

## Diversity of *mariner*-like elements in Orthoptera

K. Ustyantsev<sup>1</sup>✉, M. Biryukov<sup>1</sup>, I. Sukhikh<sup>1</sup>, N.V. Shatskaya<sup>1</sup>, V. Fet<sup>2</sup>, A. Blinov<sup>1,3</sup>, I. Konopatskaia<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

<sup>2</sup> Marshall University, Department of Biological Sciences, Huntington, USA

<sup>3</sup> Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia

✉ e-mail: ustyantsev@bionet.nsc.ru

*Mariner*-like elements (MLEs) are among the most widespread DNA transposable elements in eukaryotes. Insects were the first organisms in which MLEs were identified, however the diversity of MLEs in the insect order Orthoptera has not yet been addressed. In the present study, we explore the diversity of MLEs elements in 16 species of Orthoptera belonging to three infraorders, Acridoidea (Caelifera), Grylloidea (Ensifera), and Tettigoniidea (Ensifera) by combining data mined from computational analysis of sequenced degenerative PCR MLE amplicons and available Orthoptera genomic scaffolds. In total, 75 MLE lineages (Ortmar) were identified in all the studied genomes. Automatic phylogeny-based classification suggested that the current known variability of MLEs can be assigned to seven statistically well-supported phylogenetic clusters (I–VII), and the identified Orthoptera lineages were distributed among all of them. The majority of the lineages (36 out of 75) belong to cluster I; 20 belong to cluster VI; and seven, six, four, one and one lineages belong to clusters II, IV, VII, III, and V, respectively. Two of the clusters (II and IV) were composed of a single Orthoptera MLE lineage each (Ortmar37 and Ortmar45, respectively) which were distributed in the vast majority of the studied Orthoptera genomes. Finally, for 16 Orthoptera MLE lineages, horizontal transfer from the distantly related taxa belonging to other insect orders may have occurred. We believe that our study can serve as a basis for future researches on the diversity, distribution, and evolution of MLEs in species of other taxa that are still lacking the sequenced genomes. **Key words:** *mariner*-like elements; transposable elements; Orthoptera; Insects; horizontal transfer.

**For citation:** Ustyantsev K., Biryukov M., Sukhikh I., Shatskaya N.V., Fet V., Blinov A., Konopatskaia I. Diversity of *mariner*-like elements in Orthoptera. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2019;23(8): 1059-1066. DOI 10.18699/VJ19.581

## Разнообразие *mariner*-подобных элементов Orthoptera

К.В. Устьянцев<sup>1</sup>✉, М.Ю. Бирюков<sup>1</sup>, И.С. Сухих<sup>1</sup>, Н.В. Шацкая<sup>1</sup>, В. Фет<sup>2</sup>, А.Г. Блинов<sup>1,3</sup>, И.Д. Конопацкая<sup>1</sup>

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Университет Маршалл, Отделение биологических наук, Хантингтон, США

<sup>3</sup> Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия

✉ e-mail: ustyantsev@bionet.nsc.ru

*Mariner*-подобные элементы (МПЭ) – одни из самых распространенных мобильных элементов эукариот. Насекомые – первые организмы, в которых были обнаружены МПЭ, однако до сих пор еще не изучено разнообразие МПЭ насекомых отряда Orthoptera. В настоящей работе, совмещая результаты биоинформационного анализа секвенированных вырожденных ПЦР-ампликонов МПЭ и доступных геномных последовательностей Orthoptera, мы исследуем разнообразие МПЭ у 16 видов Orthoptera, принадлежащих к трем инфраотрядам: Acridoidea (Caelifera), Grylloidea (Ensifera) и Tettigoniidea (Ensifera). Всего среди всех изученных геномов выявлено 75 генетических линий МПЭ (Ortmar). Автоматизированная филогенетически опосредованная классификация подразделила все известное в настоящее время разнообразие МПЭ на семь кластеров со значимой статистической поддержкой (I–VII), в каждом из которых была найдена одна из генетических линий МПЭ Orthoptera. Большинство генетических линий (36 из 75) принадлежит к кластеру I; 20 линий относятся к кластеру VI; 6, 7, 4 и по 1 линии принадлежат к кластерам II, IV, VII, III и V соответственно. Два кластера, II и IV, были представлены единственными генетическими линиями МПЭ Orthoptera, Ortmar37 и Ortmar45, соответственно, распространенными среди большинства исследованных геномов Orthoptera. Мы также показали, что 16 генетических линий МПЭ Orthoptera могли участвовать в горизонтальном переносе с эволюционно удаленными таксонами насекомых из других отрядов.

**Ключевые слова:** *mariner*-подобные элементы; мобильные элементы; Orthoptera; насекомые; горизонтальный перенос.

### Introduction

DNA transposons were the first mobile genetic elements discovered in eukaryotes (McClintock, 1950). Since that time, different groups of these “selfish” elements were identified throughout the whole eukaryotic tree of life (Robertson,

2002; Feschotte, Pritham, 2007; Wicker et al., 2007). The most abundant and widespread group of DNA transposons is a *Tc1/mariner* superfamily (Plasterk et al., 1999; Tellier et al., 2015) named after its two best-studied elements, *Tc1* and *mariner* from a nematode *Caenorhabditis elegans* (Emmons

et al., 1983) and a fruit fly *Drosophila mauritiana* (Jacobson et al., 1986), respectively.

