hanem MAKNI[1,3] and Maha MEZGHANI-KHEMAKHEM[1,*]**

[1] Unité de Recherche Génomique des Insectes Ravageurs des Cultures d'intérêt agronomique (UR11ES10), Faculté des Sciences de Tunis, Université de Tunis El Manar, 2092 El Manar, Tunis, Tunisia; e-mails: wiem.benamara7@gmail.com, salma.djebbi@fst.utm.tn, dhia_bouktila2000@yahoo.fr, mdmakni98@gmail.com, hanem.makni@isajc.rnu.tn, maha.mezghani@fst.utm.tn

[2] Institut Supérieur de Biotechnologie de Béja (ISBB) – Université de Jendouba, Tunisia

[3] Institut Supérieur de l'Animation pour la Jeunesse et la Culture (ISAJC), Université de Tunis, Bir El Bey, Tunisia

**Key words.** Diptera, Cecidomyiidae, *Mayetiola destructor*, *Mayetiola hordei*, *Sitodiplosis mosellana*, *irritans* subfamily, in silico analysis, in vitro experiments, *mariner*-like element, transposons

**Abstract.** *Mariner*-like elements (MLEs) are widespread Class II transposable elements in insects that are subdivided into several subfamilies. In the current study, we carried out in silico analysis and in vitro experiments to identify MLEs belonging to the *irritans* subfamily in two cecidomyiid flies, *Mayetiola destructor* and *M. hordei*. In silico investigation of *M. destructor* genome allowed the identification of 25 *irritans*-like elements, which were mostly defective due to several mutations. These defective forms might be the remnants of active elements that ancestrally invaded the host genome. Structural analyses, including signature motifs and transposase-encoding ORFs, revealed structural heterogeneity and the presence of two full length copies. Five consensuses, reflecting the probable evolutionary groups of these elements, were constructed, based on a similarity matrix. The first consensus (*Maymarcons1*) belonged to Himar1-like elements reported in other insects, while the remaining four (*Maymarcons2* to *5*) seemed to be more specific to Cecidomyiidae. Moreover, the presence of elements belonging to the *Maymarcons4* group was ascertained by PCR amplification, in both *Mayetiola* species, and was further identified in the Transcriptome Shotgun Assembly (TSA) of the orange fly, *Sitodiplosis mosellana* (Cecidomyiidae), suggesting the existence of *irritans* elements within the Cecidomyiidae, which were derived from an ancestral species by vertical transmission during speciation. On the other hand, consensuses that are specific to *M. destructor* could be derived from a more recent invasion. This study suggests that both *M. destructor* and *M. hordei* genomes have been invaded by *irritans* elements many times with at least two different evolutionary histories.

## INTRODUCTION

Transposable elements (TEs) are repeated DNA sequences that are able to move from one site to another in a host genome. These mobile elements are ubiquitous in almost all organisms from different kingdoms and with different proportions depending on species (Chenais et al., 2012). TEs are not simply selfish DNA but rather important elements that contribute significantly to genome evolution as well as its shape architecture (Feschotte & Pritham, 2007; Bire & Rouleux-Bonnin, 2012; Hirsch & Springer, 2017). TEs are subdivided into two main classes based on their mechanisms of transposition (Finnegan, 1989; Wicker et al., 2007). Class I elements, also known as retrotransposons, transpose via an RNA intermediate according to the "copy and paste" model. Class II elements, also named transposons move via a DNA intermediate according to the "cut and paste" model. Each of these classes is subdivided into subclasses, superfamilies, families and subfamilies (Wicker et al., 2007; Piégu et al., 2015; Arensburger et al., 2016).

*Mariner*-like elements (MLEs) are Class II transposons belonging to the large IS630-*Tc1-mariner* superfamily i.e. ITm (Plasterk, 1996; Plasterk et al., 1999) and known to be widespread in most eukaryotic organisms including insects (Robertson, 1993).They are subdivided into five major subfamilies based on their sequence similarities and phylogenetic relationships: *mauritiana*, *cecropia*, *mellifera*/*capitata*, *elegans*/*briggsae* and *irritans* (Robertson & MacLeod, 1993; Bigot et al., 2005). The latter subfamily is characterized, at least, by four major characteristic lineages. The first lineage corresponds to the *Hsmar2*-like elements in chordates and primates, the second contains

* Corresponding author; e-mail: maha.mezghani@fst.utm.tn

the *Himar1*-like elements in insects, the third includes the *Bytmar1*-like elements in marine organisms and the fourth corresponds to the *Batmar2*-like elements in bats (Sinzelle et al., 2006; Bui et al., 2007).

MLEs are characterized by a typical sequence of 1300 bp in length with terminal inverted repeats (TIRs) of 20–40 bp (Halaimia-Toumi et al., 2004). The MLE TIRs have conserved motifs, such as 5'YYAGRT3' at their extremities, which correspond to the cleavage signal (Bigot et al., 2005). Nevertheless, there are exceptions recorded in at least two *irritans* transposons, namely *Hsmar2* in *Homo sapiens* (Robertson & Martos, 1997) and *Bytmar1* in the hydrothermal crab *Bythograea thermydron* (Halaimia-Toumi et al., 2004), where the distal motif is modified. The MLE TIRs flank one intronless open reading frame (ORF), which encodes a transposase of approximately 350 amino acid residues. This enzyme mediates all transposition steps and allows the integration of the excised MLE in its TA hallmark dinucleotide target site duplication (TSD) (Plasterk et al., 1999; Munoz-Lopez & Garcia-Perez, 2010). The *mariner* transposase exhibits two signature motifs WVPHEL and YSPDLAP (Robertson, 1993). It is also characterized by an N-terminal domain containing the helix-turn-helix motif (HTH), which serves to bind TIRs during the transposition process (Pietrokovski & Henikoff, 1997), as well as a C-terminal catalytic domain containing a DD34D catalytic triad catalyzing the cleavage of the TE and its integration into the TSD (Brillet et al., 2007; Yuan & Wessler, 2011). The three aspartate residues are generally anchored to three conserved motifs named respectively TG**D**EKW (TGDETW for the *irritans* subfamily), HH**D**NA and YSPDLAPS/C**D**. The *mariner* transposase is also characterized by nuclear localization signal (NLS) motifs that transport the transposase through the nuclear envelope (Brillet et al., 2007).

Each TE undergoes different steps during its life cycle. In fact, when a MLE invades a new host genome, it has to increase its copy number by many amplifications (Hartl et al., 1997; Le Rouzic & Capy, 2005). The amplification and propagation of such elements may be deleterious for the host genome, which, consequently, develops control strategies to reduce and even inhibit transposon activity. There are two main ways of control. The first is vertical inactivation, which consist of the accumulation of mutations such as frameshifts, nonsense mutations, insertions and deletions (indels) leading to inactive and fossil elements (Lohe et al., 1995; Hartl et al., 1997). The second is the stochastic loss strategy consisting in the autonomous and non-autonomous elimination of MLEs by genetic drift (Lohe et al., 1995; Kidwell & Lisch, 2001). More recently, transposon silencing has proved to be closely related to epigenetic mechanisms including small RNA molecules (siRNA and piRNA) and methylation that control transposon transcription and transposition (Rigal & Mathieu, 2011; Bucher et al., 2012). Thus, in order to escape the host genome selection pressure, MLEs may invade new host genomes by horizontal transfer (HT) as described in several insects (Lampe et al., 2003; Panaud, 2016; Peccoud et al., 2017).

The identified MLEs are mostly inactive owing to mutations affecting different parts of the elements and it has also been shown that many defective copies contain internal deletions that occur non randomly as ascertained by small direct repeats (SDRs) called microhomologies bordering deletion break points (BPs) (Brunet et al., 2002; Kharrat et al., 2015; Ben Lazhar-Ajroud et al., 2016).

Among the identified MLEs, only three elements are naturally active: *Mos1* in *Drosophila mauritiana* (Jacobson et al., 1986), *Famar1* in *Forficula auricularia* (Barry et al., 2004) and *Mboumar9* in *Messor bouvieri* (Munoz-Lopez et al., 2008). The *Himar1* element, in the horn fly *Haematobia irritans*, is also an active element that was artificially constructed from inactive copies (Robertson & Lampe 1995; Lampe et al., 1996 ).

The ability of TEs to move enabled them to be used as genetic tools for mutagenesis and transgenesis in several organisms, such as insects (Largaespada, 2003; Ryder & Russell, 2003; Handler & O'Brochta, 2012).The choice of appropriate TEs as transgenetic vectors depends on the TEs present in the target genome since the use of endogenous TEs as genetic tools could result in trans-mobilization and therefore the instability of the host genome (Ashburner et al., 1998). Thus, it is important to study and identify the different TE groups and variants existing in a given genome.

In this study, we focused on two species of Cecidomyiidae; *Mayetiola destructor* (Say, 1817) and *Mayetiola hordei* (Kieffer, 1909), which are both major pests of wheat and barley around the world. Previous studies identified a full length MLE copy with intact ORF and perfect TIRs in *M. destructor* (Russell & Shukle, 1997). This element, named *Desmar1*, belongs to the *mauritiana* subfamily and has already been used to study its insertion polymorphisms (Behura et al., 2010). Moreover, an internal region belonging to the *irritans* subfamily has been characterized and named *Des2* (Shukle & Russell, 1995).