*Mariner*-like elements (MLEs) form the *mariner* family within the *Tc1/mariner* superfamily, MLEs are relatively small in size (about 1300 bp) characterized by the presence of 28–30 bp terminal inverted repeats (TIRs), an intronless gene encoding transposase, and a typical TA dinucleotide pattern of target site duplication (TSD) (Robertson, 1993; Robertson, MacLeod, 1993; Leroy et al., 2003). MLE transposases contain a common to *Tc1/mariner* N-terminal DNA-binding domain of helix-turn-helix type and a C-terminal catalytic domain with characteristic DD(34)D motif (Brillet et al., 2007). Former studies have described nine well-defined canonical subfamilies based on the similarities of the transposase catalytic core amino acid sequences (Robertson, 1997, 2002; Robertson, MacLeod, 1993; Green, Frommer, 2001): *vertumnana*, *cecropia*, *mauritiana*, *irritans*, *mellifera*, *capitata*, *elegans*, *briggsae* and *lineata*. This list was later supplemented by the more recently identified *marmoratus* and *drosophila* subfamilies (Bui et al., 2007; Wallau et al., 2014).

Probably, the most exciting feature of *Tc1/mariner* elements and MLEs in particular is their ubiquity. There are numerous reports suggesting horizontal transfers of MLEs between highly diverged taxa (Yoshiyama et al., 2001; Lampe et al., 2003; Silva et al., 2004, 2005; Oliveira et al., 2012; Dupeyron et al., 2014). Such a tendency to horizontal transfer is explained by (1) an evolutionary necessity for a long-term persistence of a DNA transposon, which otherwise will be eventually silenced by spontaneous mutations and cell's internal defense mechanisms (Lohe et al., 1995; Schaack et al., 2010), and (2) the simplicity of MLEs structure and life cycle (Plasterk, van Luenen, 2002). Presence of cis- or trans-encoded transposase and intact TIRs are the only features required for the transposition of MLEs (Plasterk et al., 1999). In addition, the MLE transposase was demonstrated to be fully active *in vitro* (Lampe et al., 1996; Tosi, Beverley, 2000). Such a low dependence on host-encoded and other biological factors most probably underlies success of MLEs as inter-taxa colonizers and their use as transgenic vectors (Delaurière et al., 2009).

Insects were the first organisms in which MLEs were identified and for which the horizontal transfer of MLEs was first proposed (Jacobson et al., 1986; Maruyama, Hartl, 1991; Robertson, 1993; Robertson, MacLeod, 1993). The insect order Orthoptera comprises two suborders: Caelifera (grasshoppers and locusts) and Ensifera (katydid and crickets). A computational analysis of the diversity of *Tc1/mariner* elements was carried out only for the genome of the migratory locust *Locusta migratoria* (Caelifera, Acrididae, Oedipodinae), one of the largest sequenced animal genome up to date (5.7 Gb) (Wang et al., 2014). However, this study provided no classification of the identified elements below the superfamily level (Bao et al., 2015). Only a few partial sequences belonging to MLEs were experimentally identified in two other grasshopper species (both in Caelifera, Acrididae): *Eyprepocnemis plorans* (Eyprepocnemidinae) (Montiel et al., 2012) and *Trimerotropis pallidipennis* (Oedipodinae) (Tevy et al., 2007). In addition, a recently published draft genome of the Hawaiian cricket *Laupala kohalensis* (Ensifera, Grylloidea, Trigonidiidae) is now also available for the analysis (Blankers et al., 2017). In this work, we explore the diversity of MLEs in Orthoptera

from three infraorders, Acridoidea (Caelifera), Grylloidea (Ensifera), and Tettigoniidea (Ensifera). We identify new lineages of MLEs and take a first look into their horizontal transfer dynamics among insects.

## Materials and methods

**Orthoptera specimens.** Specimens of 14 species of Orthoptera were provided by Dr. A.G. Bugrov, Institute of Systematics and Ecology of Animals, the Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia. We used 10 specimens of grasshoppers (Acridoidea), two of crickets (Grylloidea), and two of katydids (Tettigoniidea). The taxonomic position of each of the studied specimens is indicated in Table 1.

**DNA extraction, PCR amplification and sequencing.** Total DNA was extracted from muscle or brain tissue of the ethanol-preserved insect specimens using DNAeasy Blood and Tissue kit (Qiagen, Germany). We used canonical degenerate PCR primers specific for the conserved part of the MLE transposase gene catalytic domain to amplify the corresponding sequences for each of the DNA specimen. The primer sequences (MAR 124F and MAR 276R) and the cycling conditions are from the previous study (Robertson, MacLeod, 1993). PCR products were visualized and separated from unincorporated primers on 1.2 % agarose gel stained with EtBr. The resulting bands were cut out and then isolated from the gel for each of the studied Orthoptera specimens using QIAquick Gel Extraction Kit (Qiagen, Germany).

**Library preparation, sequencing and assembly.** Approximately 100 ng of each of the isolated 14 PCR amplicon mixes were used for barcode library preparation with Ion Xpress Plus Fragment Library Kit (Thermo Fisher Scientific) following the manufacturer's protocol. Fragmentation time was 6 min and 8 PCR amplification cycles were done. Purification of the libraries was done with AMPure XP kit (Beckman Coulter). A DNA fraction of 280–350 bp was extracted from the libraries using LabChip XT (Caliper Life Sciences) on a DNA500 chip with subsequent purification with AMPure XP kit. The quality and molarity of the obtained libraries was evaluated with DNA High Sensivity kit on the Bionanalyzer BA2100 chip (Agilent). The final libraries were amplified in emulsion PCR with Ion PGM Hi-Q View OT2 Kit (Thermo Scientific) and then sequenced on the Ion Torrent PGM system with 316v2 chip. Raw sequence reads produced by the Ion Torrent sequencing were *de novo* assembled into contigs using MIRA v. 4.0.2 (Chevreux et al., 2004), specifying “iontor” as technology parameter and “est, denovo, accurate” as job parameters in MIRA manifest configuration files.

**Availability of supporting data.** The genomic sequences used in this work are available at NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). The genomes of *Locusta migratoria* and *Laupala kohalensis* include 5,760 Mb covering 1,397,492 scaffolds and 1,595 Mb covering 148,784 scaffolds, respectively (Wang et al., 2014; Blankers et al., 2017). The genome sequences from other 259 insect species (actual on 11.07.2018) were retrieved from NCBI GenBank using the following Entrez searching query on the Assembly database: “Insecta”[Organism] NOT “Orthoptera”[Organism] AND (latest[filter] AND “scaffold level”[filter] AND “representative genome”[filter] AND all[filter] NOT anomalous[filter] filtering option.

**Table 1.** Studied Orthoptera species and the identified diversity of MLE lineages

Family	Subfamily	Species	MLE cluster**							Number of Ortmar lineages per species
			I (36)	II (1)	III (7)	IV (1)	V (6)	VI (20)	VII (4)	
Acrididae	Acridinae	<i>Acrida</i> sp.	20	1	1	1	3	5	0	31
	Eyprepocnemidinae	<i>Eyprepocnemis plorans</i>	16	1	2	1	1	6	0	27
	Gomphocerinae	<i>Notostaurus albicornis</i>	14	1	2	1	1	4	0	23
	Melanoplinae	<i>Sinopodisma punctata</i>	18	1	3	1	1	8	0	32
		<i>Bryodemella tuberculata</i>	18	1	2	1	3	6	0	31
	Oedipodinae	<i>Locusta migratoria</i> *	19	1	1	1	3	11	0	36
		<i>Oedipoda fedtschenki</i>	21	1	2	1	3	7	0	35
	Oxyinae	<i>Oxya japonica</i>	13	1	3	0	3	4	0	24
Pezotettiginae	<i>Pezotettix anatolica</i>	11	1	2	1	3	3	0	21	
Pyrgomorphidae	Pyrgomorphinae	<i>Chrotogonus</i> sp.	6	0	1	1	0	1	0	9
Pamphagidae	Pamphaginae	<i>Paranocaracris karadagi</i>	19	1	2	0	1	6	0	29
Tettigoniidae	Tettigoniinae	<i>Gampsocleis sedakovii</i>	9	1	4	1	2	2	1	20
		<i>Tettigonia viridissima</i>	10	1	1	1	1	1	1	16
Gryllidae	Gryllinae	<i>Cardiodactylus novaeguineae</i>	11	0	6	1	4	2	3	27
	Trogonidiinae	<i>Gryllus campestris</i>	14	1	3	1	3	1	1	24
	Eneopterinae	<i>Laupala kohalensis</i> *	13	0	2	0	1	12	3	31

Note. Clusters correspond to those in Figure. Taxonomy is according to NCBI Taxonomy. See Supplementary 4 for more details.

\* Species with available genomic scaffolds.

\*\* For each cluster, (N) designate total number of different Ortmar MLE lineages found.

**Detection of MLE sequences.** We collected a set of 19 known full-length transposase protein sequences belonging to eight MLE subfamilies (Supplementary 1)<sup>1</sup> and used them as queries in a batch tBLASTn search performed on the genomic sequences of *L. migratoria* and *L. kohalensis*. The recovered BLAST hits (High Scoring Pairs) were reassembled into copies if they were following the previously defined criteria (Wallau et al., 2014). Two hits from the same scaffold are jointed together if the distance between their centers is less than 1000 bp and if they have the same orientation. Only copies longer than 400 bp were retained. In order to extract the full-length sequence of each copy, the region of each assembled copy was expanded to 1500 bp. The copies were next clustered together into lineages by iterative BLASTn, setting the similarity threshold to 75 % which was previously defined in (Bouallègue et al., 2017), retrieving a majority rule consensus sequence for each lineage during the clustering process. Singlet sequences and consensus derived from less than 10 sequences were removed from further analysis. Finally, we attempted to identify TIRs at the ends of the lineage consensus using EMBOSS EINVERTED tool (Rice et al., 2000). Each newly identified complete sequence of Orthoptera MLE lineage was added to the query set and the whole search process was repeated to achieve the higher sensitivity.

The sequences of the query set obtained after the MLE screening of *L. migratoria* and *L. kohalensis* genomes were trimmed to the size of the conserved region of the MLE transposase catalytic domain which is between the canonical MLE degenerative primers (Robertson, MacLeod, 1993). Using this

set, we performed a batch tBLASTn search against the assembled PCR amplicon sequences of 14 other Orthoptera species. Contigs of expected length (between 450 and 550 bp) which gave positive BLAST hits with more than 90 % coverage to the query sequence were kept for further analysis. The retrieved contigs were clustered into lineages as previously described. To remove redundancy and to keep (when possible) only a full-length representative of a lineage for further analyses, we did a BLASTn search using the consensus sequences of obtained lineages as queries against the database of full-length Orthoptera MLE lineages identified in this work.

Finally, consensus sequences of the obtained MLE lineages, both full-length and partial, were used as queries in a batch BLASTn search against the sequences originally retrieved by tBLASTn search from all the studied Orthoptera species to estimate the prevalence and distribution of each lineage.