Therefore, the aim of this study was to identify and characterize complete *irritans* elements in *M. destructor* and its closely related species *M. hordei*. A combination of in silico and in vitro investigations was carried out and the results used to provide a better overview of the endogenous *irritans* subfamily in these two cereal pests, which is useful in light of the estimation of these MLEs dynamics and evolutionary history.

## MATERIALS AND METHODS

### Insect sampling

Samples of *Mayetiola destructor* and *M. hordei* were collected in the third instar larvae and the flax-seed stages of development on wheat and barley. Total DNA was extracted from individual insects using the salting-out protocol (Sunnucks & Hales, 1996). Subsequently, samples from both species of *Mayetiola* species were identified, based on PCR-RFLP of the cytochrome b gene as reported by Mezghani Khemakhem et al. (2002).

### Data sources

The Mdes1.0 release of the Great Plains (GP) *M. destructor* genome was used for the identification of *irritans*-like ele-

ments. The *Mayetiola destructor* genome is available in GenBank (NCBI BioProject PRJNA45867). It consists of 26 million reads (34-fold genome coverage) sequenced using the whole genome shotgun (WGS) strategy and assembled in 36,371 contigs with a 14 kb contig N50 length and 24,475 scaffolds with a 756 kb N50 length. The sequenced fraction constitutes 153 Mb with 33 Mb of gaps between contigs, distributed across the *M. destructor*'s four chromosomes.

The transcriptome shotgun assembly (TSA) of the orange wheat blossom midge *Sitodiplosis mosellana* (Diptera: Cecidomyiidae) was also used to search for *irritans*-like elements similar to those in *M. destructor*. The *S. mosellana* transcriptome is available in GenBank (NCBI Bioproject PRJNA192921) and consists of 24383 complementary DNA (cDNA) contigs.

### In silico identification of *irritans*-like transposable elements

*Irritans*-like elements were identified in the WGS scaffolds of *M. destructor* using both the structure-based method with the *irritans* transposase typical TGDETW motif and the homology-based method using TBLASTN and BLASTN algorithms (https://blast.ncbi.nlm.nih.gov/) with reference to *irritans* transposases and transposons as queries (Table S1).

Genomic contigs exhibiting similarities with queries (*E*-value < $E^{-10}$) were identified. In order to extract complete transposon copies, TIRs and TSDs were searched for by extending DNA hits by 1000 bp upstream and downstream of the transposase open reading frame (ORF) and aligning the 5'extension with the reverse complement of the 3' extension. A filtration step was then performed by eliminating copies exhibiting an incomplete ORF because of gaps between contigs in the WGS assembly as well as fossil copies whose sizes are less than 300 bp.

### Amplification of *irritans*-like transposable elements

The *irritans*-like elements from samples of *M. destructor* and *M. hordei* were amplified using five TIR primers designed from the alignment of *irritans* copies previously identified in silico from the *M. destructor* genome (Table 1). The PCR conditions were programmed as follows: an initial denaturing step at 94°C for 5 min followed by 40 cycles of 3 steps: denaturing at 94°C for 1 min, annealing at 50°C to 56°C for 1 min, extension at 72°C for 1 min 30 s and a final extension step at 72°C for 10 min. PCR amplification products were visualized on 1% agarose gel stained with ethidium bromide.

### Cloning of PCR products and sequencing

PCR products were excised from agarose gel and purified using Wizard SV Gel and PCR Clean-Up System kits (Promega, Madison, WI, USA) according to the manufacturer's protocol. Purified DNA was then cloned in a pGEMT-easy Vector System (Promega) and used to transform chimio-competent *E. coli DH5α* strains. Colonies were then screened as described by Sambrook et al. (2011). Plasmids were extracted from positive colonies (Wizard Minipreps, Promega) and inserts were amplified using T7 and SP6 primers. The same primers were used to sequence, in both directions, the amplified inserts on an automated sequencer (ABI PRISM 3100 Genetic Analyzer, Applied Biosystems, Foster City, CA, USA). Sequences validated as MLEs were then named according to the nomenclature proposed by Robertson & Asplund (1996) and used in further analysis.

**Table 1.** Characteristics of primer sequences used for PCR amplification of cytochrome b gene and *irritans*-like elements.

| Primers | Sequence (5' → 3') | Ta (°C) |
|---|---|---|
| Cytochrome b gene | | |
| CP1 | 5'GAT GAT GAA ATT GGA TC3' | 53°C |
| CP2 | 5'CTA ATG CAA TAA CTC CTC C3' | 53°C |
| *Irritans* elements | | |
| *IrrMay1* (*Maymarcons1*) | 5'CTC GCG GTT CAT TAT ATR TTC C3' | 50°C |
| *IrrMay2* (*Maymarcons2*) | 5'CAG AAY YTW TTR AAA AAA YS3' | 50°C |
| *IrrMay3* (*Maymarcons3*) | 5'AAA ATR YTA CTA TGA WCA AAA AT3' | 54°C |
| *IrrMay4* (*Maymarcons4*) | 5'ACA TAC TAC TGT GTC CAA ATA TG3' | 56°C |
| *IrrMay5* (*Maymarcons5*) | 5'TAC TAC TGT GAT CAA ATT GAA AG3' | 56°C |

### Sequences and phylogenetic analyses

For the analysis of transposon sequences, similarities and annotations were carried out using BLAST programs (Altschul et al., 1990) in the NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Nucleotide sequences were aligned and an identity matrix was established using an iterative method for multiple sequence alignments. This method was carried out using the MAFFT program (Katoh & Standley, 2013) available in the EMBL-EBI bioinformatics Web Services (http://www.ebi.ac.uk/Tools/msa/mafft/). Nucleotide sequence alignments allowed the construction of consensuses using Jalview 2.10.0 release (Waterhouse et al., 2009). The construction was done to fit the most complete sequence. *Irritans*-like elements were conceptually translated using Mobyle SNAP Workbench web server (Monacell & Carbone, 2014) (http://mobyle.pasteur.fr, last accessed 2016) and putative transposases were manually edited for frameshift and gap insertions. Putative Helix-turn-helix (HTH) motif and nuclear localization signal (NLS) were searched for using the PRABI web server (Combet et al., 2000) (https://npsa-prabi.ibcp.fr, last accessed 2016) and SeqNLS (Lin & Hu, 2013) (http://mleg.cse.sc.edu/seqNLS/), respectively. The identity of NLS motifs was verified by reference to known *mariner* N-terminal transposases alignment in Augé-Gouillou et al. (2001). Amino acid sequences were aligned using the MAFFT program and visualized in the GeneDoc program (Nicholas et al., 1997).

Phylogenetic relationships between *irritans*-like transposable elements identified in *M. destructor* and *M. hordei* were inferred using reference elements belonging to the four *irritans* lineages as well as elements belonging to the *mauritiana*, *cecropia*, *mellifera*, *elegans* subfamilies. The tree was constructed using the Maximum Likelihood (ML) method with bootstrap analysis of 1000 replicates using the MEGA 6 program (Tamura et al., 2013). Phylogenetic tree management was carried out using the iTOL v3 program (Letunic & Bork, 2007).

## RESULTS

### In silico identification of *irritans*-like elements

In silico investigation of the *M. destructor* genome resulted in the identification of 25 *irritans*-like elements with sizes ranging from 474 bp to 1590 bp. These elements named Md1 to Md25 were mapped to 22 different scaffolds, among which three were found to contain 2 *irritans* copies (Table 2).

Most of the copies were defective due to mutations occurring in all parts of the elements and exhibited ORFs encoding truncated transposases lacking or containing some modified signature motifs (Table S2). Only 17 sequences were flanked by a TA dinucleotide target site duplication (TSD) on one or both sides. Among the 25 *irritans* elements, two full length copies (Md14 and Md24) exhibited perfect or near perfect TIRs flanked by the TA dinucleotide TSD and an ORF encoding a transposase. The Md24 transposase is inactive due to two frameshift mutations, while the Md14 transposase bears only a transversion in the start