**Classification and phylogenetic analysis.** Corresponding full-length and partial protein sequences of MLE transposases were identified in lineage consensus using BLASTx search against the original transposase tBLASTn query set. The obtained protein sequences were aligned together with full-length transposase proteins of known MLEs from the previous work (Bouallègue et al., 2017) using MAFFT v7.312 (Katoh, Standley, 2013). The resulted alignment was used to build a Bayesian phylogeny of MLEs using the MrBayes v3.2.5 (Ronquist et al., 2012). The best-fitting model LG+I+G with four gamma rate categories was selected using SMS tool (Lefort et al., 2017) based on Akaike information criterion (Akaike, 1998). Bayesian posterior probabilities were estimated by running ten Markov chain Monte Carlo chains for 2,500,000 generations in two parallel runs, sampling each

<sup>1</sup> Supplementary Materials 1–6 are available in the online version of the paper: <http://www.bionet.nsc.ru/vogis/download/pict-2019-23/appx23.pdf>

250 generations, and discarding first 25 % of samples in the end. Automatic phylogeny-based classification of the sequences on the resulted tree was performed using Cluster Picker tool (Ragonnet-Cronin et al., 2013). The minimal statistical support was set to 90 % and various genetic distance thresholds from 40 to 60 % were tested.

**Horizontal gene transfer detection.** The full-length nucleotide consensus sequences of the identified Orthoptera lineages were used in MEGABLAST search against the genomic scaffolds of 259 insect species from different orders available in the NCBI. Cases of potential horizontal transfer between the studied Orthoptera genomes and other insect taxa were considered when MEGABLAST hit showed more than 90 % of identity covering more than 90 % of the query sequences as it was proposed by several authors (Filée et al., 2015; Wallau et al., 2016; Bouallègue et al., 2017).

## Results and discussion

Screening for MLE sequences was based on a homology approach (tBLASTn) with a set of 19 known full-length amino acid transposase sequences used as queries (see Supplementary 1). First, we initialized the search in the available genomic scaffolds of *L. migratoria* and *L. kohalensis* in order to identify Orthoptera full-length MLE sequences. A total of 4290 MLE copies from *L. migratoria* clustered in 27 lineages and 992 MLE copies from *L. kohalensis* clustered in 24 lineages were identified. Here, one lineage corresponds to a group of at least 10 sequences which have more than 75 % identity to each other detected by BLASTn comparisons.

Partial sequences of the conserved region of the MLE transposase catalytic domain were identified among the *de novo* sequenced PCR amplicons obtained from 14 other Orthoptera species. The sequences were then clustered into lineages following the same rules as described above (see Table 1). For each lineage, obtained both from the full-length or partial sequences, a majority-rule consensus sequence was derived (Supplementary 2). A total of 75 lineages were identified in all the studied Orthoptera genomes. For 49 of the lineages there was a full-length representative either from *L. migratoria* or *L. kohalensis* genome (Supplementary 3). There were no lineages distributed in all the genomes studied. Sixteen lineages were found only in the Caelifera, and 26, only in the Ensifera genomes. Thirty-three lineages were present both in Caelifera and Ensifera. Sixteen lineages were identified in a single genome each (Supplementary 4).

On average, 26 lineages were found for a studied orthopteran genome. The maximum (36 lineages) was identified in the genome of the migratory locust (*L. migratoria*) and the minimum (9 lineages) in the genome of an African grasshopper (*Chrotogonus* sp.) (see Table 1). It should be noted that the increase in the number of MLE lineages detected in *L. migratoria* (from 27 to 36) and *L. kohalensis* (from 24 to 31) genomes is due to the fact that these lineages are present as singlets or in less than 10 copies in the genomes and therefore were filtered out during the initial stringent mining process. Nevertheless, their minor presence is noted in Table 1 and Supplementary 4.

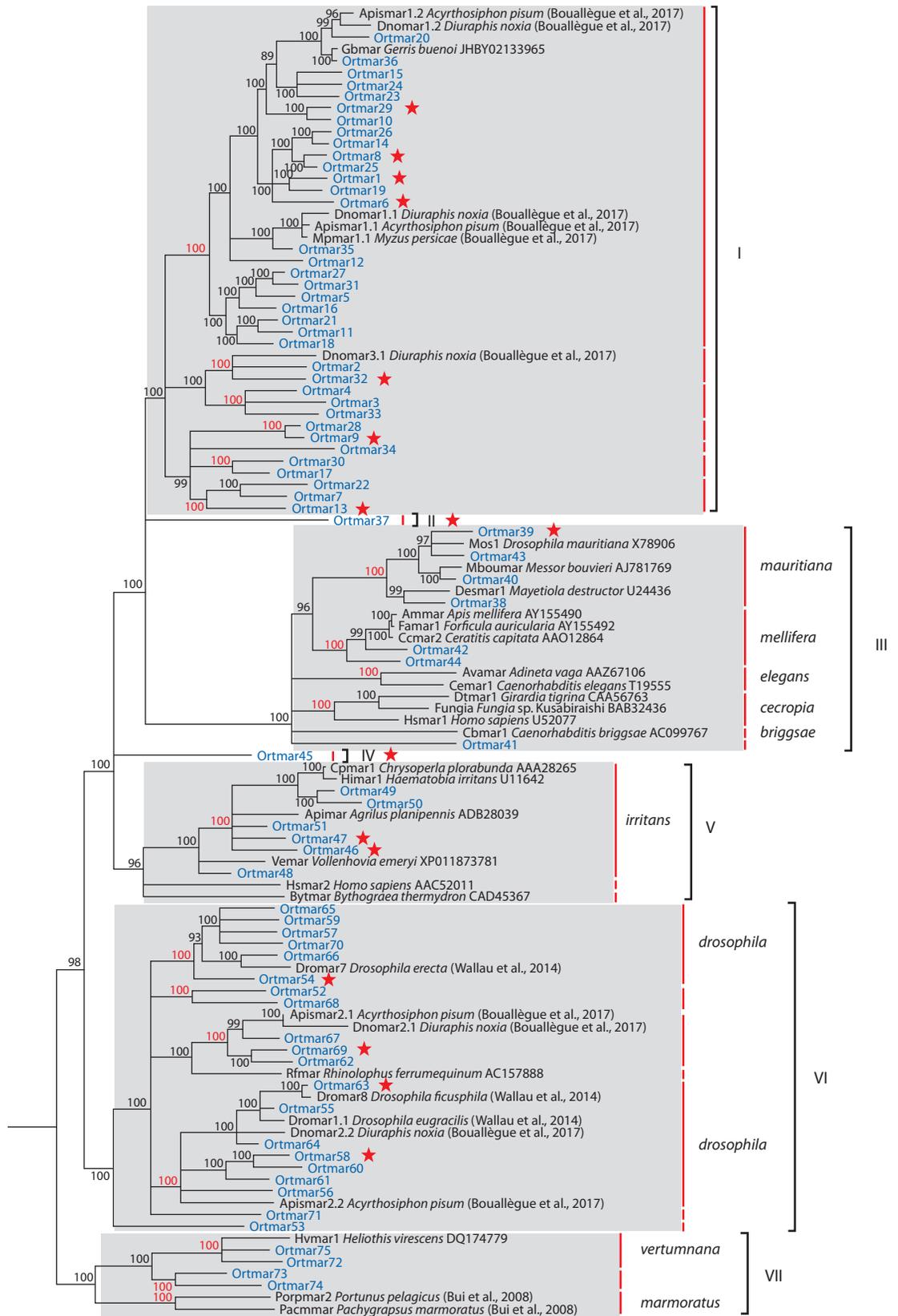
For the MLE lineage consensuses identified during the first stringent search in the *L. migratoria* and *L. kohalensis* genomes, TIRs as well as the characteristic to MLEs TA di-