**Table 2.** Features of the 25 in silico *irritans*-like elements identified in *Mayetiola destructor*.

| *Irritans* copy names | Contig accession numbers | Contig names | Contig length | Position | *Irritans* copy length | Chromosome/scaffold | ITR size | ITR sequences (5' → 3') | TSD | Consensus |
|---|---|---|---|---|---|---|---|---|---|---|
| Md1 | AEGA01000423 | Contig 423 | 23869 bp | 13500–14531 | 1032 bp | chromosome X1 unlocalized genomic scaffold X1.7 | 5':28 bp<br>3':36 bp | 5':CTCGGCGGTTCATTATATATCCCCGGAA<br>3':CTTATTGTTCCATT**CTAGTTCA**ATATATTCCCGGAA | 5'TA<br>3'AA | *Maymarcons1* |
| Md2 | AEGA01009687 | Contig 9693 | 14450 bp | 13812–14436 | 625 bp | chromosome X1 unlocalized genomic scaffold X1Random.5 | 3':30 bp | 3':TTCAGCGGTTCAATATATTCCCGGAATGAG | 3':TA | |
| Md3 | AEGA01017243 | Contig 17256 | 5230 bp | 1–608 | 608 bp | Contig17256 | 5':26 bp | 5':CTCGACGGTTCAATATATTCCCAGAA | 5':TT | |
| Md4 | AEGA01033719 | Contig 33750 | 3676 bp | 1597–2070 | 474 bp | unplaced genomic scaffold Un.22201 | 3':32 bp | 3':ACCAATGAGACCAATTTAGTCTCATTAGTCAT | 3':TC | |
| Md5 | AEGA01028651 | Contig 28675 | 10119 bp | 5717–6751 | 1035 bp | chromosome X2 unlocalized genomic scaffold X2.10 | 5':30 bp | 5':CAAATTTATTTTGGAGCGATCTATTTAAAA | 5':AA | *Maymarcons2* |
| Md6 | AEGA01026913 | Contig 26932 | 8787 bp | 6123–7686 | 1564 bp* (+151ins) | chromosome A1 unlocalized genomic scaffold A1.27 | 5':32 bp<br>3':31 bp | 5':CGCCACTGATTCTGTGCAAATGTCATTTAATT<br>3':CCGAACTTTTTAAAAGTAAAACAAAGTATCT | 5':TA<br>3':TTA | |
| Md7 | AEGA01030236 | Contig 30262 | 3619 bp | 1600–3190 | 1590 bp* (+276ins) | unplaced genomic scaffold Un.20086 | 5':32 bp<br>3':31 bp | 5':GCGTATTTATTACGCAAAAATGCCATTTGCAT<br>3':CCAAATTATTTAATTTAAATGCCAATTCTTT | 5':TA<br>3':CA | |
| Md8 | AEGA01022147 | Contig 22165 | 15673 bp | 14539–15536 | 998 bp | unplaced genomic scaffold Un.16291 | 5':35 bp | 5':TCTGGTGGTGCGGATA**TACGTTCG**AGAGCCCGATC | 5':CA | |
| Md9 | AEGA01021976 | Contig 21994 | 3551 bp | 1915–2921 | 1007 bp | Contig21994 | 5':29 bp<br>3':29 bp | 5':CACAGTCTGTTGTAAAAATTCCCGAACTT<br>3':TACAGTCTGTCGCAAAAGTTCCCGAACTT | 5':TA<br>3':TA | |
| Md10 | AEGA01024833 | Contig 24851 | 24102 bp | 1–1109 | 1109 bp | Unplaced genomic scaffold Un.17377 | 3':30 bp | 3':CACCCTCCACCCCAAAAATAACCGAACTT | 3':TA | |
| Md11 | AEGA01019908 | Contig 19925 | 3325 bp | 1958–2677 | 720 bp | Contig 19925 | 3':30 bp | 3':ATAATACTACTATGAACAAAAATAAGGGGA | 3':TA | *Maymarcons3* |
| | | Contig 19925 anti-sens | | 2646–2886 | 241 bp | | 3':31 bp | 3':AAAATACTACTATGATCAAAAAATAAGGGGA | 3':GA | |
| Md12 | AEGA01027503 | Contig 27522 | 39061 bp | 31190–31900 | 711 bp | chromosome X1 unlocalized genomic scaffold X1Random.8 | 3':30 bp | 3':ACATACTACTATGATCAAAAAATAAGGGGA | 3':CA | |
| | | Contig 27522 anti-sens | | 31869–32120 | 252 bp | | 3':31 bp | 3':TACATACTACTATGATCAAAAAATAAGGGGA | 3':TG | |
| Md13 | AEGA01016065 | Contig 16078 | 4463 bp | 3114–3824 | 711 bp | unplaced genomic scaffold Un.11547 | 3':31 bp | 3':CAACTGTCCAATTTTCAGTTACATTTTGGCA | 3':AT | |
| | | Contig 16078 anti-sens | | 3795–4083 | 289 bp | | 3':31bp | 3':GAATTGTGAAAGTACAAGATATTGGTTCTCA | 3':AA | |
| Md14 | AEGA01027044 | Contig 27063 | 4903 bp | 2864–4188 | 1325 bp | unplaced genomic scaffold Un.18414 | 5':32 bp<br>3':32 bp | **5':ACATACTACTGTGTCCAAATATGAGTAAGACT**<br>**3':ACATACTACTGTGTCCAAATATGAGTAAGACT** | 5':TC<br>3':TC | *Maymarcons4* |
| Md15 | AEGA01014101 | Contig 14112 | 1784 bp | 761–1621 | 861 bp | unplaced genomic scaffold Un.16484 | 5':31 bp<br>3':31 bp | **5':CTACTGTGTCCAAATATGAGTAAGACTTT**<br>**3':CTACTGTGTCCAAATATGAGTAAGACTTT** | 5':TA<br>3':TA | |
| Md16 | AEGA01022488 | Contig 22506 | 14651 bp | 2787–3683 | 897 bp | unplaced genomic scaffold Un.16430 | 5':31 bp<br>3':31bp | 5':CTACTGTGTCCAAAAAATAGTAAGACTTTTT<br>3':CTACTGTGTCCAAAAAATAGTAAGACTTTGT | 5':TA<br>3':TA | |
| Md17 | AEGA01036148 | Contig 36180 | 1729 bp | 358–1267 | 910 bp | unplaced genomic scaffold Un.17398 | 5':31 bp<br>3':31 bp | 5':CTACTGTGGGCAAAAAATAGTAAAACTTTTT<br>3':CTACTGTGGGCAAAAAATAGTAAGACTTTGT | 5':TA<br>3':TA | *Maymarcons5* |
| Md18 | AEGA01024909 | Contig 24927 | 19334 bp | 12103–13016 | 914 bp | unplaced genomic scaffold Un.17398 | 5':31 bp<br>3':31 bp | **5':TACTACTGTGATCAAATTGAAAGGTGAATTT**<br>**3':TACTACTGTGATCAAATTGAAAGGTGAATTT** | 5':TA<br>3':TA | |
| Md19 | AEGA01024835 | Contig 24853 | 14841 bp | 10330–11230 | 901 bp | unplaced genomic scaffold Un.17377 | 5':24 bp<br>3':24 bp | **5':GTGATCAAATTGAAAGGTGAATTT**<br>**3':GTGATCAAATTGAAAGGTGAATTT** | 5':TA<br>3':AG TA | |
| Md20 | AEGA01006984 | Contig 6989 | 1189 bp | 1–738 | 738 bp | Contig 6989 | 3':30 bp | 3':ATACTGCATTTTCATAAATTTTGTCGAAAT | 3':TA | |
| Md21 | AEGA01002443 | Contig 2444 | 1025 bp | 243–894 | 652 bp | unplaced genomic scaffold Un.1210 | 5':31 bp<br>3':31 bp | 5':CTACTGTGATCAAATTGAAAGGTGAATTTTT<br>3':TTACTGTGATCAAATTGAAAGGTGAATTTGT | 5':TA<br>3':TA | |
| Md22 | AEGA01022322 | Contig 22340 | 3738 bp | 202–1302 | 1101 bp | Contig 22340 | 5':29 bp<br>3':30 bp | 5':TGAGTTTAATCATATTGAAATAAATATTT<br>3':CTACTGTGATCAAATTGAAAGGTGAATTTG | 5'AA<br>3'TA | |
| Md23 | AEGA01027488 | Contig 27507 | 11595 bp | 3994–5580 | 1587 bp* (+ 437 bp ins) | chromosome X1 unlocalized genomic scaffold X1Random.8 | 5':38 bp<br>3':39 bp | 5':CTGTGATCAATTTCCGGTAGGATTTTCAAATTTAAAAT<br>3':CTCTGTGATTAATATCCGGAAGGACCTCATTAAAAAAAT | 5':TA<br>3':TA | |
| Md24 | AEGA01006767 | Contig 6772 | 4506 bp | 646–1926 | 1281 bp | chromosome A1 unlocalized genomic scaffold A1.28 | 5':39 bp<br>3':39 bp | 5':CTACTGTGATCAAATATAACCTGGAATTTGCAATTTAAA<br>3':CTACTGTGATCAAATATAACCCGGAATTTGCTTATAAAA | 5':TA<br>3':TA | – |
| Md25 | AEGA01033657 | Contig 33688 | 7130 bp | 4256–5570 | 1315 bp | unplaced genomic scaffold Un.22167 | 5':24 bp<br>3':22 bp | 5':ACTACATGAACAGAAAATATTCTC<br>3':AGTATGAACAACATATCTCCCC | 5':TC<br>3':TA | – |

Perfect TIRs are in bold and insertions are underlined. Copies including insertions in their ORFs are labeled with an asterix (*). Full length copies are marked in red.

codon (ATG → ATT) changing methionine into isoleucine, resulting in its dysfunctionality.

The searches in GenBank database using the BLASTX algorithm revealed that identified elements shared the best amino acid homologies, ranging from 51% to 85%, with *irritans*-like elements from the tephritid fruit fly *Bactrocera tryoni* (APL98287.1) and the green lacewing *Chrysoperla plorabunda Cpmar1* (AAC46945.1).

Nucleotide sequence alignment of the identified *irritans*-like elements allowed the establishment of a similarity matrix from which five consensuses designated by *Maymarcons1–Maymarcons5* were constructed. The comparison of these consensuses with the active *Himar1* element of the horn fly *H. irritans* (U11642) revealed a nucleic acid similarity ranging from 47% (*Maymarcons3*) to 58% (*Maymarcons1*). A diagram of the consensuses of the *irritans*-like nucleic acid sequences and their conceptual transposases is shown in Fig. 1a.

TIRs alignment of elements belonging to each consensus revealed different mirror and palindromic motifs centered at positions ranging from 12 to 19. The 5'YYAGRT3' motif cleavage signal sequence described by Bigot et al. (2005) was altered by frameshift mutations in *Maymarcons1* and missing in the remaining consensuses. However,

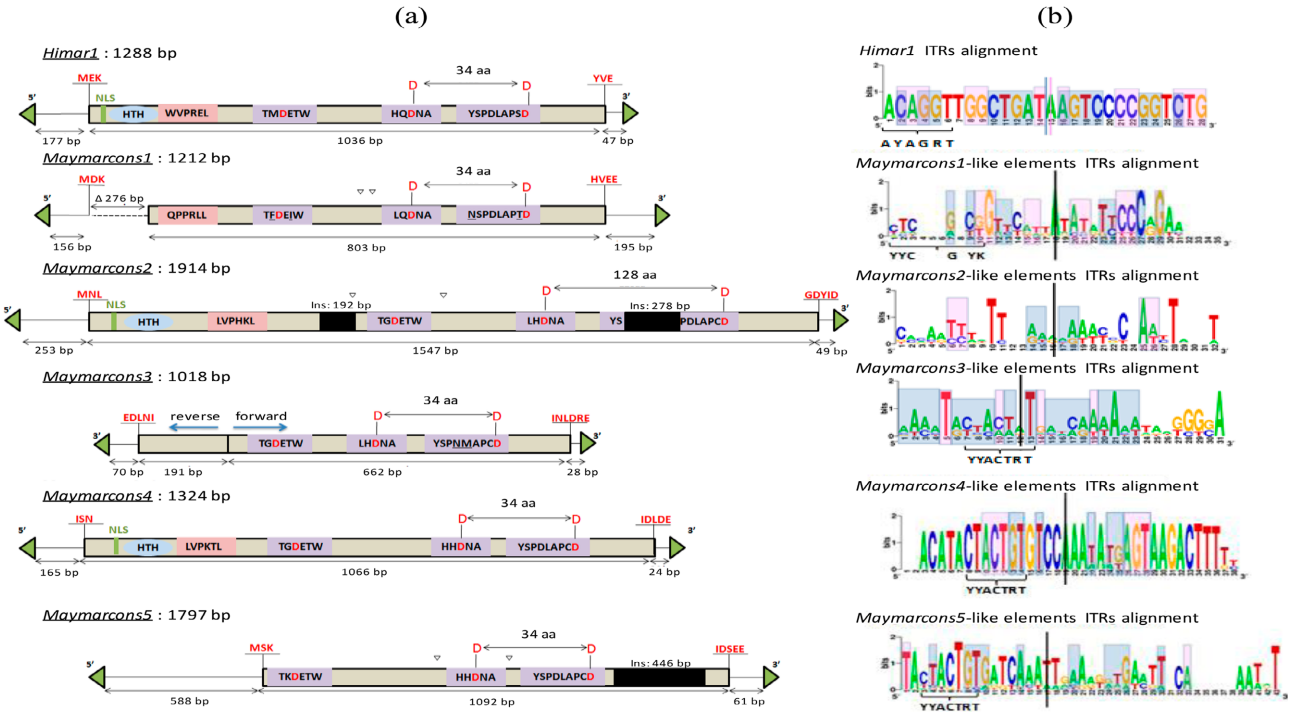


**Fig. 1.** Diagram of the five *irritans*-like consensuses and the logo of their corresponding ITRs. (a) The five consensuses are compared to the full length *irritans* element *Himar1* (U11642) as a reference. ITRs are indicated by green triangles and UTRs by a continuous black line. HTH and NLS motifs are indicated, respectively, by a blue circle and green rectangle. Motifs of the catalytic triad are boxed in purple rectangles and modified residues are underlined. The aspartate residues are marked in red (with red capital D). The WVPHEL signature motif is indicated by a pink rectangle. The first start residues and last residues are indicated in red. Deletions are represented by dashed lines, whereas insertions (Ins) are indicated by black rectangles. Frameshifts are indicated by empty upside-down triangles. (b) Weblogo representing the ITRs of the five *irritans* groups identified (*Maymarcons*-like elements) compared to ITRs of *Himar1*. The vertical axis is in bits with a maximum of two bits, which is proportional to the nucleotide level conservation at each position. Palindromic and mirror motifs are shown in pink and blue rectangles, respectively. Vertical black lines correspond to symmetry axes. In the *Himar1* logo, pink and blue axes correspond to the symmetry of palindromic and mirror motifs, respectively.

a conserved motif 5'CTACTRT3' was detected in *Maymarcons3*, *Maymarcons4* and *Maymarcons5* at positions 7–13; 8–14 and 3–9, respectively (Fig. 1b).

For comparative purposes, the transcriptome shotgun assembly (TSA) of the orange wheat blossom midge *S. mosellana* available in GenBank was investigated using as queries the five built consensuses of *Maymarcons*. Four cDNA contigs similar to *Maymarcons4* were identified with a 79% to 82% nucleotide identity, among which there were two with full length copies of *irritans*-like elements with an intact ORF. This suggests that these elements are potentially active in the orange cecidomyiid fly whereas the two others that correspond to incomplete *irritans* elements have internal deletions spanning the first two motifs of the catalytic triad and occur at positions 684–1071 bp (Fig. 2).

### In vitro identification of *irritans* like transposons in *M. destructor* and *M. hordei*

To validate the presence of *irritans*-like elements in the two species of *Mayetiola* studied, five primers were designed from the TIRs sequence logos. Results indicate that for *M. destructor*, PCR products were obtained with primers designed from *Maymarcons4* and *Maymarcons5*, whereas the *M. hordei* amplifications were obtained only with a primer specific to *Maymarcons4*. Cloning and sequencing of these products allowed the identification of 17 *irritans*-like elements ranging from 802 bp to 929 bp.

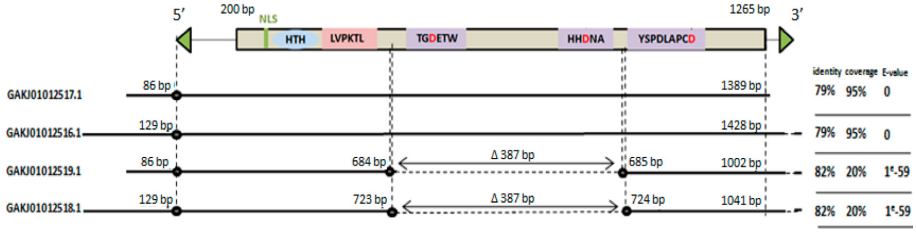


**Fig. 2.** Comparison of the *Maymarcons4* consensus and the four cDNA sequences detected in *Sitodiplosis mosellana*. Deletions are indicated by dashed lines. Accession numbers of cDNA sequences are shown on the left and their identity statistics with *Maymarcons4* are shown on the right.

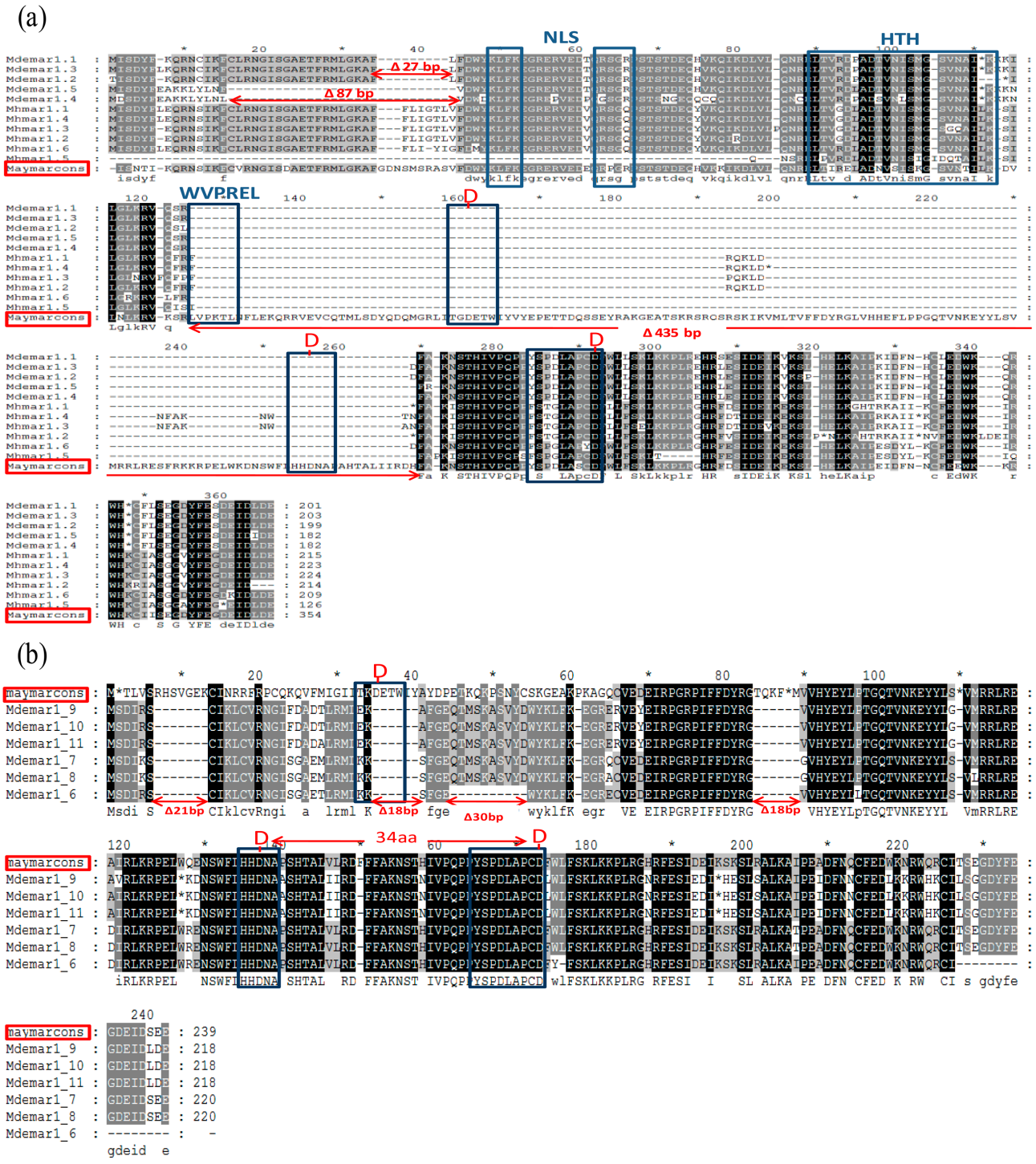


**Fig. 3.** Alignment of the conceptual translation of *Mdemar1* and *Mhmar1* elements with (a) the putative transposase of *Maymarcons4* consensus (b) the putative transposase of *Maymarcons5* consensus. Black and grey blocks correspond to identical and homologous regions. Deletions are represented by discontinuous lines and marked by double pointed arrows. Asterix correspond to stop codons. Binding regions, catalaytic triad domains and signature motifs are boxed in blue.