nucleotide TSD have been identified. Among the 27 *L. migratoria* lineages TIRs were found in 24 lineages, and the TSD was found in 17. TIRs were found in all the 24 lineage consensuses from the *L. kohalensis* genome, 19 of which also had the TSD (see Supplementary 3).

Transposase amino acid sequences of the lineage consensuses were then used in phylogenetic analysis together with other known MLE transposases. To implement a phylogeny-based automatic classification of the identified lineages, we used the Cluster Picker tool (Ragonnet-Cronin et al., 2013) which assigns sequences into a cluster if (1) there is a substantial statistical support for their clustering on a tree and (2) a distance threshold between sequences in a cluster is satisfied. Testing different distance thresholds we found that at 43 % level all the “canonical” (historically first identified) MLE subfamilies (like *mauritiana*, *mellifera*, *cecropia*, etc.) still can be distinguished from each other, while increasing the threshold to 44 % already results in collapsing *mauritiana* and *mellifera* subfamilies into a single cluster. Using the 43 % threshold results in 28 clusters (including those represented by a single lineage), which can be supposedly of a subfamily level. However, these clusters at the same time are the parts of the higher-level ones (see Figure). For further convenience and the redundancy removal, we arbitrarily chose 55 % similarity threshold at which seven higher-level and statistically pronounced clusters (I–VII) are defined by Cluster Picker (see Figure, Table 1, and Supplementary 4). Four of these clusters (III, V, VI, and VII) include previously known subfamilies. Cluster III is comprised of *mauritiana*, *mellifera*, *cecropia*, *elegans* and *briggsae* subfamilies. Cluster V is represented solely by the *irritans* subfamily, even more supporting its status as such. Members of a recently defined *drosophila* subfamily (Wallau et al., 2014) were found within cluster VI. Cluster VII includes *vertumnana* and *marmoratus* subfamilies (Bui et al., 2007, 2008). Although no previously defined subfamilies were found grouped in cluster I, it includes elements recently identified in the insect order Hemiptera (Bouallègue et al., 2017) and assigned to *irritans* subfamily (which, however it is not supported by our phylogenetic analysis). Clusters II and IV are comprised of single Orthoptera lineage each (Ortmar37 and Ortmar45, respectively). However, these lineages were found in the vast majority (13 out of 16) of the studied genomes (see Figure, Table 1, and Supplementary 4).

The identified Orthoptera MLE lineages were distributed among all the seven clusters. The majority of the lineages (36 out of 75) belong to cluster I; 20 to cluster VI; and seven, six, four, one and one lineages, to the clusters II, IV, VII, II, and V, respectively (see Figure, and Supplementary 4). Interestingly, lineages of the cluster VII were identified only among the studied Ensifera genomes.

Horizontal transfer was suggested to be one of the main features explaining an extremely widespread occurrence of MLEs (Robertson, Lampe, 1995). To ensure the long-term survival, a transposable element has to be transferred horizontally or otherwise it would be eventually “fossilized” in the genome by vertical inactivation mechanisms (Lohe et al., 1995). Therefore, to fully understand the evolution of a transposable element lineage it is important to evaluate its horizontal transfer rate. For 16 lineages identified in the studied Orthoptera genomes, horizontal transfer between dis-



Phylogeny of MLE transposases reconstructed by a Bayesian analysis using the LG+I+G amino acid substitution model.

The tree was rooted with transposases from *rosa* and *LTIR Tc1/mariner* families (see Supplementary 6 for the names and accessions). Posterior probability is indicated for each node. Nodes with posterior probability below 90% are collapsed and not shown. Names of the 75 Orthoptera MLE lineages identified in this study start with "Ortmar" and highlighted in blue. Clusters corresponding to original MLE subfamilies and predicted by Cluster Picker (Ragonnet-Cronin et al., 2013) with 43% genetic distance threshold are indicated by bold red lines to the right of the tree, and their corresponding names (where applicable) are shown. Posterior probabilities of the nodes corresponding to subfamily-level clusters are highlighted in red. Higher-level clusters predicted by Cluster Picker with 55% genetic distance threshold are indicated with black bold brackets and assigned with Roman numerals. Red stars indicate Orthoptera lineages which may have been horizontally transferred between Orthoptera species and species from other insect orders (see Table 2 for more details).