Elements from *M. destructor* similar to *Maymarcons4*-like elements were named *Mdemar1.1* to *Mdemar1.5* and those similar to *Maymarcons5*-like elements were named *Mdemar1.6* to *Mdemar1.11*. The *Maymarcons4*-like elements of *M. hordei* were named *Mhmar1.1* to *Mhmar1.6*. All the sequences were deposited in the DNA Data Bank of Japan (DDBJ: http://www.ddbj.nig.ac.jp/) under accession numbers: LC218006– LC218022.

Alignment of the 11 *Maymarcons4*-like elements obtained from *M. destructor* and *M. hordei* showed nucleotide similarities with *Maymarcons4* ranging from 87.03% to 90.27% and 80.66% to 86.02%, respectively. Furthermore, the conceptual translation of *Mayamarcons4*-like elements was performed and aligned with the putative transposase of *Maymarcons4*. As shown in Fig. 3a, all elements have a deletion spanning the first signature motif WVPREL and the

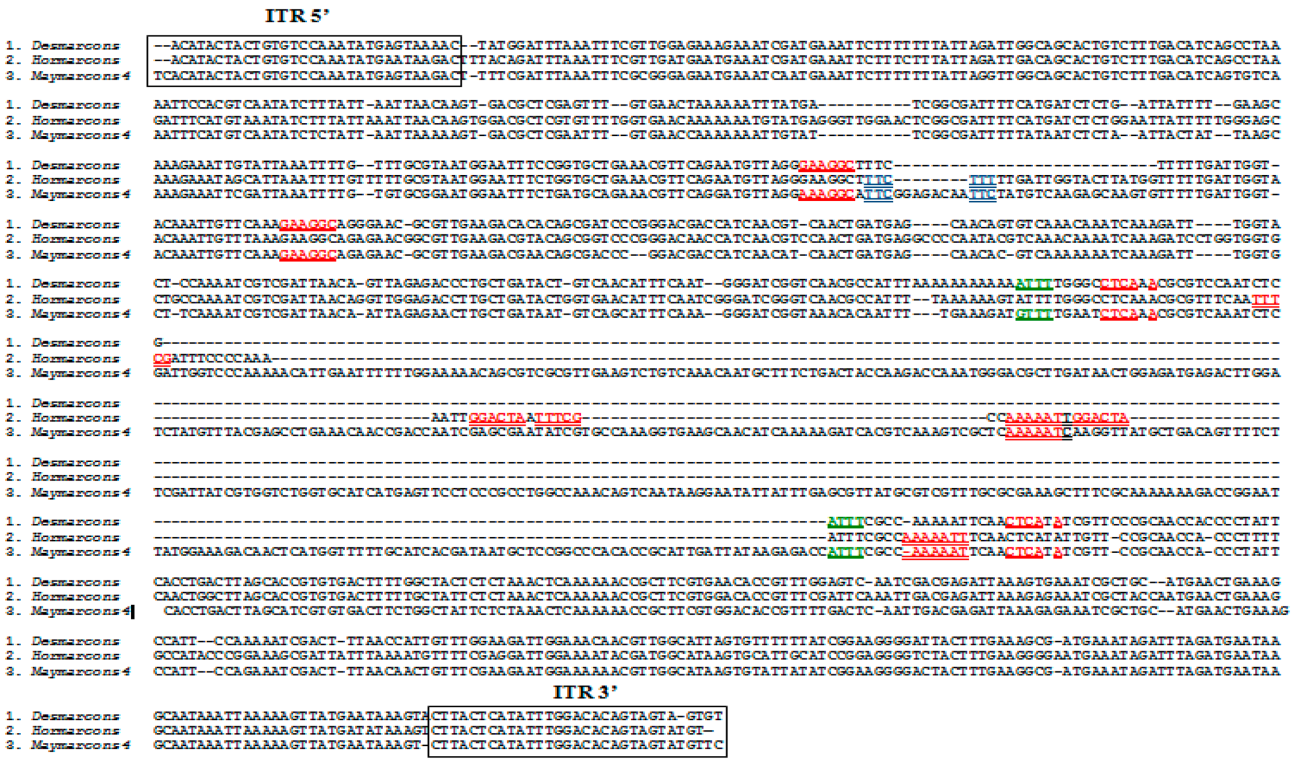


**Fig. 4.** Nucleic acidalignment of *Maymarcons4* with *Desmarcons* and *Hormarcons* generated from in vitro elements in *M. destructor* and *M. hordei*, respectively. Short direct repeats (SDRs) microhomologies, flanking deletions and Breaking Points (BPs) are in bold and underlined by a single or a double line in *Desmarcons* and *Hormarcons*, respectively. Microhomologies localized near the BPs (BPNN i.e. breaking point near near) are in red, microhomologies exact near the BPs (BPNE i.e. breaking point near exact) are green and microhomologies localized exactly at BPs (BPEE i.e. breaking point exact exact) are blue. Boxed regions correspond to 5' and 3' TIRs of the 3 consensuses sequences.

two first aspartic residues (D) of the catalytic triad DD(34) D. The third aspartic residue motif (YSPDLAPCD) is conserved in *M. destructor*, whereas in *M. hordei* it is replaced by the FS(T/P)GPLACD motif.

Moreover, comparison of the six *Maymarcons5*-like elements identified in *M. destructor* revealed 84.26% to 96.93% nucleic acid similarity with the consensus of *Maymarcons5*. The 5' and 3' TIRs of these elements differ in their inner region while the 5'CTACTRT3' motif is conserved in its outer region. The alignment of the putative transposases of these elements with the conceptual transposase of *Maymarcons5* revealed a deletion spanning the first motif of the catalytic core as shown in Fig. 3b.

Noteworthy, the nucleic acid alignments of several *Maymarcons4*-like elements identified in the two *Mayetiola* species and *Sitodiplosis* transcripts, revealed deletions spanning the same positions. Given that, microhomology analyses have been performed to verify whether these deletions are random or not, two consensuses were established from the identified elements and designated *Desmarcons* and *Hormarcons* for *M. destructor* and *M. hordei*, respectively. The alignment of both consensuses with *Maymarcons4* sequence revealed a total of six deletions sized from 8 bp to 431 bp (Fig. 4). The *Desmarcons* has a deletion of 28 bp flanked by a short direct repeat (SDRs), which occurs near both BPs (Breaking Points Near Near, BPNN) and a 431 bp deletion bordered by SDRs, which are exactly at the BP on one side and near the BP on the other side (BPs Exact Near, BPEN). The *Hormarcons* has four deletions of 8 bp, 137 bp, 43 bp and 136 bp flanked by SDRs localized exactly at BPs on both sides (BPs Exact Exact, BPEE) and/or BPNN microhomologies.

The phylogenetic tree (Fig. 5) indicates two major groups; the first belongs to the *Himar1*-like lineage and contains the *Maymarcons1* consensus, while the second diverges from the four known *irritans* lineages and is divided into three subgroups, one corresponding to the *Maymarcons2* consensus, one to the *Maymarcons3* consensus and a third that includes *Maymarcons4* and *Maymarcons5*-like elements. In the latter subgroup, *Maymarcons4*-like elements of *M. destructor* diverge from those of *M. hordei*.

## DISCUSSION

In the *M. destructor* genome, two MLEs named *Desmar1* and *Des2* were described (Shukle & Russell, 1995; Russell & Shukle, 1997). The *Desmar1* is a full length *mauritiana*-like element with an intact ORF and perfect TIRs, while *Des2* is an internal region belonging to *irritans*-like elements. To date, no complete *irritans* copies have been identified.

In the current study, complete copies (from TIR to TIR) of *irritans*-like transposable elements were identified and characterized for the first time in *M. destructor* and *M. hordei* using a combination of in silico and in vitro approaches. In silico analysis of the *M. destructor* genome revealed 25 *irritans*-like elements from which five con-

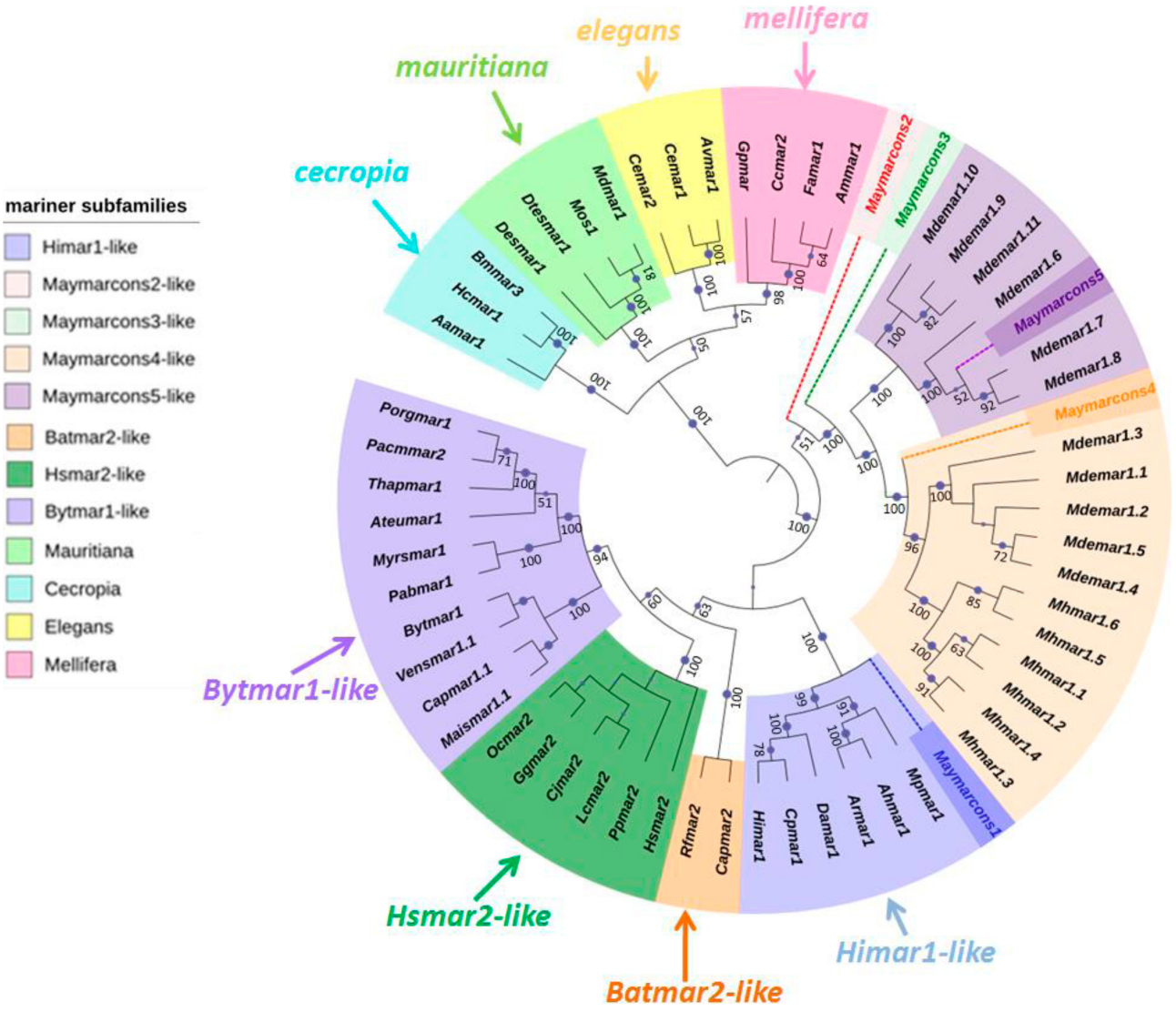


**Fig. 5.** Phylogenetic relationships based on the nucleic acid sequences of *Maymarcons* consensuses, in vitro *irritans* elements of *M. destructor* and *M. hordei*, and the other subfamilies of mariner elements. The tree was inferred using the maximum likelihood method with a bootstrap of 1000 replicates. The sizes of the blue circles depend on bootstrap values. Values less than 50% are hidden.The reference elements from the five mariner subfamilies were downloaded from Genbank and the accession numbers are: *Portunus granulates Porgmar1* (AM906133.1), *Pachygrapsus marmoratus Pacmmar2* (AM231072.1), *Thalamita poissoni Thapmar1* (AM906153.1), *Atelecyclus undecimdentatus Ateumar1* (AM906094.1), *Myra subgranulata Myrsmar1* (AM906111.1), *Paromola bathyalis Pabmar1* (AM906119.1), *Bythograea thermydron Bytmar1* (AJ507219.1), *Ventiella sulfuris Vensmar1.1* (AJ507232.1), *Cancer pagurus Capmar1.1* (AJ507245.1), *Maia squinado Maismar1.1* (AJ507238.1) from *Bytmar1*-like *irritans* lineage, *Oryctolagus cuniculus Ocmar2* (AC147588.2), *Gorilla gorilla Ggmar2* (AC145402.3), *Callithrix jacchus Cjmar2* (AC191240.1), *Lemur catta Lcmar2* (AC133072.1), *Pongo pygmaeus Ppmar2* (DQ480417.1), Human mariner2 *Hsmar2* (U49974.1) from *Hsmar2*-like *irritans* lineage, *Rhinolophus ferrumequinum Rfmar2* (AC163264.3) and *Carollia perspicillata Capmar2* (AC148202.3) from *Batmar2*-like *irritans* lineage, *Haematobia irritans Himar1* (U11642.1), *Chrysoperla plorabunda Cpmar1* (U11650.1), *Drosophila ananassae Damar1* (U11646.1), *Ascogaster reticulates Armar1* (AB020618.1), *Adoxophyes honmai Ahmar1* (AB020617.1), *Mantispa pulchella Mpmar1* (U11649.1) from *Himar1*-like *irritans* lineage, *Apis mellifera Ammar1* (AY155490), *Forficula auricularia Famar1* (AY155492.1), *Ceratitis capitata Ccmar2* (AY155493), *Glossina palpalis Gpmar* (U18308.1) from *Mellifera* subfamily, *Caenorhabditis elegans Cemar1* (ZC132.1), *Cemar2* (Y39A3A.1), *Adineta vaga Avmar1* (AF014939.1), *Musca domestica Mdmar1* (AF373028.1), *Drosophila mauritiana Mos1* (M14653), *Drosophila teissieri Dtesmar1* (AF035566.1), *Mayetiola destructor Desmar1* (U24436.1), *Bombyx mori Bmmar3* (D88671), *Hyalophora cecropia Hcmar1* (M63844.1), *Attacus atlas* Aamar1 (AB0064).

sensuses were built. This low copy number of elements is congruent with previous studies made by Shukle & Russell (1995).

This study revealed that most of the *irritans*-like copies were defective and damaged, due to a frameshift, nonsense or indel mutations spanning all the parts of the elements, which indicate an ancient invasion of the genome by these elements, which might be in the senescence stage (Kidwell & Lisch, 2001).

Likewise, the deletions occur mainly in the N-terminal region and the first domain of the catalytic triad, which are crucial for an efficient MLE mobilization (Lohe & Hartl, 2002). Thus, these deleted elements could act as inhibitors of trans-mobilization by the full-length copies as described for *Botmar1*-like copies (Rouleux-Bonnin et al., 2005) or as repressors like the *KP* deleted form, reported in the *P* element (Black et al., 1987; Andrews & Gloor, 1995).

Furthermore, analysis of the *M. destructor* genome revealed chimerical elements with 3'–3'extremities that might be generated by either an ectopic recombination replacing 5' extremity by 3' extremity or an internal deletion of an initial head-to-tail *mariner* close copies as proposed by Filée et al. (2015).

The analysis of TIRs revealed specific conserved motifs that are different from those described by Bigot et al. (2005) suggesting specific interactions between these elements and their protein products. It is noteworthy that such conserved motif modifications were previously reported in the two *irritans* elements, *Bytmar1* and *Hsmar2* (Bigot et al., 2005). These observations provide evidence of high diversity in the *irritans* TIRs compared to those of other *mariner* subfamilies.

The molecular analysis revealed *Maymarcons4*-like elements in both species, whereas *Maymarcons5*-like elements were detected only in *M. hordei*. This could be explained by these elements invading the *M. destructor* genome following speciation. Conversely, the non amplification of other *irritans*-like elements detected in the in silico investigation could be related to the high nucleotide variability of *mariner* TIRs (Bigot et al., 2005) or to the non occurrence of these elements in the Tunisian strains analyzed. Another explanation could be that an eventual ancient invasion of some of these elements (*Maymarcons1* and *Maymarcons2* like elements) led to the accumulation of mutations in their whole sequences, including ITRs. This would be due to the independent evolution of these copies.

The occurrence of *Maymarcons4*-like elements in two species of *Mayetiola* and even in the TSA of the orange blossom midge *S. mosellana* indicate an ancient invasion of these *irritans* elements in a common ancestral species of cecidomyiid, which would have been followed by a vertical transmission into derived species, in which it took the form of independently-differentiated, heterologous elements, as is hypothesized for the YSPDLAPCD motif in *M. hordei.* Likewise, it is also likely that a horizontal transfer between *M. destructor* and *S. mosellana* occurred, since they share the same host plant and have full length copies of *irritans* elements in their genomes.