**Table 2.** Cases of potential horizontal transfer of MLE lineages identified in the studied Orthoptera genomes between other insects detected by MEGABLAST search

Orthoptera lineage*	Non-Orthoptera insect orders**							
	Lepidoptera (4/45)	Hymenoptera (15/67)	Hemiptera (3/23)	Diptera (3/100)	Coleoptera (2/10)	Odonata (1/2)	Phasmatodea (1/2)	Blattodea (2/4)
Ortmar1 (13)	–	–	–	–	–	–	1/320/94.8	–
Ortmar6 (10)	–	–	–	–	1/10/94.0	–	–	1/2/92.4
Ortmar8 (10)	–	–	–	–	1/43/92.0	–	–	–
Ortmar9 (10)	–	–	–	–	–	–	–	1/80/92.9
Ortmar13 (9)	–	–	–	–	–	1/1/90.2	–	–
Ortmar29 (3)	–	–	–	–	–	–	–	1/12/91.4
Ortmar32 (1)	–	1/1/90.6	–	–	–	–	–	1/78/95.3
Ortmar37 (13)	–	–	–	–	–	1/10/97.8	–	–
Ortmar39 (11)	–	9/351/93.2	–	–	–	–	–	–
Ortmar45 (13)	–	–	–	–	–	–	–	1/1/90.5
Ortmar46 (11)	–	–	1/5784/95.4	–	–	–	–	–
Ortmar47 (10)	–	–	1/1/90.6	–	–	–	–	–
Ortmar54 (9)	–	–	2/19/97.0	–	–	–	–	–
Ortmar58 (4)	–	3/8/96.5	–	–	–	–	1/51/96.8	–
Ortmar63 (2)	4/6/97.4	8/135/98.4	1/56/95.9	3/49/99.3	–	–	–	–
Ortmar69 (1)	–	–	1/50/95.6	–	–	–	–	–

Note. See Supplementary 5 for more details.

\* For each lineage, (N) – number of studied Orthoptera genomes in which the lineage was identified;

\*\* for each non-Orthoptera order, (X/Y) – (number of genomes with detected potential horizontal transfers/number of genomes in the database); for the corresponding Orthoptera MLE lineage, values X/Y/Z designate number of genomes with MEGABLAST hit with more than 90 % identity/total number of MEGABLAST hits among the genomes/the maximum MEGABLAST identity (%) found.

tantly related species (divergence time over 250 Mya (<http://www.timetree.org/home>)) from other insect orders may have occurred (see Table 2, and Supplementary 5). Among these potential transfers, some draw an additional attention. For instance, high-identity (up to 99 %) MEGABLAST hits with the full-length sequence of the Ortmar63 lineage consensus from the *drosophila* subfamily (see Figure, cluster VI) were identified simultaneously in 16 genomes belonging to Diptera, Hemiptera, Lepidoptera, and Hymenoptera, while the lineage itself was identified in ten and two copies in two crickets only: *L. kohalensis* and *Cardiodactylus novaeguineae*, respectively (see Supplementary 4). A less prominent signal was found for the lineage Ortmar39 of the *mauritiana* subfamily identified in eleven Orthoptera genomes with 10 copies per genome on average (see Figure, cluster III). MEGABLAST hits to Ortmar39 (up to 92 % identity) were found for nine Hymenoptera genomes with the number of matches ranging from 1 to 257 (see Table 2, Supplementary 5). Two substantially diverged lineages from the cluster I (Ortmar8 and Ortmar9, each was found in ten studied Orthoptera genomes) gave 80 and 78 high scoring hits with a single genome of a termite *Cryptotermes secundus*. A substantial amount of high-identity hits with a genome of a stick insect *Clitarchus hookeri* (Phasmatodea) were found for the lineages Ortmar1 (320 hits up to 95 %, cluster I, distributed in 13 Orthoptera genomes) and Ortmar58 (51 hits up to 97 %, cluster VI, distributed in the 4 Orthoptera genomes). Another example is even more extraordinary: 5784 high-identity (up to 95 %) hits with the sequence of the Ort-

mar46 lineage of the *irritans* subfamily, widespread within the Orthoptera (see Figure, cluster V, Supplementary 4 and S5), were found in a single genome of the blood-sucking bug *Rhodnius prolixus* (Hemiptera) (see Table 2, and Supplementary 5).

Interestingly, previous studies showed that the *R. prolixus* genome is extremely saturated with members of a single lineage of the *irritans* subfamily, suggesting a relatively recent transposition burst with the estimated expansion occurred around the time of speciation of *R. prolixus*, approximately 1.4 Mya (Fernández-Medina et al., 2016). With Orthoptera species as a potential source, it can be hypothesized that the horizontal transfer of the Ortmar46 lineage could be a cause of the *irritans* subfamily expansion in the *R. prolixus* genome. Although in all the identified potential horizontal transfer cases the exact mechanism(s) and direction of the transfer remain elusive, these findings can later help in deciphering of each particular cases of the process in insects, as suggested by some authors (Silva et al., 2004; Loreto et al., 2008). Overall, the horizontal transfer cases identified here are in line with the recent report on massive horizontal transfer of transposable elements between insects which was based on more sophisticated genome to genome comparisons (Peccoud et al., 2017).

## Conclusion

In this study, we used a non-standard approach to investigate the variability of MLEs. First of all, we took advantage of modern high-throughput sequencing technologies to avoid the bias of non-exhausting sampling which would occur if

the standard molecular cloning of the PCR amplicons with the subsequent Sanger sequencing was used. Here, the Ion Torrent PGM was a platform of choice, however, any of the currently available next-generation sequencing technologies can be utilized. The robustness of the obtained results was additionally controlled by the results of computational analysis of the two genome sequences from distantly related taxa of Orthoptera, *L. migratoria* (suborder Caelifera) and *L. kohalensis* (suborder Ensifera). Secondly, the automatic phylogeny-based classification of MLEs with Cluster Picker (Ragonnet-Cronin et al., 2013) allowed the definition of clusters on a rational, researcher-independent basis. In all previous studies on MLEs, it was a choice of the researchers whether to include newly identified MLE lineages to previously defined clusters, or designate new ones. Using this approach, we were able to identify and classify MLE lineages in the diverse 16 species of Orthoptera (see Table 1 and Figure), for which so far the genomic sequences were available only for the two species. We believe that our study can serve as a basis for future researches on diversity, distribution, and evolution of MLEs in species of other taxa that are still lacking the sequenced genomes.

## References

- Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. New York: Springer, 1998;199-213.
- Bao W., Kojima K.K., Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*. 2015;6(1):11. DOI 10.1186/s13100-015-0041-9.
- Blankers T., Oh K.P., Bombarely A., Shaw K.L. The genomic architecture of a rapid island radiation: mapping chromosomal rearrangements and recombination rate variation in *Laupala*. *bioRxiv*. 2017; 160952. DOI 10.1101/160952.
- Bouallègue M., Filée J., Kharrat I., Mezghani-Khemakhem M., Rouault J.D., Makni M., Capy P. Diversity and evolution of *mariner*-like elements in aphid genomes. *BMC Genomics*. 2017;18(1):1-12. DOI 10.1186/s12864-017-3856-6.
- Brillet B., Bigot Y., Augé-Gouillou C. Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica*. 2007;130(2):105-120. DOI 10.1007/s10709-006-0025-2.
- Bui Q.T., Casse N., Leignel V., Nicolas V., Chénais B. Widespread occurrence of *mariner* transposons in coastal crabs. *Mol. Phylogenet. Evol.* 2008;47(3):1181-1189. DOI 10.1016/j.ympev.2008.03.029.
- Bui Q.T., Delaurière L., Casse N., Nicolas V., Laulier M., Chénais B. Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*. *Gene*. 2007;396(2):248-256. DOI 10.1016/j.gene.2007.03.004.
- Chevreaux B., Pfisterer T., Drescher B., Driesel A.J., Müller W.E.G., Wetter T., Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res*. 2004;14(6):1147-1159. DOI 10.1101/gr.1917404.
- Delaurière L., Chénais B., Hardivillier Y., Gauvry L., Casse N. *Mariner* transposons as genetic tools in vertebrate cells. *Genetica*. 2009; 137(1):9-17. DOI 10.1007/s10709-009-9370-2.
- Dupeyron M., Leclercq S., Cerveau N., Bouchon D., Gilbert C. Horizontal transfer of transposons between and within crustaceans and insects. *Mob. DNA*. 2014;5(1):4. DOI 10.1186/1759-8753-5-4.
- Emmons S.W., Yesner L., Ruan K.S., Katzenberg D. Evidence for a transposon in *Caenorhabditis elegans*. *Cell*. 1983;32(1):55-65.
- Fernández-Medina R.D., Granzotto A., Ribeiro J.M., Carareto C.M.A. Transposition burst of *mariner*-like elements in the sequenced genome of *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 2016;69: 14-24. DOI 10.1016/j.ibmb.2015.09.003.
- Feschotte C., Pritham E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 2007;41(1):331-368. DOI 10.1146/annurev.genet.40.110405.090448.
- Filée J., Rouault J.-D., Harry M., Hua-Van A. *Mariner* transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. *BMC Genomics*. 2015;16:1061. DOI 10.1186/s12864-015-2060-9.
- Green C.L., Frommer M. The genome of the Queensland fruit fly *Bactrocera tryoni* contains multiple representatives of the *mariner* family of transposable elements. *Insect Mol. Biol.* 2001;10(4): 371-386.
- Jacobson J.W., Medhora M.M., Hartl D.L. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 1986;83(22):8684-8688.
- Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013;30(4):772-780. DOI 10.1093/molbev/mst010.
- Lampe D.J., Churchill M.E., Robertson H.M. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* 1996; 15(19):5470-5479.
- Lampe D.J., Witherspoon D.J., Soto-Adames F.N., Robertson H.M. 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.* 2003;20(4):554-562. DOI 10.1093/molbev/msg069.
- Lefort V., Longueville J.-E., Gascuel O. SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 2017;34(9):2422-2424. DOI 10.1093/molbev/msx149.
- Leroy H., Castagnone-Sereno P., Renault S., Augé-Gouillou C., Bigot Y., Abad P. Characterization of Mcmar1, a *mariner*-like element with large inverted repeats (ITRs) from the phytoparasitic nematode *Meloidogyne chitwoodi*. *Gene*. 2003;304:35-41.
- Lohe A.R., Moriyama E.N., Lidholm D.A., Hartl D.L. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Mol. Biol. Evol.* 1995;12(1):62-72. DOI 10.1093/oxfordjournals.molbev.a040191.
- Loreto E.L.S., Carareto C.M.A., Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity (Edinb.)*. 2008; 100(6):545-554. DOI 10.1038/sj.hdy.6801094.
- Maruyama K., Hartl D.L. Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J. Mol. Evol.* 1991;33(6):514-524.
- McClintock B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA*. 1950;36(6):344-355. DOI 10.1073/pnas.36.6.344.
- Montiel E.E., Cabrero J., Camacho J.P.M., López-León M.D. *Gypsy*, *RTE* and *Mariner* transposable elements populate *Eyprepocnemis plorans* genome. *Genetica*. 2012;140(7-9):365-374. DOI 10.1007/s10709-012-9686-1.
- Oliveira S.G., Bao W., Martins C., Jurka J. Horizontal transfers of *Mariner* transposons between mammals and insects. *Mob. DNA*. 2012; 3(1):14. DOI 10.1186/1759-8753-3-14.
- Peccoud J., Loiseau V., Cordaux R., Gilbert C. Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci. USA*. 2017;114(18):4721-4726. DOI 10.1073/pnas.1621178114.
- Plasterk R.H., Izsvák Z., Ivics Z. Resident aliens: the *Tc1/mariner* superfamily of transposable elements. *Trends Genet.* 1999;15(8):326-332.
- Plasterk R.H.A., van Luenen H.G.A.M. The *Tc1/Mariner* family of transposable elements. In: *Mobile DNA II*. Washington, DC: American Society of Microbiology, 2002;519-532.
- Ragonnet-Cronin M., Hodcroft E., Hué S., Fearnhill E., Delpech V., Brown A.J., Lycett S. Automated analysis of phylogenetic clusters. *BMC Bioinformatics*. 2013;14(1):317. DOI 10.1186/1471-2105-14-317.
- Rice P., Longden I., Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000;16(6):276-277.