Strikingly, the deleted regions in the defective forms of *Maymarcons4*-like elements in *M. destructor* and *M. hordei* are the same, suggesting a possible occurrence of these deletions in the ancestor of the two species. Moreover, these gaps are flanked by microhomologies. The association of microhomology with deletion breakpoints is reported in *Mos1* (Brunet et al., 2002), *mauritiana* (Kharrat et al., 2015) and *irritans* elements (Ben Lazhar-Ajroud et al., 2016). These deletions do not occur randomly and could result from a host genome control, as well as from additional mechanisms, such as abortive gap repair (Rubin & Levy, 1997) and/or ectopic recombination between homologous short sequences leading to different deletion forms (Negoua et al., 2013; Kharrat et al., 2015).

The phylogenetic analysis grouped the *Maymarcons1* consensus within *Himar1*-like lineage and revealed a novel *irritans* group, distinct from the four *irritans* lineages previously reported by Sinzelle et al. (2006). Thus, we recommend that the original classification should be broadened to include the *irritans* elements characterized in this study, as well as the two *irritans* elements *Tvmar1* (Claudianos et al., 2002) and *Pacmmar1* (Bui et al., 2007), which also differ from the four known *irritans* lineages.

Moreover, the phylogenetic tree revealed a divergence in the *Maymarcons4*-like elements with respect to *Mayetiola* species, which favours an independent evolution of these elements after speciation and supports the vertical transfer from an ancestral species.

The high diversity recorded in *M. destructor* suggests that its genome was invaded many times by different types of *irritans* elements, as reported in species of *Drosophila* by Wallau et al. (2014).

In conclusion, the combined results of the in silico and in vitro analyses give an outline of the evolutionary dynamics of the *irritans*-like elements in the genomes of the two species of *Mayetiola*. The knowledge of the TE content might be helpful to explore the genome for a better understanding the seeking behavior of these insects with their host and in the case of transposon-based biological pest management for a better vector choice.

**ACKNOWLEDGEMENTS.** The authors are grateful for the funding support of the Tunisian Ministry of Higher Education and Scientific Research.

## REFERENCES

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. 1990: Basic local alignment search tool. — *J. Mol. Biol.* **215**: 403–410.

Andrews J.D. & Gloor G.B. 1995: A role for the *KP* leucine zipper in regulating *P* element transposition in *Drosophila melanogaster*. — *Genetics* **141**: 587–594.

Arensburger P., Piégu B. & Bigot Y. 2016: The future of transposable element annotation and their classification in the light of functional genomics – what we can learn from the fables of Jean de la Fontaine? — *Mob. Genet. Elem.* **6**(6): e1256852.

Ashburner M., Hoy M.A. & Peloquin J.J. 1998: Prospects for the genetic transformation of arthropods. — *Insect Mol. Biol.* **7**: 201–213.

Augé-Gouillou C., Hamelin M.H., Demattei M.V., Periquet G. & Bigot Y. 2001: The ITR binding domain of the *mariner Mos1*-transposase. — *Mol. Genet. Genom.* **265**: 58–65.

Barry E.G., Witherspoon D.J. & Lampe D.J. 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.

Behura S.K., Shukle R.H. & Stuart J.J. 2010: Assessment of structural variation and molecular mapping of insertion sites of *Desmar*-like elements in the Hessian fly genome. — *Insect Mol. Biol.* **19**: 707–715.

Ben Lazhar-Ajroud W., Caruso A., Mezghani M., Bouallegue M., Tastard E., Denis F., Rouault J.D., Makni H., Capy P., Chenais B., Makni M. & Casse N. 2016: Characterization of irritans *mariner*-like elements in the olive fruit fly *Bactrocera oleae* (Diptera: Tephritidae): evolutionary implications. — *Naturwissenschaften* **103**: 64, 14 pp.

Bigot Y., Brillet B. & Auge-Gouillou C. 2005: Conservation of palindromic and mirror motifs within inverted terminal repeats of *mariner*-like elements. — *J. Mol. Biol*. **351**: 108–116.

Bire S. & Rouleux-Bonnin F. 2012: Transposable elements as tools for reshaping the genome: it is a huge world after all! — *Meth. Mol. Biol*. **859**: 1–28.

Black D.M., Jackson M.S., Kidwell M.G. & Dover G.A. 1987: KP elements repress P-induced hybrid dysgenesis in *Drosophila melanogaster*. — *EMBO J*. **6**: 4125–4135.

Brillet B., Bigot Y. & Auge-Gouillou C. 2007: Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. — *Genetica* **130**: 105–120.

Brunet F., Giraud T., Godin F. & Capy P. 2002: Do deletions of *Mos1*-like elements occur randomly in the Drosophilidae family? — *J. Mol. Evol.* **54**: 227–234.

Bucher E., Reinders J. & Mirouze M. 2012: Epigenetic control of transposon transcription and mobility in *Arabidopsis*. — *Curr. Opin. Plant Biol.* **15**: 503–510.

Bui Q.T., Delauriere L., Casse N., Nicolas V., Laulier M. & Chenais B. 2007: Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*. — *Gene* **396**: 248–256.

Chenais 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.

Claudianos C., Brownlie J., Russell R., Oakeshott J. & Whyard S. 2002: *maT* – a clade of transposons intermediate between *mariner* and *Tc1*. — *Mol. Biol. Evol.* **19**: 2101–2109.

Combet C., Blanchet C., Geourjon C. & Deleage G. 2000: NPS@: Network protein sequence analysis. — *Trends Biochem. Sci.* **25**: 147–150.

Feschotte C. & Pritham E.J. 2007: DNA transposons and the evolution of eukaryotic genomes. — *Annu. Rev. Genet*. **41**: 331–368.

Filée J., Rouault J.D., Harry M. & Hua-Van A. 2015: *Mariner* transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. — *BMC Genom.* **16**: 1061, 17 pp.

Finnegan D.J. 1989: Eukaryotic transposable elements and genome evolution. — *Trends Genet.* **5**: 103–107.

Halaimia-Toumi N., Casse N., Demattei M.V., Renault S., Pradier E., Bigot Y. & Laulier M. 2004: The GC-rich transposon *Bytmar1* from the deep-sea hydrothermal crab, *Bythograea thermydron*, may encode three transposase isoforms from a single ORF. — *J. Mol. Evol.* **59**: 747–760.

Handler A.M. & O'Brochta D.A. 2012: *Transposable Elements for Insect Transformation.* Acadamic Press, London, pp. 90–133.

Hartl D.L., Lohe A.R. & Lozovskaya E.R. 1997: Modern thoughts on an ancyent *marinere*: function, evolution, regulation. — *Annu. Rev. Genet*. **31**: 337–358.

Hirsch C.D. & Springer N.M. 2017: Transposable element influences on gene expression in plants. — *Biochim. Biophys. Acta* **1860**: 157–165.

Jacobson J.W., Medhora M.M. & Hartl D.L. 1986: Molecular structure of a somatically unstable transposable element in *Drosophila*. — *Proc. Natl. Acad. Sci. USA* **83**: 8684–8688.

Katoh K. & Standley D.M. 2013: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. — *Mol. Biol. Evol*. **30**: 772–780.

Kharrat I., Mezghani M., Casse N., Denis F., Caruso A., Makni H., Capy P., Rouault J.D., Chenais B. & Makni M. 2015: Characterization of *mariner*-like transposons of the *mauritiana* subfamily in seven tree aphid species. — *Genetica* **143**: 63–72.

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

Lampe D.J., Churchill M.E. & Robertson H.M. 1996: A purified *mariner* transposase is sufficient to mediate transposition in vitro. — *EMBO J.* **15**: 5470–5479.

Lampe D.J., Witherspoon D.J., Soto-Adames F.N. & Robertson H.M. 2003: Recent horizontal transfer of *mellifera* subfamily *mariner* transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. — *Mol. Biol. Evol*. **20**: 554–562.

Largaespada D.A. 2003: Generating and manipulating transgenic animals using transposable elements. — *Reprod. Biol. Endocrinol.* **1**: 80, 10 pp.

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

Letunic I. & Bork P. 2007: Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. — *Bioinformatics* **23**: 127–128.

Lin J.R. & Hu J. 2013: SeqNLS: nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. — *PLoS ONE* **8**(10): e76864, 12 pp.

Lohe A.R. & Hartl D.L. 2002: Efficient mobilization of *mariner* in vivo requires multiple internal sequences. — *Genetics* **160**: 519–526.

Lohe A.R., Moriyama E.N., Lidholm D.A. & Hartl D.L. 1995: Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. — *Mol. Biol. Evol.* **12**: 62–72.

Mezghani Khemakhem M., Makni H. & Marrakchi M. 2002: Identification par PCR-RFLP de marqueurs mitochondriaux chez deux espèces de *Mayetiola* nuisibles aux cultures de céréales (Diptera: Cecidomyiidae). — *Ann. Soc. Entomol. Fr.* **38**: 277–282.

Monacell J.T. & Carbone I. 2014: Mobyle SNAP Workbench: a web-based analysis portal for population genetics and evolutionary genomics. — *Bioinformatics* **30**: 1488–1490.

Munoz-Lopez M. & Garcia-Perez J.L. 2010: DNA transposons: nature and applications in genomics. — *Curr. Genom.* **11**: 115–128.

Munoz-Lopez M., Siddique A., Bischerour J., Lorite P., Chalmers R. & Palomeque T. 2008: Transposition of *Mboumar-9*: identification of a new naturally active *mariner*-family transposon. — *J. Mol. Biol.* **382**: 567–572.

Negoua A., Rouault J.D., Chakir M. & Capy P. 2013: Internal deletions of transposable elements: the case of *Lemi* elements. — *Genetica* **141**: 369–379.