- Robertson H.M. The *mariner* transposable element is widespread in insects. *Nature*. 1993;362(6417):241-245. DOI 10.1038/362241a0.
- Robertson H.M. Multiple *mariner* transposons in flatworms and hydras are related to those of insects. *J. Hered.* 1997;88(3):195-201. DOI 10.1093/oxfordjournals.jhered.a023088.
- Robertson H.M. Evolution of DNA transposons in eukaryotes. In: *Mobile DNA II*. Washington, DC: American Society of Microbiology, 2002;1093-1110.
- Robertson H.M., Lampe D.J. Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. *Mol. Biol. Evol.* 1995;12(5):850-862.
- Robertson H.M., MacLeod E.G. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol. Biol.* 1993;2(3):125-139.
- Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 2012;61(3):539-542. DOI 10.1093/sysbio/sys029.
- Schaack S., Gilbert C., Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* 2010;25(9):537-546. DOI 10.1016/j.tree.2010.06.001.
- Silva J.C., Bastida F., Bidwell S.L., Johnson P.J., Carlton J.M. A potentially functional *mariner* transposable element in the protist *Trichomonas vaginalis*. *Mol. Biol. Evol.* 2005;22(1):126-134. DOI 10.1093/molbev/msh260.
- Silva J.C., Loreto E.L., Clark J.B. Factors that affect the horizontal transfer of transposable elements. *Curr. Issues Mol. Biol.* 2004;6(1):57-71.
- Tellier M., Bouuaert C.C., Chalmers R. *Mariner* and the *ITm* superfamily of transposons. *Microbiol. Spectr.* 2015;3(2):MDNA3-0033-2014. DOI 10.1128/microbiolspec.MDNA3-0033-2014.
- Tevy F., Guzman N., Gonzalez G., Lia V., Poggio L., Confalonieri V.A. Mobile elements and inverted rearrangements in *Trimerotropis pallidipennis* (Orthoptera: Acrididae). *Caryologia*. 2007;60(3):212-221. DOI 10.1080/00087114.2007.10797939.
- Tosi L.R., Beverley S.M. Cis and trans factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Res.* 2000;28(3):784-790.
- Wallau G.L., Capy P., Loreto E., Hua-Van A. Genomic landscape and evolutionary dynamics of *mariner* transposable elements within the *Drosophila* genus. *BMC Genomics*. 2014;15(1):1-19. DOI 10.1186/1471-2164-15-727.
- Wallau G.L., Capy P., Loreto E., Le Rouzic A., Hua-Van A. VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the *Mariner* family within *Drosophila*. *Mol. Biol. Evol.* 2016;33(4):1094-1109. DOI 10.1093/molbev/msv341.
- Wang X., Fang X., Yang P., Jiang X., Jiang F., Zhao D., Li B., Cui F., Wei J., Ma C., Wang Y., He J., Luo Y., Wang Z., Guo X., Guo W., Wang X., Zhang Y., Yang M., Hao S., Chen B., Ma Z., Yu D., Xiong Z., Zhu Y., Fan D., Han L., Wang B., Chen Y., Wang J., Yang L., Zhao W., Feng Y., Chen G., Lian J., Li Q., Huang Z., Yao X., Lv N., Zhang G., Li Y., Wang J., Wang J., Zhu B., Kang L. The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* 2014;5:2957. DOI 10.1038/ncomms3957.
- Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhou B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 2007;8(12):973-982. DOI 10.1038/nrg2165.
- Yoshiyama M., Tu Z., Kainoh Y., Honda H., Shono T., Kimura K. Possible horizontal transfer of a transposable element from host to parasitoid. *Mol. Biol. Evol.* 2001;18(10):1952-1958. DOI 10.1093/oxfordjournals.molbev.a003735.

#### ORCID ID

K. Ustyantsev orcid.org/0000-0003-4346-3868  
I. Sukhikh orcid.org/0000-0002-3548-4354

**Acknowledgements.** The work of K. Ustyantsev and A. Blinov was funded by the Russian State Budget (Project No. 0324-2019-0040). The work of M. Biryukov was funded by the Russian State Budget (Project No. 0259-2019-0010). We thank Dr. A.G. Bugrov for providing the insect specimens and the Center of Genome Investigation of the Institute of Cytology and Genetics SB RAS for the sequencing assistance.

**Conflict of interest.** The authors declare no conflict of interest.

Received May 28, 2019. Revised June 25, 2019. Accepted June 28, 2019.