Nicholas K.B., Nicholas H.B.J. & Deerfield D.W. 1997: GeneDoc: analysis and visualization of genetic variation. — *Embnew. News* **4**(1): 14 pp.

Panaud O. 2016: Horizontal transfers of transposable elements in eukaryotes: the flying genes. — *C. R. Biol.* **339**: 296–299.

Peccoud J., Loiseau V., Cordaux R. & Gilbert C. 2017: Massive horizontal transfer of transposable elements in insects. — *Proc. Natl. Acad. Sci. USA* **114**: 4721–4726.

Piégu B., Bire S., Arensburger 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.

Pietrokovski S. & Henikoff S. 1997: A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. — *Mol. Gen. Genet*. **254**: 689–695.

Plasterk R.H. 1996: The *Tc1/mariner* transposon family. — *Curr. Top. Microbiol. Immunol.* **204**: 125–143.

Plasterk R.H., Izsvak Z. & Ivics Z. 1999: Resident aliens: the *Tc1/mariner* superfamily of transposable elements. — *Trends Genet.* **15**: 326–332.

Rigal M. & Mathieu O. 2011: A "mille-feuille" of silencing: Epigenetic control of transposable elements. — *Biochim. Biophys. Acta* **1809**: 452–458.

Robertson H.M. 1993: The *mariner* transposable element is widespread in insects. — *Nature* **362**: 241–245.

Robertson H.M. & Asplund M.L. 1996: *Bmmar1*: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*. — *Insect Biochem. Mol. Biol*. **26**: 945–954.

Robertson H.M. & Lampe D.J. 1995: Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. — *Mol. Biol. Evol.* **12**: 850–862.

Robertson H.M. & MacLeod E.G. 1993: Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. — *Insect Mol. Biol.* **2**: 125–139.

Robertson H.M. & 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.

Rouleux-Bonnin F., Petit A., Demattei M.V. & Bigot Y. 2005: Evolution of full-length and deleted forms of the *mariner*-like element, *Botmar1*, in the genome of the bumble bee, *Bombus terrestris* (Hymenoptera: Apidae). — *J. Mol. Evol.* **60**: 736–747.

Rubin E. & Levy A.A. 1997: Abortive gap repair: underlying mechanism for *Ds* element formation. — *Mol. Cell. Biol.* **17**: 6294–6302.

Russell V.W. & Shukle R.H. 1997: Molecular and cytological analysis of a *mariner* transposon from Hessian fly. — *J. Hered*. **88**: 72–76.

Ryder E. & Russell S. 2003: Transposable elements as tools for genomics and genetics in *Drosophila*. — *Brief. Funct. Genom. Proteom.* **2**: 57–71.

Sambrook J., Fritsch E.F. & Maniatis N. 2011: Screening of bacterial recombinants: Strategies and preventing false positives. In Brown P.G. (ed.): *Molecular Cloning – Selected Applications in Medicine and Biology*. Intech, Rijeka, 324 pp.

Shukle R.H. & Russell V.W. 1995: *Mariner* transposase-like sequences from the Hessian fly, *Mayetiola destructor*. — *J. Hered.* **86**: 364–368.

Sinzelle L., Chesneau A., Bigot Y., Mazabraud A. & Pollet N. 2006: The *mariner* transposons belonging to the *irritans* subfamily were maintained in chordate genomes by vertical transmission. — *J. Mol. Evol.* **62**: 53–65.

Sunnucks P. & Hales D.F. 1996: Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). — *Mol. Biol. Evol.* **13**: 510–524.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. 2013: Mega6: Molecular evolutionary genetics analysis version 6.0. — *Mol. Biol. Evol.* **30**: 2725–2729.

Wallau G.L., Capy P., Loreto O. & Hua-Van A. 2014: Genomic landscape and evolutionary dynamics of *mariner* transposable elements within the *Drosophila* genus. — *BMC Genomics* **15**: 727, 19 pp.

Waterhouse A.M., Procter J.B., Martin D.M., Clamp M. & Barton G.J. 2009: Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. — *Bioinformatics* **25**: 1189–1191.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P. & Schulman A.H. 2007: A unified classification system for eukaryotic transposable elements. — *Nat. Rev. Genet.* **8**: 973–982.

Yuan Y.W. & Wessler S.R. 2011: The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. — *Proc. Natl. Acad. Sci. USA* **108**: 7884–7889.

Received May 8, 2017; revised and accepted September 4, 2017
Published online October 5, 2017

**Table S1.** Reference sequences used as queries to search for *irritans* transposable elements in the genomic scaffolds of *Mayetiola destructor*.

| Query | Species | Subphylum/Order | Nucleotide accession number | Amino acid accession number |
|---|---|---|---|---|
| *Bytmar1* | *Bythograea thermydron* | Crustacea | | CAD45367.1 |
| *Erivmar1* | *Eriphia verrucosa* | Crustacea | | CAP20022.1 |
| *Cpmar1* | *Chrysoperla plorabunda* | Neuroptera | U11650 | AAC46946.1 |
| *Hsmar2* | *Homo sapiens* | Primates | U49974.1 | AAC52011.1 |
| *Apmar1* | *Agrilus planipennis* | Coleoptera | GQ398105.1 | ADB28039.1 |
| *Himar1* | *Haematobia irritans* | Diptera | U11642 | |
| *Ag5* | *Anopheles gambiae* | Diptera | U11658.1 | |
| *Damar1* | *Drosophila ananassae* | Diptera | U11646.1 | |
| *Mpmar1* | *Mantispa pulchella* | Neuroptera | U11649.1 | |
| *Himar1* | *Haematobia irritans* | Diptera | U11642.1 | |
| *Xtmar1* | *Xenopus tropicalis* | Anura | AJ852524.1 | |
| *Diamar19* | *Diachasmimorpha longicaudata* | Hymenoptera | AY601745.1 | |
| *Pfmar3* | *Psyttalia fletcheri* | Hymenoptera | AY601746.1 | |
| *Ahmar1* | *Adoxophyes honmai* | Lepidoptera | AB020617.1 | |

**Table S2.** Features of the 25 transposases conceptually translated from the in silico identified elements in *Mayetiola destructor*.

| Transposase | Length | Start motif | Presence/absence of WVPHEL signature | Stop condons number | Catalytic triad characteristics | | |
|---|---|---|---|---|---|---|---|
| | | | | | TGDETW | HHDNA | YSPDLAPSD |
| Md1 | 257aa | MDK | WLPRLL | 2 | TFDEIW | FLQDNA | YSPDLAPAD |
| Md2 | 185aa | abs | abs | 4 | abs | FLQDNA | SFRVLASSD |
| Md3 | 142aa | MNR | WVRLL | 1 | TIDETW | abs | abs |
| Md4 | 132aa | abs | abs | 4 | abs | PENAP | *SLDVAPSD |
| Md5 | 293aa | MNF | WN | 3 | TGE | LLHDNAP | YSPDLATCD |
| Md6 | 354aa $^{+ins}$ | MNF | LVPHKL | 3 | TGD*TW | ILHYYA | CSPDLAPCD |
| Md7 | 339aa$^{+ins}$ | MNL | LVPHKL | 2 | TGDETW | LLHDSS | YS($^{ins}$)APDLAPCD |
| Md8 | 289aa | MNF | FVPHKL | 4 | *GDETW | RLLHDNA | YSPGLAPCD |
| Md9 | 218aa$^{+ins}$ | MSV | abs | 2 | abs | LLHNNA | YSPDMAPCD |
| Md10 | 334aa | MLG | abs | 4 | TGGETW | LLHDNS | YSPDFAPCD |
| Md11 | 219aa<br>64aa | abs | abs | 3 | TGDETW | ILHHENA | YSPNMAPCD |
| Md12 | 227aa<br>60aa | abs | abs | 2 | TGDETL | ILHHDNA | YSPNMAPCV |
| Md13 | 206aa<br>61aa | abs | abs | 1 | TGDETW | ILHHDNA | YSPNMVPCD |
| Md14 | 354aa | ISN | LVPKTL | 0 | TGDETW | FLHHDNA | YSPDLASCD |
| Md15 | 203aa | MISD | abs | 1 | abs | abs | YSPDLAPCD |
| Md16 | 215aa | MPK | abs | 0 | abs | LHHDNA | YSPDLAPCD |
| Md17 | 220aa | MSDI | abs | 2 | abs | LHHDNA | YSPDLAPCD |
| Md18 | 221aa | MSDI | abs | 0 | abs | FLHHDNA | YSPDLAPCD |
| Md19 | 211aa | MSDI | abs | 1 | abs | FLHHDNA | YSPDLAPCD |
| Md20 | 215aa | abs | abs | 0 | abs | FLHHDNA | YSPDLAPCD |
| Md21 | 153aa | abs | abs | 1 | abs | FSHHDNA | YSSELASCD |
| Md22 | 144aa | abs | abs | 0 | abs | FLHHDNA | QSPSSPDLAPCD |
| Md23 | 223aa$^{+ins}$ | abs | abs | 3 | TKDETW | FKDSA | YSPYLAPCD |
| Md24 | 342aa | MVR | LVPKTL | 0 | TGDETW | LHHDNA | YSPDLAPCD |
| Md25 | 220aa | abs | abs | 1 | TGDETW | FLHHDNA | YLPDLASCA |

"abs" indicates that the motif is missing. Asterix in motifs designs stop codon occurrence. "$^{ins}$" indicates insertion in the predicted transposase ORF.